# Network effects in systemic risk propagation

**Barjašić, Irena**

University of Zagreb

FACULTY OF SCIENCE

Irena Barjašić

# NETWORK EFFECTS IN SYSTEMIC RISK PROPAGATION

DOCTORAL DISSERTATION

Zagreb, 2022

University of Zagreb

FACULTY OF SCIENCE

Irena Barjašić

# NETWORK EFFECTS IN SYSTEMIC RISK PROPAGATION

DOCTORAL DISSERTATION

Supervisor:
Vinko Zlatić, PhD

Zagreb, 2022

Sveučilište u Zagrebu

PRIRODOSLOVNO MATEMATIČKI FAKULTET

Irena Barjašić

# MREŽNI UTJECAJ U PROPAGACIJI SISTEMSKOG RIZIKA

DOKTORSKI RAD

Mentor:
dr. sc. Vinko Zlatić

Zagreb, 2022.

## Supervisor's biography

Vinko Zlatić is a senior research associate and a head of Condensed Matter and Statistical Physics group in the Division of Theoretical Physics Division at the Ruđer Bošković Institute in Zagreb, Croatia. He has a background in theoretical physics of complex systems and has received his doctorate degree from the University of Zagreb, Croatia. After finishing his doctorate degree, he was a postdoctoral researcher at the Sapineza University of Rome, Italy 2008-2010. In 2012 he became a research associate at the Ruđer Bošković Institute and since 2018 he is a senior research associate.

Vinko's research focuses on probabilistic models of complex networks with special emphasis on topological connectivity and new types of percolations. He is also working on a number of applied problems, with special focus on systemic risk in financial systems. Besides coordinating Ruđer Bošković nodes on a number of FP7 and H2020 projects, Vinko has coordinated the Twinning project for the Division of Theoretical Physics. He has also applied his research in risk in commercial projects with Croatian Banking Union and Croatian Depositary Agency.

# Acknowledgements

First and foremost, I would like to thank my supervisor Vinko Zlatić, PhD for his guidance, support and invaluable discussions. I would also like to thank Prof. Stefano Battiston from the University of Zurich for the successful collaboration, and new perspectives. Additionally, this endeavour would not have been possible without the support from the QuantiXLie Centre of Excellence, a Project co-financed by the Croatian Government and European Union through the European Regional Development Fund - the Competitiveness and Cohesion Operational Program (Grant KK.01.1.1.01.0004), who financed my research. Finally, I would like to thank my friends and family for the support and encouragement.

## Abstract

The role of the network in the systemic risk propagation is investigated from two different perspectives, one being the detection of the default propagation from available data and the other the quantification of its effect on the value of debt. In the first part of this thesis, a methodology that uses causal network motifs is devised to distinguish the existence of network propagation from the minimal required data on the process. It is extensively tested against artificial data and the conditions when it is applicable are found. An example with the real Croatian default data is provided. In addition to that, analytical results on the temporal evolution of the causal motifs are obtained for a better understanding of the methodology. In the second part, we study the valuation of debt in a credit exposure network that is also exposed to external shocks on equity, when default propagation is taken into account. From the default process with external shocks, which is both simulated and theoretically modeled, we obtain the expected fraction of defaulted banks. We provide the result depending on the process and network parameters and compare it to the case where the network propagation is not taken into account. We show that the introduction of correlation on the shocks makes diversification of risk less efficient, and a network contribution to the probability of default exists even in the case of an infinite fully diversified network. We draw a direct parallel with the calculation of the Credit Valuation Adjustment and stress the need for including the network effects into the existing methodology.



Keywords: complex networks, systemic risk, financial contagion

# Prošireni sažetak rada

## Uvod

Smjerovi razmišljanja i istraživanja u fizici mogu se podijeliti na dva filozofska pravca: redukcionizam i holizam [1]. Prvi je stoljećima vodio fizičare u pokušajima da raščlane i objasne mnoštvo raznolikih prirodnih pojava. Raznolikost materije prvo je atomska teorija sažela u periodni sustav elemenata, a kasnije je dobiven standardni model, koji svodi trenutni opis prirode na konačan broj elementarnih čestica te obuhvaća tri od četiri poznate sile kojima međudjeluju. Zakoni gibanja koje je otkrio Newton povezali su gibanje na Zemlji s gibanjem nebeskh tijela, a Einstein je generalizacijom principa relativnosti na neinercijalne sustave izveo gravitaciju kao posljedicu geometrije prostor-vremena. Također, redukcionizam postoji i u ostalim prirodnim znanostima, tendencija svođenja bioloških procesa na kemijske procese koji su im u pozadini nastavlja se objašnjavanjem kemijskih procesa fizikom.

S druge strane, neki fenomeni, poput topline ili tlaka, pojavljuju se tek kad se u obzir uzme kolektivno gibanje čestica, te se ne mogu iščitati iz pojedinih čestica koje prate Newtonove zakone gibanja. Holistički pristup istraživanju fokusira se na prirodnim pojavama u kojima je fenomen kao cjelina više od samog zbroja njegovih dijelova. Korelacije u fizikalnim sustavima, koje se pojavljuju kad je sustav u kritičnom stanju, povezuju sustav na svim skalama te zahtjevaju opis pomoću cijelog sustava umjesto uprosječivanja fizikalnih veličina. Uz klasične sustave gdje su fazni prijelazi pod utjecajem vanjskih parametara, postoje i sustavi koji spontanom evolucijom dolaze do kritičnog stanja. Takva pojava naziva se samoorganizirana kritičnost, te je karakteriziraju rijetki, ali značajni događaji koji nastaju kombinacijom jednostavnih interakcija, nelinearnosti sustava te dugodosežnih vremenskih i prostornih korelacija [2].

Područje koje se bavi istraživanjem sustava sa kolektivnim svojstivma naziva se teorija kompleksnih sustava. Preciznije, ako kompleksnost pojava proizlazi iz topološki netrivijalnih interakcija, kao formalizam za proučavanje koristi se teorija mreža. Matematička struktura na kojoj se temelji teorija mreža je graf, koji komponente sustava promatra kao skup čvorova, a njihove interakcije kao parove čvorova, koji čine bridove. Bridovi na mreži mogu biti usmjereni ili neusmjereni, binarni ili težinski, ovisno o sustavu koji opisujemo.

Najpoznatiji model stohastičkih grafova naziva se Erdős-Rényi graf [3], po matematičarima Paulu Erdősu i Alfrédu Rényiju koji su prvi počeli koristiti stohastičke grafove za opis realnih kompleksnih sustava. Prednost modela je u njegovoj jednostavnosti, budući da ga određuju samo broj čvorova i srednji stupanj povezanosti čvora [4] (ili broj bridova [3]), no zbog toga ne reproducira neka svojstva realnih mreža, poput relativno male udaljenosti među čvorovima, nastajanja klastera te distribucije stupnja vezanja koja slijedi zakon potencija [5]. Modeli poput Watts-Strogatz modela [6] i Barabási-Albert modela [7] koji su predloženi kasnije, bolje reproduciraju navedena svojstva, no Erdős-Rényi zbog svoje jednostavnosti te lako dohvatljivih teorijskih rezultata ostaje i dalje u upotrebi.

Kompleksne mreže područje su koje povezuje razne znanstvene discipline, od fizike i spinskih stakala, preko mreža proteina i hranidbenih lanaca u biologiji, prometnih mreža do interneta i tehnoloških mreža. Multidisciplinarni pristup spaja doprinose iz teorije grafova, statističke fizike, mrežnih metrika razvijenih u psihologiji, računalnih znanosti, statističkih metoda, itd.

U ovoj tezi bavit ćemo se proučavanjem financijskih mreža. Slom financijskog sustava može nastati kao posljedica vanjskog šoka, no može i proizaći kao makroskopski efekt koji je posljedica kolektivnog ponašanja komponenti u sustavu. Nakon globalne financijske krize 2008. godine, opis financijskih sustava kao kompleksnih, otvorio je mogućnost predviđanja nestabilnosti u sustavu [8]. Razumijevanje načina na koji se šokovi šire kroz sustav, te svojstava mreže koja doprinose pojačavanju ili gušenju šokova, od značajne je važnosti regulatorima koji se brinu za stabilnost financijskog sustava [9, 10].

Doprinosi provedenog doktorskog istraživanja su slijedeći: **i)** razvijena je metodologija koja detektira prisutnost endogenog širenja u sustavu iz dostupnih podataka, **ii)** napravljen je teorijski izračun očekivanih vrijednosti kauzalnih motiva, **iii)** napravljena je analiza mrežnog doprinosa vjerojatnosti bankrota, direktno povezana s izračunom valuacije duga, te **iv)** izveden je teorijski opis širenja koreliranih šokova kroz financijski sustav.

### Pregled literature

Matematički, graf je definiran kao par $G = (V, E)$, gdje je $V$ skup elemenata koje nazivamo čvorovi, a $E$ skup parova čvorova, nazvanih bridovi. Teorija mreža nastala je kad su nakon opisa realnih mreža pomoću grafova pronađena univerzalne statističke značajke, kao što

su "mali svijet", nastajanje klastera, te specifična svojstva distribucije stupnjeva vezanja. Terminom "mali svijet" opisano je svojstvo realnih mreža da se od nekog čvora do bilo kojeg drugog može doći u relativno malom broju koraka, to jest da prosječni najkraći put na mreži raste s logaritmom broja čvorova $N$. Visoki stupanj klasteriranja svojstven je za realne mreže, kao i distribucija stupnjeva vezanja koja slijedi zakon potencija. Prema tome, generativni modeli iz teorije mreža teže reproducirati ova svojstva [11]. Modeli koje koristimo u ovoj disertaciji su Erdős-Rényi graf [3] te stohastički k-regularni graf. Erdős-Rényi graf prvi je konstruirani model stohastičkog grafa, a parametri koji određuju distribuciju iz koje se izvlače pojedine realizacije su broj čvorova $N$ i srednji stupanj vezanja čvora $\langle k \rangle$, ili broj čvorova $N$ i broj bridova $m$. Unatoč tome što nema visok stupanj klasteriranja niti zakon potencija u distribuciji stupnjeva vezanja, Erdős-Rényi graf se i dalje koristi zbog svoje jednostavnosti i dostupnosti analitičkih rezultata. Stohastički k-regularni graf definira stupanj vezanja čvorova, koji je jednak točno $k$ za svaki čvor, dok su bridovi među čvorovima nasumično postavljeni. Podgrafovi koji se pojavljuju u češće nego očekivano na realnim mrežama nazivaju se mrežni motivi, te mogu pružiti uvid u svojstva mreža i mrežnih procesa. Prvi put su upotrijebljeni u [12], u pokušaju da se pomoću njih definiraju klase univerzalnosti kompleksnih mreža. U istraživanju predstavljenom u ovoj disertaciji, definiramo kauzalne mrežne motive kako bismo pomoću njih klasificirali dinamičke procese na mreži.

Proučavanje dinamičkih procesa na mreži pokušaj je da se povežu kolektivni makroskopski efekti s mikroskopskom mehanikom koja ih pokreće. Određuju ih tranzicije između mikroskopskih stanja sustava. Budući da je egzaktan pristup problemu nemoguć zbog veličine sustava, pri računanju putanja sustava koristi se probabilistički opis [13]. Prva dva dinamička procesa koje koristimo u ovoj disertaciji pripadaju klasi međudjelujućih sustava čestica [14, 15, 16, 17], koja se sastoji od stohastičkih procesa na mrežama. Za simulaciju širenja propadanja koristimo kontaktni proces, koji je u epidemiologiji poznatiji kao SI proces. Kod kontaktnog procesa, vjerojatnost prenošenja zaraze proporcionalna je broju zaraženih susjeda, što dobro opisuje uvjete širenja bankrota. Kod drugog procesa iz klase međudjelujućih sustava čestica koji koristimo, nazvanog model glasovanja [18], stanje čvora mijenja se s intenzitetom proporcionalnim udjelu susjeda u određenom stanju. Osim stohastičkih procesa na mrežama, koristmo i deterministički proces, model praga [19], u kojem su stanja čvorova na mreži određena pragom koji udio susjeda u drugom stanju

mora prijeći kako bi i središnji čvor promijenio stanje. Probabilistički model kojim se opisuje putanja mikroskopskih stanja diskretnih sustava naziva se master jednadžba [20]. Zbog diskretnosti stanja na čvorovima, ovaj opis se prirodno nameće za dinamičke sustave na kompleksnim mrežama, no egzaktno rješenje master jednadžbe u većini slučajeva nije moguće. Aproksimacijom srednjeg polja [13], gdje pretpostavljamo da je sustav homogen te da nema korelacijama među mikrovarijablama, moguće je pojednostavniti master jednadžbu te dobiti rješenje. Međutim, pretpostavke potrebne za aproksimaciju srednjeg polja ne vrijede u svim slučajevima, te najčešće rješenje nije jako precizno. Aproksimativne master jednadžbe [21, 22] umjesto vremenskog razvoja stanja jednog uprosječenog čvora promatraju skup srednišnjeg čvora s njegovim prvim susjedima, te dobivaju puno preciznije rješenje. Proširenje aproksimativnih master jednadžbi na usmjerene mreže s vremenskim uređenjem dio je teme je kojom smo se bavili u prvom dijelu doktorske disertacije.

Dinamičke procese na mreži može se klasificirati preko uzroka zbog kojih se promjena stanja čvorova događa, a po njihovoj prirodi ti uzroci mogu se podijeliti na egzogene i endogene. Klasifikacija dinamičkih procesa na egzogene i endogene tema je brojnih znanstvenih radova [23, 24, 25, 26], koji iz oblika distribucije šokova te pretpostavljenih funkcionalnih ovisnosti procijenjuju je li se dogodilo endogeno širenje. Doprinos ove teze detekciji endogenih procesa je neparametarska metoda koja ne pretpostavlja postojanje širenja na mreži *a priori*.

Sistemski rizik definira se kao rizik da veliki dio sustava propadne. Zbog isprepletenosti kompleksnih sustava, takav događaj može biti uzrokovan i propadanjem samo jedne jedinke u sustavu, ako to propadanje pokrene kaskadni proces. Doprinos teorije mreža proučavanju sistemskog rizika dan je kroz radove koji: predlažu mrežnu metriku za računanje važnosti pojedinih institucija [27], proučavaju dvosmislen utjecaj diversifikacije na smanjenje sistemskog rizika [28, 29], uvode Pigouovo oporezivanje kao odvraćanje institucija od rizičnog ponašanja [30, 31], te istražuju optimalnu topologiju sustava i nivo asortativnosti za stabilnost sustava [32, 33, 34]. Širenje sistemskog rizika može se odvijati direktnim širenjem bankrota na financijskoj mreži, no do povećanje rizika može doći i zbog širenja devaluacije duga među institucijama [35]. Uslijed povećanja vjerojatnosti propadanja neke institucije, vrijednost njenih dugova na tržištu duga se smanjuje, te je

važno za njene vjerovnike da uračunaju takve gubitke. U drugom dijelu ove disertacije bavimo se proučavanjem mrežnog doprinosa vjerojatnosti propadanja neke institucije, s direktnom primjenom na računanje prilagodbe kreditne valuacije (Credit Valuation Adjustment) [36] koja se koristi u financijskoj industriji za izračun gore spomenutih gubitaka.

**Detekcija endogenih procesa**

Svojstveno je za kompleksne sustave da iz jednostavne mikroskopske dinamike mogu nastati makroskopska svojstva koja utječu na cjelokupno stanje sustava. Primjere takvog ponašanja vidimo u rapidnom širenju epidemija, ili kaskadnom širenju bankrota na financijskoj mreži. Za neometano funkcioniranje te stabilnost takvog sustava nužno je detektirati efekte nagle propagacije u samom začetku. Sličan problem se pojavljuje pri modeliranju nekog određenog dinamičkog procesa na mreži, budući da ne možemo nužno pretpostaviti da postoji endogena komponenta procesa samo zato jer je podatke koje imamo moguće predstaviti kao mrežu. Dakle, u prvom dijelu ove doktorske disertacije zanima nas, trebamo li koristiti interaktivne čestične sustave za modeliranje procesa iz nekih podataka samo zato jer ih je moguće predstaviti mrežom, ili je nekad dovoljno predstaviti proces kao utjecaj vanjskog polja. Kako bismo odgovorili na ovo pitanje, razvili smo neparametarsku metodu koja ne pretpostavlja unaprijed postojanje endogenog procesa, te ga detektira iz podataka. Podatke koje koristimo mora biti moguće predstaviti mrežom, što znači da nam trebaju agenti koji su nekom trenutku doživjeli promjenu stanja, vrijeme te promjene te relacije među tim agentima koje mogu prenositi promjene stanja. Iako je moguće uvrstiti više podataka na mrežu, naš cilj je prvo izgraditi metodu na minimalnom potrebnom skupu podataka. Metoda se temelji na opservaciji da se pojavnost bridova s kauzalnim uređenjem vremena, u slučaju procesa vođenog vanjskim poljem, ne bi trebala puno razlikovati od pojavnosti nakon što permutiramo vremena bankrota na čvorovima. S druge strane, ako postoji endogena komponenta procesa koja se širi po mreži, pretpostavljamo da bi takvi bridovi trebali biti učestaliji nego na mreži s permutiranim vremenima. Definiramo testne statistike u obliku broja kauzalnih motiva na mreži - acikličkih motiva koji se sastoje isključivo od kauzalnih bridova, te ih rangiramo po broju bridova. Uz kauzalne motive kao testnu statistiku uzimamo i veličinu najveće kauzalne mrežne komponente. Za generiranje nul-distribucije koristimo randomizirane referentne modele (RRM) [37], koji čine mikrokanonski ansambl mreža koji permutira

vremena bankrota po čvorovima, a fiksira sva ostala svojstva mreže. Metodu testiramo na simuliranim procesima, na kojima možemo kontrolirati doprinos endogene i egzogene komponente. Proces za simulaciju definiramo pomoću Poissonovih procesa s intenzitetima $\alpha$ i $\beta$, za egzogenu i endogenu komponentu, te definiramo njihov omjer $\zeta \coloneqq \alpha/\beta$. Strukturu endogenog procesa modeliramo prema SI modelu te prema modelu glasovanja. Nakon simulacija, kao rezultat dobijemo distribuciju frekvencija testnih statistika iz originalnog procesa te iz permutiranih RRM ansambala. Analizu podataka počnemo primjenom Kolmogorov-Smirnov testa na te dvije distribucije kako bismo utvrdili razlikuju li se međusobno. Zatim koristimo z-statistiku kako bismo ustvrdili postoji li statistički signifikantna razlika između pojedinog procesa i njegovog RRM-modela. U slučaju kad je testna statistika kauzalni motiv višeg reda, možemo iskoristiti činjenicu da se motivi višeg reda sastoje od podmotiva, koji se i u slučaju potpuno egzogenog procesa pojavljuju na mreži s različitim vjerojatnostima. Stoga koristimo općenitu verziju z-statistike, Mahalanobisovu udaljenost [38], koju možemo shvatiti kao korijen zbroja kvadrata pojedinih z-statistika u slučaju kad je matrica kovarijacije dijagonalna. Kad je endogena komponenta modelirana kao SI proces, iz rezultata vidimo da je kauzalni motiv s jednim bridom najbolji u detekciji endogenih procesa, ako se koristi z-statistika, te ako se nalazimo u sredini procesa. Povećanjem broja bridova u kauzalnom motivu povećava se razmak između broja motiva u originalnom procesu i srednje vrijednosti nul-distribucije, no usporedno raste i varijanca nul-distribucije, što otežava detekciju. Korištenjem kauzalnih motiva s dva brida u kombinaciji s Mahalanobisovom udaljenošću signifikantnost rezultata pomiče se prema kraju procesa. Kod korištenja modela glasovanja za endogeni proces, rezultati su kvalitativno slični, no detekcija endogenih procesa je slabija za sve statistike, te prestaje biti moguća već za $\zeta > 1$. Kako bismo bolje razumjeli ovaj rezultat, riješili smo jednadžbe srednjeg polja za omjer broja egzogeno i endogeno propalih čvorova, za oba tipa endogenih procesa. Iz rješenja vidimo da je za isti parametar procesa $\zeta$, omjer čvorova uvijek veći za proces glasovanja, što objašnjava otežanu detekciju. Također, za oba tipa procesa vidimo da se povećavanjem mreže prostor signifikantnih rezultata širi i za z-statistiku i za Mahalanobisovu udaljenost. Nakon validacije na simuliranim podacima, primjenili smo metodu na realnim podacima hrvatskih firmi koje su ušle u predstečajnu nagodbu. Selekcionirali smo firme koje su i vjerovnici i dužnici, te 549 firmi sa 1198 dugova predstavili kao mrežu. Rezultati analize kauzalnim motivima ukazuju na to da je endogeno širenje postojalo u

ranijim fazama procesa, te kad je proces bio na 70-85% razvoja. Korištenje naše metode na podacima u realnom vremenu ukazalo bi na začetak kaskadnog procesa, te potaknulo regulatore na daljnje istraživanje uzroka bankrota. Uz ekstenzivno provođenje simulacija, te kako bismo bolje razumjeli rezultate, proširili smo aproksimativne master jednadžbe [21, 22] na usmjerene mreže s vremenskim uređenjem, i dobili izraz za vremensku evoluciju očekivanih vrijednosti kauzalnih motiva.

**Mrežni doprinos u strukturi vjerojatnosti bankrota**

Osim kaskadnog širenja bankrota po mreži, utjecaj na porast sistemskog rizika može imati i širenje devaluacije duga. Nakon sklapanja kreditnog ugovora, moguće je dalje trgovati njime na sekundarnom tržištu, gdje se njegova vrijednost procjenjuje preko raznih faktora, jedan od kojih je i vjerojatnost bankrota dužnika. Dakle, samo povećanje vjerojatnosti da će neki dužnik bankrotirati, za posljedicu ima smanjenje vrijednosti svih njegovih dugova, što se dalje reflektira na kapitalu vjerovnika, te može pokrenuti kaskadu devaluacija i time povećati cjelokupni sistemski rizik. Upravo ovakva vrsta gubitaka nalazila se iza dvije trećine gubitaka [39] vezanih za kreditne rizike druge ugovorne strane u financijskoj krizi 2008. Stoga je u regulatornom sporazumu Basel III uvedena nova vrsta rizika, nazvana kreditna valuacijska prilagodba (Credit Valuation Adjustment - CVA) [40]. Kreditna valuacijska prilagodba definira se kao razlika između nominalne vrijednosti duga i vrijednosti duga kad se uračuna mogući kreditni rizik druge ugovorne strane [36]. Od svih faktora koji čine formulu za CVA, u drugom dijelu ove disertacije fokusirat ćemo se na izračun vjerojatnosti bankrota. Dvije uhodane vrste modela bave se izračunom vjerojatnosti bankrota - strukturalni i intenzitetni modeli [41]. Strukturalni modeli koriste kapitalnu strukturu firme kako bi procijenili vjerojatnost da će ta firma bankrotirati, dok intenzitetni modeli kalibriraju stohastičke procese, pomoću kojih modeliraju vremena bankrota, preko tržišnih instrumenata. Međutim, svi navedeni modeli promatraju samo individualne institucije, ne uračunavajući pritom rizik od bankrota do kojeg može doći kaskadnim širenjem putem mreže kreditnih izlaganja. Cilj našeg istraživanja je pokazati da postoje slučajevi kad nije opravdano zanemariti doprinos kaskadnih procesa širenja bankrota na mreži u izračunu CVA. Niz radova koji se bave ovom temom, [42, 43, 44] izračunali su egzaktnu vjerojatnost bankrota, uz pretpostavke malog broja institucija, komonotonosti šokova, te nedostupnosti svih informacija o mreži. Naš pristup problemu

ne bazira se na egzaktnom izračunu stanja svake institucije u sustavu, što za vrlo velike sustave nosi sa sobom "kletvu dimenzionalnosti" (curse of dimensionality), već pomoću simulacija šokova i širenja financijskih oštećenja po mreži kao rezultat dobivamo distribuciju očekivanog udjela bankrotiranih institucija na mreži. Također, ovakav pristup daje nam slobodu oko modeliranja korelacija na vanjskim šokovima, za što koristimo Gaussovu kopulu [45]. Mehanizam širenja oštećenja po mreži uzimamo iz [46], te pretpostavljamo da su vanjski šokovi usmjereni na kapital. Vanjskim šokovima pripisujemo moguće diskretne vrijednosti, s pripadajućim vjerojatnostima, te uvodimo korelacije među njima, koristeći Gaussovu kopulu. Model mrežne topologije koji koristimo je stohastički k-regularni graf, budući da je na njemu bilo moguće zadovoljiti uvjete koje smo zadali za kapitalnu vrijednost i financijsku polugu svake institucije, bez većih komplikacija kod simulacija. Također, ovaj izbor topologije odgovara empirijskim podacima za jezgre bankovnih mreža [47]. Uz simulativni pristup, dobiveno je i popratno teorijsko rješenje, korištenjem modela praga [19]. Distribucija institucija pogođenih određenim šokom dobivena je uz pomoć svojstava Gaussove kopule, te nakon uvrštavanja u aproksimaciju srednjeg polja za model praga [19], možemo analitički izraziti očekivani udio bankrotiranih institucija za dane parametre. Prednost ovakvog pristupa je u mogućnosti jednostavnog dohvaćanja limesa beskonačne mreže. Rezultati simulacija dobiveni su u obliku distribucije udjela bankrotiranih institucija na svakoj simuliranoj mreži, koja ima multimodalni oblik, te za konačni rezultat uzimamo njenu očekivanu vrijednost. U rezultatima za sustave bez korelacija na vanjskim šokovima vidimo dvije moguće faze sustava nakon diversifikacije mreže, nadkritičnu i podkritičnu, ovisno o financijskoj poluzi. Nadkritična faza podrazumijeva da će sve institucije propasti, a podkritična da će se očekivana vjerojatnost bankrota spustiti na vrijednost dobivenu samo od vanjskog šoka. Međutim, kad se na vanjske šokove uvedu korelacije, ponašanje sustava se promijeni kvantitativno i kvalitativno. U slučaju kad su korelacije prisutne, diverzifikacija više ne može ukloniti doprinos mreže vjerojatnosti bankrota. Također, za nadkritični režim sustava bez korelacija, uvođenjem korelacije vjerojatnost bankrota se spušta, što objašnjavamo promjenom oblika distribucije šokova, koji uvođenjem korelacija dobiva nagib prema nižim vrijednostima, te time šokovi slabiji od šoka dovoljnog za potpuni bankrot postaju vjerojatniji. Konačno, provjerili smo utjecaj veličine sustava na rezultate, te zaključili da veličina sustava igra ulogu kad korelacije nisu prisutne na šokovima, dok u slučaju s koreliranim šokovima konačni rezultat ne ovisi

o veličini mreže, što je potkrijepljeno i teorijskim izračunom za slučaj beskonačne mreže.

**Zaključak**

U ovoj doktorskoj disertaciji proučavali smo mrežne efekte u širenju sistemskog rizika iz dvije različite perspektive. Jedna se odnosila na detektiranje mrežne propagacije iz podataka, dok je druga imala naglasak na kvantifikaciji doprinosa mreže vrijednosti duga. U prvom dijelu razvili smo metodologiju za detektiranje endogene propagacije iz podataka koji se mogu prikazati kao temporalna mreža. Testirali smo metodologiju na velikom rasponu parametara simuliranih procesa, te dobili statistički značajne rezultate do $\zeta \sim 1$, za mreže od 1000 čvorova. Razliku simuliranih distribucija smo prvo testirali Kolmogorov-Smirnov testom te potvrdili njeno postojanje. Zatim smo uspoređivali individualne procese s odgovarajućim RRM ansamblima te koristili z-statistiku i Mahalanobisovu udaljenost za detekciju endogene komponente procesa. U slučajevima koji su na granici prostora parametara gdje je detekcija bila moguća, signifikantnost rezultata ovisi o postotku mreže koji je bankrotirao, srednjem stupnju čvora, te veličini mreže. Povećanjem mreže, povećava se i raspon parametra $\zeta$ za koji je moguće detektirati endogeno širenje. Kao najbolje testne statistike, pokazali su se kauzalni motiv s jednim bridom kad se koristi z-statistika, te kauzalni motiv s dva brida kad se koristi Mahalanobisova udaljenost. Na prostoru parametara, ove dvije mjere pokazale su se komplemetarnima; dok motiv s jednim bridom najbolje funkcionira u srednjim fazama procesa, motiv s dva brida postaje signifikantan pri kraju procesa. Kako bismo bolje razumjeli vremensku evoluciju testnih statistika, proširili smo aproksimativne master jednadžbe na usmjerene mreže s vremenskim uređenjem, te dobili ovisnost očekivane vrijednosti kauzalnih motiva o vremenu, koja se slaže sa simulacijama. Metoda je također primjenjena na empirijskim podacima hrvatskih firmi u predstečajnim nagodbama, te je ukazala na mogućnost mrežne propagacije predstečajnih nagodbi, koju bi trebalo dalje istražiti u financijskim izvještajima firmi.

Drugi dio disertacije sadrži istraživanje strukture vjerojatnosti bankrota, te kvantifikaciju doprinosa toj vjerojatnosti od strane mreže kreditnih izlaganja na kojoj se institucije nalaze. Ovo istraživanje donosi popravku na račun prilagodbe kreditne valuacije, koji trenutno ne uzima u obzir moguću propagaciju bankrota po mreži. Simulirali smo sustav koji se sastoji od svih informacija o mreži kreditnih izlaganja. Nakon izlaganja mreže

nekoreliranim šokovima, sustav je nakon diverzifikacije zauzeo ili nadkritično stanje, gdje su sve institucije bankrotirale, ili podkritično stanje, u kojem se vjerojatnost bankrota smanjila do određenog nivoa. Uvođenjem korelacija na šokove, ponašanje sustava se mijenja, te više nije moguće potpuno ukloniti mrežni utjecaj diverzifikacijom. Također, za sustave koji su prethodno bili u nadkritičnom stanju, uvođenjem korelacija vjerojatnost bankrota se smanjuje. Provjeren je i utjecaj veličine sustava na ove efekte, te je pokazano da nisu posljedica konačne veličine sustava. Uz simulirane rezultate, dobili smo i analitičko rješenje, pomoću modela praga te distribucija vanjskih šokova koje smo izveli. Iz analitičkih rezultata izveli smo i asimptotsko ponašanje za beskonačne mreže. Konačno, možemo potvrditi da postoje slučajevi kad nije moguće zanemariti utjecaj mreže na vjerojatnost bankrota, te bi ih trebalo uvrstiti u formalizam za računanje CVA.

# Contents

# List of Tables

# List of Figures

4.4 **Switching between sub- and supercriticality.** On x-axis the degree of the network starts from 252 and goes up to a complete network, $k = 19998$, and on the y-axis we plot the expected fraction of default, $\langle q \rangle$. In the first row the initial probability of default is fixed, $\vec{p}_{up} = (0.09, 0.083, 0.827)$ and we change the shock sizes from left $\vec{\sigma}_{left} = (-1.1, -0.5, 0)$ to right $\vec{\sigma}_{right} = (-1.1, -0.25, 0)$. In the second row we use a different initial probability of default $\vec{p}_{down} = (0.05, 0.086, 0.864)$ for the same pair shock sizes. Interbank leverage takes values from the range in $\Lambda_b = \{0.2, 1, 2, 4, 5, 6, 9, 12, 16\}$ and it is shown that with the decrease of the initial probability, or the shock size, the system behaviour stays the same if we increase $\Lambda_b$ appropriately. In the panel insets, we plot the dependence of $\langle q \rangle$ on the interbank leverage $\Lambda_b$, for the chosen values of $k$ that are marked on the x-axis of the plot containing the inset. . . . . 81

4.5 **Probability of defaults with correlation between shocks**. The degree of the nodes $\langle k \rangle$ starts from 211 and goes up to 19998 (complete network), for the number of nodes $N = 10000$. The plots do not show a smaller $\langle k \rangle$ than 211, as the transitional effects in that range are due to the choice of the network, and not relevant to real systems. At time $t = 0$ the shock values are $\vec{\sigma} = (-1.1, -0.75, 0)$ , with probabilities $\vec{p} = (0.02, 0.09, 0.89)$, respectively. We vary the interbank leverage $\Lambda_b = \{0.2, 1, 2, 4, 5, 8, 12, 14\}$. The red horizontal line represents the initial probability of default (without taking into account the network of liabilities). The grey dashed line represents double the values of the initial probability of default. Subfigures **a)-d)** show different correlation coefficients $\rho = \{0, 0.1, 0.2, 0.3\}$. Values where the introduced correlation increases the probability of default by 100% or more are shown in colour, values below that are shown in greyscale. 84

4.6 **Probability of default and the system scale** Probability of default is plotted on the y-axis for maximally diversified (complete) networks with the numbers of banks $N = \{300, 500, 800, 1000, 3000, 5000, 8000, 10000\}$ on the x-axis. At time $t = 0$ the shock values are $\vec{\sigma} = (-1.1, -0.75, 0)$ , with probabilities $\vec{p} = (0.02, 0.09, 0.89)$, respectively. In different subplots (**a)-d)**) we vary the correlation coefficient from 0 to 0.3. The subplots indicate that the correlation-induced probability of default does not depend on the system size, which is further supported with horizontal lines that represent the analytical result for the limit of an infinite network. . . . . 86

# Chapter 1

# Introduction

## 1.1 Physics, complexity and financial systems

The directions of thought and research in physics, and science in general, follow two complementary philosophical principles: reductionism and holism [1]. An elegant and tempting principle such as reductionism, that aims to reduce the myriads of natural phenomena to a set of simple principles, has been followed by scientists over centuries. The diversity of matter was simplified through the atomic theory to the periodic system of elements, and then later even further to the standard model, that includes a finite number of elementary particles and three fundamental forces. The laws of motion, that Newton introduced, unified the motion on Earth with the motion of celestial objects, whereas Einstein generalized the principle of relativity to the non-inertial systems and described gravitation as an effect of the geometry of spacetime. Outside physics, a reductionist tendency decomposed biological processes to chemical processes, while chemical laws revealed to have physics in the background. Even for contemporary physicists, the ultimate aim is to find the set of equations that govern all the phenomena in the nature - a theory of everything.

A different perspective, however, was to start with some basic principles and try to build up from them to the observed phenomena. It is well known that individual particles, each simply obeying well known Newton's laws of motion, give rise to new physical phenomena of heat and pressure when observed as a collective. Collective or cooperative effects appear

in many natural processes, showing that "the whole is more than the sum of the parts" [2]. One of the most basic expressions of the cooperative phenomena is the central limit theorem; it demonstrates how random variables from various distributions, in the infinite limit, add up to a sum that forgets about the individual distributions' properties, and instead arises with a Gaussian shape [2].

Continuous scale invariance exhibited by physical systems stands behind the definition of a critical state, one between order and disorder. As the correlations exist at all scales and exclude the possibility of averageing of physical properties, this state needs to be described by taking the entire system into account [2]. The distributions of variables in such systems universally emerge as power laws. Renormalization group [48], the mathematical formalism developed to tackle critical phase transitions, uses coarse graining of the system to provide a map between different scales, and describe the macroscopic behaviour. When the continuous scale invariance is reduced to discrete scale invariance, the power law exponent becomes a complex number, thus introducing log-periodic oscillations [2]. Such oscillations have been found to have a predictive power in earthquakes [49, 50, 51], ruptures of engineering structures [52] and financial crashes [53, 54, 55, 56]. Percolation theory [57] provides yet another view on phase transitions. A simple mathematical model describes in the first approximation a wide generality of phenomena from various disciplines, such as contact processes [15], branching and annihilating random walks with odd parity [58, 59], population dynamics, transport phenomena, theory of liquid water, of stock market price fluctuations, the human vasculature system, earthquake nucleation, rupture in random media, etc.

In addition to the classical phase transitions driven by parameters, some dynamical systems, that are out of equilibrium or heterogeneous, have the property that the attractor of their evolution is a critical point. Such a behaviour is termed self-organized criticality, and is characterized by rare and large events, that appear out of a simple set of rules combined with non-linear interactions and long-range spatial and temporal correlations. The term was first introduced in Bak, Tan and Wiesenfeld's paper [60], where they present a simple sandpile model exhibiting spontaneous criticality and complexity. Seismicity, river networks, propagation of forest fires and biological evolution processes [61, 62] are some of the systems proposed to be described by self organized criticality, although the

mechanisms and the universality of the phenomena are still under debate.

Quenched randomness in some media contradicts the concept that disordered systems can be represented by effective and equivalent homogenized counterpart systems [63]. The full probability distribution of the quantities is necessary for the full description [1]. The physical systems exhibiting such behaviour are spin glasses [64], which show the breaking of ergodicity due to the the multi-valley structure of the energy landscape.

Complex systems theory is a broad field that emerged from the latter direction of research, in an attempt to generalize the behaviour of the various systems that exhibit emerging phenomena. Network theory, more specifically, studies the complexity that arises from topologically nontrivial interactions. The mathematical structure used to represent such systems is a graph, with the constituents classified as a set of vertices, and the interactions as pairs of those vertices - edges. Depending on the system being described, the interactions can be directed or undirected, binary or weighted.

A simple stochastic graph [3] was proposed by Erdős and Rényi as an elementary tool for studying complex networks. This marked the beginning of the use of random graphs for modeling real-world networks. Erdős-Rényi graph is defined by fixing either the number of vertices and the number of edges [3], or the number of vertices and the average number of edges per vertex - the mean degree [4], while randomizing all the other properties. Its properties enabled researchers to study the characteristics of random graphs in an analytically tractable way, and it is an important baseline model in statistical comparisons. Nevertheless, due to its simplicity, Erdős-Rényi stochastic graph does not reproduce some features exhibited by the real networks, for example a relatively short path between two nodes, cluster formation and power law degree distribution [5].

Watts-Strogatz model [6], generated by the interpolation between an ordered finite-dimensional lattice and a random graph, was proposed to reproduce the property of clustering on a network, as well as the short path between two nodes. These two network properties together define a family of "small-world" network models developed subsequently [65], and are an important tool in the study of social networks.

Power law distribution of node degrees was attained by “scale free” models [7], which form by following the dynamics of network assembly. The Barabási-Albert model mimics the

dynamics of the real networks in two principles: growth and the preferential attachment. The network obtains a new node in every time step, and the edge between the new node and the old ones is created with the probability proportional to every node's degree, with the final degree distribution arising as a power law.

Systems of interest come from all kinds of different disciplines, starting from some of the aforementioned physical systems, e.g. spin glasses [64], protein [66] and food networks [67] from biology, social networks [68], traffic networks [69], the Internet [70], etc. The non-triviality of the topological features places the study of complex networks outside the scope of the graph theory, and it invites a multidisciplinary approach. Tools from statistical mechanics such as maximum entropy principle serve for network reconstruction and pattern validation [71, 5] and various computer algorithms are needed to perform calculations on large networks [72]. Notable experiments such as Milgram's small world experiment [73] belong to psychology, and metrics such as centrality measure [74, 75] and clustering coefficients [76] were inspired by social networks. Finally, machine learning [77], statistical methods and hypothesis testing are absolutely necessary for the empirical part of the discipline.

Several authors have combined networks with game theory to study under which conditions human cooperation arises over defection [78]. Patterns formed in neuronal networks have served for studying and predicting the onset of mental illnesses in individuals [79]. The neural networks have been studied under the scope of network theory, and the graph structure of the networks with the best predictive performance is shown to have a very similar structure with biological neural networks [80].

In this thesis we will focus our attention on financial networks in particular. A financial crash can be a result of an extreme external shock or emerge as a macroscopic effect from the behaviour of the multitude of participants in the financial system. If the system loses resilience internally, i. e. enters a different "phase", the shock that drives it into collapse needs not be large. As it was considering only the participants as individuals, and leaving out the system they constitute as a whole, the traditional economic theory was not able to predict collapses of the financial system. After the 2008 crisis, the introduction of the view of complexity into the financial systems opened the door to the possible detection of a coming financial crash. Introducing the interactions into the picture, and representing

the system as a complex network, enabled researchers to find properties of the system that indicate a change in the system's stability, and could serve as early warning signals of a coming crisis [8]. Understanding the way that an external shock propagates through the network, which properties of the network contribute to the amplification, and which to the damping of the shock is an important matter regulators need to understand if they want to keep the system stable. Due to the confidentiality of the financial data, complete knowledge of the system cannot be obtained, however, statistical tools that reconstruct network exist [9]. Access to larger amounts of data and the consensus of the institutions to disclose more detailed data would enable researchers to build models that encompass more heterogeneity and lead in the direction of a more stable and resilient financial system [10].

## 1.2 Summary of contributions

We present original contributions to the field of complex networks, and in particular financial networks. The list of contributions, in the order in which they will be presented in the thesis is:

- the development of a methodology that helps distinguish a presence of an endogenous propagation in the system in the available data
- the extension of the approximate master equations to directed time-ordered networks and the calculation of the expected numbers of causal motifs
- the analysis of the contribution of network contagion to the probability of default of an institution in the system, directly applicable to the calculation of the Credit Valuation Adjustment
- a theoretical description of financial contagion with correlated exogenous shocks

## 1.3 Dissertation outline

This dissertation is organized in 5 chapters. In Chapter 2 we put the research we conducted into context, and write a literature review of the topics and concepts that we used in our research. Chapter 3 presents the methodology that was devised to detect an endogenous contribution in the propagation of defaults, using causal motifs. For a better understanding of the methodology, an analytical description of the temporal evolution of the motif frequency is provided. In Chapter 4 we present the research on the contribution of the contagion on the financial network to the default probability of institutions. We study the case with and without correlations of exogenous shocks. In addition to that, we include an analytical description of the financial contagion when shocks on institutions are correlated. All the results are discussed in Chapter 5, and the directions for future research are suggested.

# Chapter 2

# Complex networks and systemic risk: Background

We introduce the main topics necessary for understanding the work done in the thesis: graph theory, contact processes on complex networks and systemic risk. We provide a literature review for these topics and explain how they intertwine in the description of financial systems.

## 2.1 From graph theory to complex networks

### 2.1.1 Graph theory

Mathematically, a graph is defined as a pair $G = (V, E)$, where $V$ is a set of of elements named vertices, and $E \subseteq \{e_{ij} = \{v_i, v_j\} | v_i, v_j \in V\}$ is a set of pairs of vertices, which are called edges. In the case that the pairs of vertices are unordered, the graph is called undirected, whereas, if the pairs of vertices are ordered, the graph is said to be directed. In this Chapter, all the formulas will refer to undirected graphs, unless stated otherwise. Additionally, numerical values can be assigned to edges. Depending on whether they are all equal to 1, or have positive real values, the graph is either unweighted or weighted.

In case all possible edges are present, the graph is said to be complete, and if there are no edges it is empty. These properties are indicated as $K^n$ and $E^n$, respectively. A loop

on a graph is defined as an edge that has the same vertex as both elements of the pair, $e_{jj} = (v_j, v_j)$. A multigraph is a graph that contains loops and multiple edges between the same vertices.

A walk on a graph is a finite or an infinite sequence of vertices, joined by a sequence of edges. If a walk has all the different edges, it is called a trail, and if all the vertices are different, we call it a path. A trail with the first and the last vertices equal is called a cycle. In case a graph contains no cycles, it is called acyclic. If a path exists between any two vertices in a graph, we say that the graph is connected.

The graph of the size $N$ (containing $N$ vertices) is represented with an $N \times N$ adjacency matrix, defined as:

$$A_{ij} = \begin{cases} 1, & \text{if an edge exists between } i \text{ and } j \\ 0, & \text{otherwise.} \end{cases} \tag{2.1}$$

If the graph is directed, an edge is defined to point from vertex $i$ to vertex $j$, which makes the matrix $A_{ij}$ asymmetric. In case of multiple edges, their values are added up, and for loops the convention is $A_{ii} = 2$, as it makes the calculations neater. For weighted networks, the weights are assigned to elements $A_{ij}$ [72, 11].

The components of a graph are defined as the connected subgraphs that are not contained in any larger connected subgraph. An important metric of a graph is the size of the largest component. If the graph is directed, we can define a largest strongly connected component, which requires a directed path to exist between any two vertices, and a largest weakly connected component, which would be the largest component of the same graph with undirected edges.

### 2.1.2 Complex networks

An important step from graph theory towards network science was in the realization that real networks universally share some statistical features, namely the small-world, the clustering and the degree distribution properties.

The shortest path (or a graph geodesic) between two vertices on a network is defined as the path for which the sum of the weights of its edges is minimized. A global statistical measure is represented by an average of all the shortest paths $\langle l \rangle$ on a network. It was observed on many natural networks, that the average shortest path scales logarithimically or slower with the size of the network, and the phenomenon was named the "small-world" effect [11].

In many real world networks, the small-world property is observed together with the high level of clustering. Clustering is a concept that refers to the tendency of natural networks to form cliques. Locally, it is defined as a ratio of all the existing edges between the neighbours of a vertex $i$ (with $k_i$ neighbours) and all the possible edges between them. If we define the neighbourhood of a vertex $i$ as $N_i = \{v_j : e_{ij} \in E, e_{ji} \in E\}$, we can define the clustering coefficient for that vertex as:

$$C_i = \frac{2|\{e_{jk} : v_j, v_k \in N_i, e_{jk} \in E\}|}{k_i(k_i - 1)}. \tag{2.2}$$

In order to construct the clustering coefficient globally, we employ triplets, which are defined as three vertices that have two or three edges between them, and named an open and closed triplet respectively. The global clustering coefficient is then:

$$C = \frac{\text{number of closed triplets}}{\text{number of both open and closed triplets}}. \tag{2.3}$$

A high global clustering coefficient (higher than if the network ties were made randomly) is a well known property of real networks, and it is especially pronounced in social networks [72, 11].

The distribution of the vertex degrees $P(k)$ provides us with an important characterization of the network. Depending on the functional form of the degree distribution, we can broadly form two classes of real networks, homogeneous and heterogeneous networks. Homogeneous networks have degree distributions with light-tails, such as Poisson or Gaussian distributions. The presence of the light-tail, and the mean value that overlaps with maximum indicates that there exists a characteristic scale in the network. On the contrary, heavy-tailed distributions, such as power-law, emerge from the heterogeneity of the network connectivity pattern. It is specific for the power law distributions with the

exponent $\gamma \leq 2$ that they don't have a well defined mean, whereas for the exponent $\gamma \leq 3$ the variance is not well defined. Therefore, the heavy tail of a distribution, for $\gamma$ less than 3, implies that there is a finite probability of finding vertices with a degree much larger than the mean value. Since these distributions are not bell-shaped, the mean value does not represent a characteristic scale of the system [11].

The giant component $\mathcal{G}$ [13] of a network is defined as a connected component that contains a finite fraction of the network's vertices and thus diverges in the limit $N \to \infty$. The problem of finding a giant component in a network is analogous to infinite-dimensional edge percolation in the limit $N \to \infty$. The critical point divides the regime of a fragmented network from the regime where a giant component arises. A simple calculation can provide the condition for a giant component on graphs with a local tree structure with no cycles. For an undirected, uncorrelated network with the degree distribution $P(k)$, we can write the probability $q$ that a randomly picked edge does not lead to the giant component. It is stated as a sum over all possible degrees $k$ of a product of the probability that an edge leads to a vertex with degree $k$, and the probability that none of the other $k-1$ edges of that vertex are connected to the giant component in a self-consistent equation:

$$q = \sum\nolimits_k \frac{kP(k)}{\langle k \rangle} q^{k-1}. \tag{2.4}$$

Then, the probability that a given vertex belongs to the giant component is $P_G$ can be deduced from the complementary probability that a vertex does not belong to the giant component, $\sum_k P(k)q^k$:

$$P_G = 1 - \sum\nolimits_k P(k)q^k. \tag{2.5}$$

Except for the solution $P_G = 0$ for $q = 1$, which always exists, we can check for the existence of another solution by rewriting the Equation (2.4) as $q = F(q)$ and looking for a point where $F(q)$ intersects the line $q$. Considering the properties of the function $F(q)$, $F(0) = P(1)/\langle k \rangle$, $F(1) = 1$, $F'(q) > 0$ and $F''(q) > 0$ for $0 < q < 1$, the point of intersection can be obtained from the condition:

$$\frac{d}{dq}\left(\sum\nolimits_k \frac{kP(k)}{\langle k \rangle} q^{k-1}\right)\Big|_{q=1} > 1, \tag{2.6}$$

which can be simplified to:

$$\frac{\langle k^2 \rangle}{\langle k \rangle} > 2. \tag{2.7}$$

Therefore, the critical point separates the phase for $\langle k^2 \rangle / \langle k \rangle < 2$, in which the size of the largest component scales as $\ln N$, from the phase for $\langle k^2 \rangle / \langle k \rangle > 2$, which contains a giant component. Exactly at the critical point, where $\langle k^2 \rangle / \langle k \rangle = 2$, the largest component follows the power law $N^{2/3}$.

### 2.1.3 Erdős-Rényi graph

Paul Erdős and Alfred Rényi [3] were the first ones to apply probabilistic methods to the theory of graphs, which can be considered as the beginning of the random graph theory. They defined the graph $G(N, m)$, that was later named after them, by fixing $N$ vertices and $m$ randomly placed edges. A probability space with $\binom{\binom{N}{2}}{m}$ equiprobable graph realizations is obtained that way. An equivalent, but a mathematically more tractable, model was introduced in [4]. Instead of fixing the number of edges $m$, they fixed the probability $p$ that an edge will appear, consequently fixing the expected number of edges $E(m) = p\binom{N}{2}$. The probability of a graph realization that has $m$ edges then becomes:

$$P(G(N, m)) = p^m (1-p)^{\binom{N}{2} - m}. \tag{2.8}$$

The degree distribution can be simply derived, if we notice that the probability that the vertex is connected to $k$ vertices, and not connected to the remaining $N - k - 1$ is equal to:

$$P(k) = \binom{N-1}{k} p^k (1-p)^{N-1-k}. \tag{2.9}$$

If we fix the mean degree to be constant $\langle k \rangle = pN$ and take the limit of an infinite network $N \to \infty$, the binomial distribution can be approximated by the Poisson distribution:

$$P(k) = e^{-\langle k \rangle} \frac{\langle k \rangle^k}{k!}. \tag{2.10}$$

Since the Poisson distribution has very light tails, Erdős-Rényi graph is a typical representation of homogeneous networks with a very well defined scale and small degree fluctuations.

It is possible to easily derive the expression for the clustering coefficient $\langle C \rangle$, since the probability of an edge between two vertices is independent of the existence of any other edge:

$$\langle C \rangle = p = \frac{\langle k \rangle}{N}. \tag{2.11}$$

In the limit of infinitely large networks, if $\langle k \rangle$ is fixed, the clustering coefficient goes to zero, which shows that Erdős-Rényi network does not represent the real networks well in this property.

The calculation of the average shortest path $\langle l \rangle$ is straightforward as well. We assume that the cycles are negligible, and that every vertex has approximately $\langle k \rangle$ neighbours. Starting from one vertex and iterating $\langle l \rangle$ times, we can reach the entire network, i.e. $\langle k \rangle^{\langle l \rangle} \approx N$. It follows that:

$$\langle l \rangle \approx \frac{\ln N}{\ln \langle k \rangle}. \tag{2.12}$$

It is clear from Eq. (2.12) that the Erdős-Rényi graph recovers the small-world property of the real networks [11, 5, 13].

The condition for the emergence of the giant component in an Erdős-Rényi network can be easily obtained by inserting the equality that stands for the Erdős-Rényi network, $\langle k^2 \rangle = \langle k \rangle + \langle k \rangle^2$ into the Equation (2.7), which then provides us with the critical point $\langle k \rangle_c = 1$. Therefore, an Erdős-Rényi network will have a giant component as soon as the mean degree $\langle k \rangle$ becomes greater than 1. That corresponds to the critical probability $p_c = 1/N$. Since, in this thesis, we are concerned with contagion processes on networks, we need to have the existence giant components on networks we use guaranteed, in order for the network to support the propagation of such a process.

### 2.1.4 Random k-regular graph

A k-regular graph is a graph in which every vertex has degree $k$. Then, a random k-regular graph is a graph $G(N, k)$ selected from the probability space of all the k-regular graphs with $N$ vertices, for which the condition $3 \leq k < N$ with $N \cdot k$ even, is satisfied. The regularity condition affects the properties of this kind of random graphs, and the mathematical results are less tractable compared to the Erdös-Rényi graph. Nevertheless,

in some cases, as in Chapter 4, where we constrain the sum of the edge weights to be the same for each vertex, they are more convenient to use for simulations than Erdős-Rényi graphs.

### 2.1.5 Network motifs

A subset of a graph $G(V, E)$, is called a subgraph, and it is defined as a graph $G' = (V', E')$ all of whose vertices $V'$ are contained in $V$, and edges $E'$ are contained in $E$. If the size of a subgraph is $n < N$, and it is complete, we call it a clique. In real networks, the abundance of some types of connected subgraphs can serve as a tool for studying its properties. Such small, statistically significant, connected subgraphs were named network motifs, and they represent building blocks in network evolution [13]. A few examples are shown in Fig. 2.1. The methodology for the detection of basic structural elements of

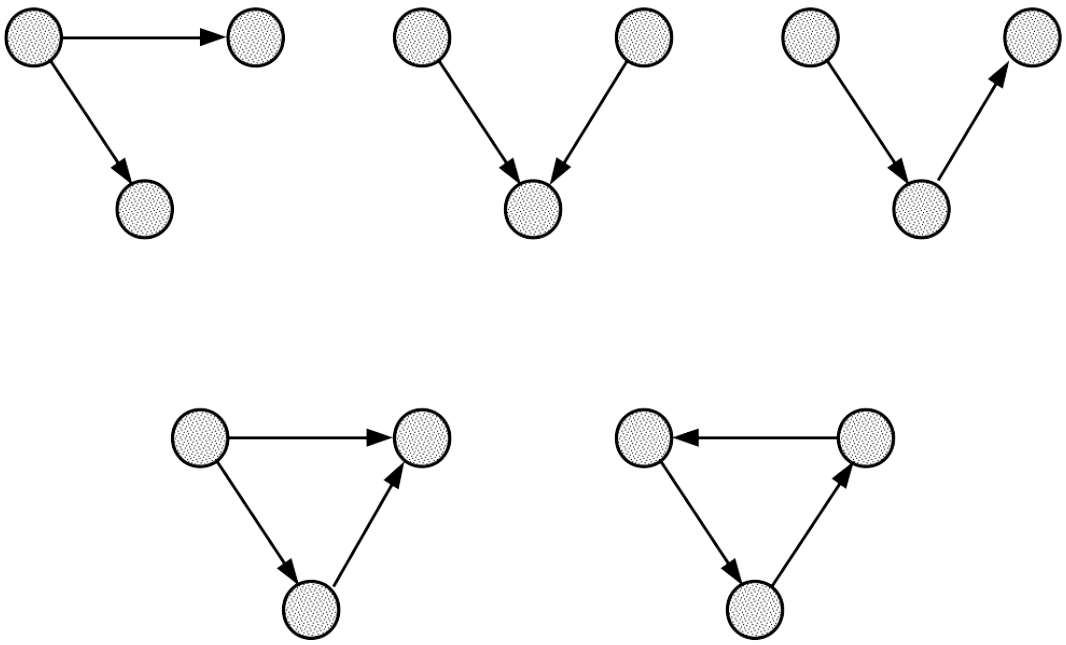


Figure 2.1: Examples of network motifs

networks was first developed by [12]. The authors counted the occurrences of motifs with $n$ vertices on the networks of interest, and then compared it to the count on an

ensemble of randomized networks. The ensemble of randomized networks was created by preserving the degree of each vertex and the appearances of $(n-1)$-vertex motifs, and randomizing all the other properties. The original counts were compared to the counts from the randomized ensemble, and the subgraphs which had a significantly higher count were recognized as structurally important network motifs. The main idea of the authors was to define universality classes of networks using motifs.

Instead of obtaining the distribution of the motif counts by simulating the randomized ensemble, several attempts were made at providing the analytical approximation. The expression for the first two moments (mean and variance) was derived simply from knowing the degree sequence of the network in [81]. This provides a way to calculate z-scores of network motifs without computationally expensive simulations. However, as the p-value characterization is impossible without the entire distribution, Pólya-Aeppli distribution is proposed in [82] as the approximate distribution for the motif count. The important thing to note, both in the calculation of motifs and in choosing the approximate distribution, was the particularity of the network motifs that they tend to overlap. That introduces the covariance term in the calculation of variance, and requires a distribution that accounts for the existence of clumps.

However, if there is a dynamical process occurring on a network, the temporal evolution of the motif appearance is intractable using the existing work. In Chapter 3, we extend Approximate Master Equations [21, 22] to derive the temporal evolution of the first moment of motifs with causal ordering, for a binary state dynamics on a network with a given degree distribution.

## 2.2 Dynamical processes on networks

The theory of dynamical processes on networks is a framework that bridges the emergence of macroscopic properties with the microscopic interactions between the network's elements. The approach is analogous to the concept of statistical physics, from which it borrows some methods. The dynamical description of the system attributes a variable $\sigma_i$ to each vertex $i$ that corresponds to its dynamical state. In an epidemic process it can refer to the individual being susceptible, infected, or recovered; in a contagion on a financial

network it can describe an institution as defaulted or non-defaulted. The set of all possible states of a vertex can be defined as a state space $S = \{1, 2, \ldots, \kappa\}$, $\sigma_i \in S$. If the states of all the vertices in a network of size $N$ are known at time $t$, $\sigma(t) = (\sigma_1(t), \ldots, \sigma_N(t))$, we say that the microscopic state of the system is known.

The dynamical process is contained in the transitions of the system from one microscopic state to another, $\sigma^a \rightarrow \sigma^b$. We describe the process evolution in the phase space of the variable $\sigma(t)$, consisting of all the possible configurations. However, due to the typically large size of complex systems, the evolution of all the microscopic variables $\sigma_i$ is intractable. We present an approach that focuses on a probabilistic description of the system trajectory, the master equation, in the Subsection 2.2.3 [13]. In the following subsections, we present the microscopical definition of all the dynamical processes that will be used in this thesis.

### 2.2.1 Interacting particle systems

Dynamical processes on networks that are stochastic in nature can be tracked back to a class of stochastic processes called interacting particle systems [14, 15, 16, 17]. An interacting particle system is defined as a continuous-time Markov process $(X(t))_{t\geq 0}$ on some configuration space $\Omega$. The configuration space is a Cartesian product space $\Omega = S^G$, where $G$ is a countable set, called a lattice, that represents the site space, and $S$ is a local state space. Therefore, the configuration space contains $G$ copies of $S$, and the Markov process defined on its elements represents the microstate of a dynamical process:

$$\sigma(t) \rightarrow X(t) = (X_i(t))_{i\in G}, \quad X_i(t) \in S \quad \forall \quad i \in G. \tag{2.13}$$

The continuous-time Markov process is a stochastic process generated by the conditional probabilities that satisfy the Chapman-Kolmogorov equation [83]. The best known examples of interacting particle systems are the contact process, voter model, stochastic Ising model and stochastic Potts model. We will familiarize ourselves with the contact process and the voter model, since they are used in Chapter 3 to simulate synthetic default propagation processes which we use to test our method.

### The contact process

The contact process [16] is a stochastic process widely used in biology to model bacterial growth, and in epidemiology to track infectious disease spreading. The local state space of the contact process is $S = \{0, 1\}$. Markov chain states are thus $X_i(t) = 1$ if the site $i$ is infected or occupied, and $X_i(t) = 0$ if it is recovered, or empty. The transition rates on the state space $S$, $r_{01}$ and $r_{10}$, at a site $i$ are given as:

$$0 \to 1, \quad r_{01}(i) = \lambda \sum_{j:j\sim i} X_j(t), \tag{2.14}$$

$$1 \to 0, \quad r_{10}(i) = \delta, \tag{2.15}$$

where the relation $j \sim i$ denotes that $i$ and $j$ are neighbours. A quantity of interest for this process is the number of the infected sites $N(t)$. For an infinite graph $G$ it was shown that, for a fixed $\delta$, there exists a critical value $\lambda_c$, so that for $\lambda > \lambda_c$ the process has a positive value of $N(t)$ for all times, and for $\lambda < \lambda_c$ it gets extinct almost certainly. Such an abrupt change in behaviour with the change in the parameter $\lambda$ is recognized as a second order phase transition. We will use the special case of the contact process, with $\delta = 0$, to model the default process. The increase in the probability of default is linear with the infected neighbours, which makes it a suitable approximation for the propagation of default.

### The voter model

The voter model [18, 17, 16] was built to model the dynamics of opinion spreading, but it is often used as well for modeling a spatial conflict, e.g. biological populations or nations in conflict. The vertex $i$ copies the state of a vertex $j$ with probability $p(i, j)$. The local state space can be in general $S = \{0, \dots n\}$, but we describe the case when $n = 1$. The rate of change between the two states is symmetrical $r_{01} = r_{10} = r$, and for the site $i$ it follows:

$$r(i) = \sum_{j:X_i(t)\neq X_j(t)} p(i, j), \quad \sum_j p(i, j) = 1, \tag{2.16}$$

where $p(i,j)$ are transition probabilities, such that the random walk with those transition probabilities is irreducible. This definition points to the duality between the voter model and coalescing random walks [84]. A coalescing random walk considers a set of particles making independent random walks on a graph, with the property that, when two particles meet on a vertex, they coalesce and proceed with the walk as a new particle . If we consider a simple random walk, that can take one step at a time, on a lattice, $S = \mathbb{Z}^d$, the transition probability is [17]:

$$p(i,j) = \begin{cases} 1/2d, & \text{if } i \sim j \text{ and } X_i(t) \neq X_j(t) \\ 0, & \text{otherwise,} \end{cases} \tag{2.17}$$

where $2d$ is the number of the sites neighbouring the vertex $i$. In this case we can say that the site $i$ flips with the rate proportional to the number of neighbours with a different opinion. As the system evolves, patches of sites with same opinions appear, and the system ends up fixating one of the opinions. Such clustering behaviour appears only if the lattice has the dimension less than 3. In our research, simulations are performed on networks, and not on lattices. When applied to model a default process, the voter model has the default rate proportional to the fraction of the defaulted neighbours, instead of their absolute number as in the contact process, which makes the process more similar to actual defaults in the economy.

### 2.2.2 The threshold model

In an attempt to model large, rare cascades in social and economic systems, Watts [19] introduced a simple threshold rule on a random network of interacting agents. Such a model belongs to the wider class of game-theoretic economic models, known as binary decisions with externalities [85], used to study how individuals make choices based on the decisions of others. Unlike the models from interacting particle systems, the threshold model is defined deterministically. The state space is again $S = \{0,1\}$, but instead of a transition rate, there is a transition rule:

$$0 \to 1, \quad \text{if} \quad \frac{1}{k} \sum_{j:j\sim i} \sigma_i \geq \phi. \tag{2.18}$$

A site $i$ in state $\sigma_i = 0$, that has $k$ neighbours, will switch to state 1, if the fraction of its neighbours in state 1 is larger than some threshold $\phi$. The model introduces local dependencies, fractional thresholds, and the heterogeneity of the thresholds and the network. These properties are shown to be essential in recovering the two qualitative features of global cascades; rare occurrence and large size.

### 2.2.3 The master equation and approximate solutions

We have now defined all the three dynamical process models that we will use in this thesis on a microscopic level. As it was noted at the beginning of this section, tracking the trajectory of all the microscopic variables $\sigma_i$ is impossible in large systems. Therefore, we will approach the problem from a different perspective; instead of looking for a mechanistic picture the system, we will focus on a probabilistic description, and resort to tracking the evolution of the probability distribution of the system configuration, $P(\sigma, t)$. As opposed to the Fokker-Planck equation, that describes systems whose state changes continuously, the master equation governs the evolution of systems with discontinuous states, which makes it a natural choice for the representation of dynamical processes on networks. For processes that are continuous in time, it reads [13]:

$$\frac{\partial P(\sigma, t)}{\partial t} = \sum_{\sigma'} \left[ P(\sigma', t) W(\sigma' \to \sigma) - P(\sigma, t) W(\sigma \to \sigma') \right]. \tag{2.19}$$

The sum goes over all the possible microscopic states $\sigma'$, and the terms $W(\sigma' \to \sigma)$ are the transition rates between the microscopic states $\sigma'$ and $\sigma$. The master equation is written in the form that stresses the existence of the terms that represent the inflow of the probability that any state $\sigma'$ changes to $\sigma$, and the outflow obtained by the rate at which the $\sigma$ microstate changes to any other possible state [20]. Even though the transition rates can generally depend on the entire microscopic configuration $\sigma$, taking the network structure into account simplifies them significantly. After noticing that the change of the vertex $i$ state can be influenced only by local interactions with neighbouring vertices, $j \in V(i)$ of $i$, the transition rates $W(\sigma' \to \sigma)$ can be decomposed as:

$$W(\sigma' \to \sigma) = \prod_i w({\sigma'}_i \to \sigma_i | \sigma_j). \tag{2.20}$$

It is through this equation that the network topology enters the dynamical process description.

A solution of the master equation, $P(\sigma, t)$ would, in principle, allow us to calculate the expected value of any quantity of interest in the system. For some quantity $A(\sigma)$, that depends on the microstate $\sigma$, the temporal evolution of the expected value is simply:

$$\langle A(t) \rangle = \sum_{\sigma} A(\sigma) P(\sigma, t). \tag{2.21}$$

However, in most cases it is impossible to obtain a complete solution for the master equation, even for simple dynamical processes. Approximations schemes that must be employed, such as mean-field theory or approximate master equations, are next described.

### Mean-field approximation

The evolution of the number of vertices in a specific state of the process is an intuitively important quantity of interest. The average number of vertices in the state $x \in S$, as a function of time, can be obtained from a projection of the master equation:

$$N_x(t) = \sum_{\sigma} \sum_{i} \delta_{\sigma_i, x} P(\sigma, t). \tag{2.22}$$

In order to solve this equation, a mean field approximation is used on the probability $P(\sigma, t)$. The assumptions that it introduces are the homogeneity of the system, and the lack of all correlations between the microstate variables $\sigma_i$. Therefore, the probability that a system is in some microstate $\sigma$, can be simplified to:

$$P(\sigma, t) = \prod_{i} p_{\sigma_i}. \tag{2.23}$$

The probability that vertex $i$ is in some state $\sigma_i$ is now independent of that vertex, as in a homogeneous system there is no difference between the vertices. Since all the correlations between vertices are neglected, $P(\sigma, t)$ is factored as a product of independent probabilities for each vertex.

A general form of the equations obtained after using the mean field approximation looks

like:

$$\frac{\partial N_x}{\partial t} = F_x(N_1, \dots N_\kappa), \tag{2.24}$$

where $N_1, \dots N_\kappa$ are average numbers of vertices in all the possible states $x \in S$ [13].

We will present a more specific mean-field equation, for a dynamical process with $\kappa = 2$, which is called binary dynamics, and has a wide application in modeling fundamental processes. The state space of the binary dynamics is $S = \{0, 1\}$. The default process, which we focus on in this thesis, is an example of binary dynamics. We denote with $k$ the degree of a vertex, and with $m$, the number of neighbours of that vertex that are in state 1. The mean degree of the network is denoted with $\langle k \rangle$. The rates of transition between the states are generally denoted as $w(0 \to 1|k, m) = F_{k,m}$ and $w(1 \to 0|k, m) = R_{k,m}$. The rate $F_{k,m}$ is the rate of the transition from the state 0 to state 1, for a vertex with degree $k$ and $m$ neighbours in state 1. The rate $R_{k,m}$ denotes the rate of the opposite transition, $1 \to 0$. A vertex $i$, with $m$ out of $k$ neighbours in state 1, will be in state 1 or state 0 with the respective probabilities:

$$P(\sigma_i = 1; k, m) = \rho_k B_{k,m}\left(\frac{k}{\langle k \rangle}\rho_k\right), \quad P(\sigma_i = 0; k, m) = (1 - \rho_k) B_{k,m}\left(\frac{k}{\langle k \rangle}\rho_k\right),$$
$$B_{k,m}(q) = \binom{k}{m} q^m (1 - q)^{k-m}. \tag{2.25}$$

The differential equation describing the density of the vertices with degree $k$ in state 1, $\rho_k(t)$ then follows [21, 22]:

$$\frac{d}{dt}\rho_k = \sum_{m=0}^{k} \left[ F_{k,m}(1 - \rho_k) B_{k,m}\left(\frac{k}{\langle k \rangle}\rho_k\right) - R_{k,m}\rho_k B_{k,m}\left(\frac{k}{\langle k \rangle}\rho_k\right) \right]. \tag{2.26}$$

The initial value of the density is $\rho_k(0) = \rho(0)$. The fraction of the vertices in state 1 in the whole network is easily obtained if we know the degree distribution:

$$\rho(t) = \sum_k P_k \rho_k(t). \tag{2.27}$$

If we wish to use the mean field equations for the microscopic dynamical processes defined above (the contact process, also known as SI model, when it is modeled on a network, and with $\delta = 0$, the voter model and the threshold model) we only need to state their

transition rates in terms of $k$ and $m$ (Table 2.1).

Table 2.1: Transition rates for the SI, voter model and threshold model

| Process | $F_{k,m}$ | $R_{k,m}$ |
|---|---|---|
| SI | $\lambda m$ | 0 |
| Voter model | $\frac{m}{k}$ | $\frac{k-m}{k}$ |
| Threshold | $\begin{cases} 0, & \text{if } m < M_k \\ 1, & \text{if } m \geq M_k \end{cases}$ | 0 |

**Approximate master equations**

Although Eq. (2.26) is simple to derive, the assumptions used in its derivation are not always valid, nor is the equation very accurate, especially on sparse networks, or near the critical points of the dynamics. One way of improving the accuracy of the description would be to take subsets of vertices larger than 1 (pairs, triplets) as the model units [86]. However, increasing the number of vertices in the unit causes the number of equations needed to describe its evolution to explode. Another type of approach is to focus on one vertex, but to keep track of all of its first neighbours. Such models are named approximate master equations [87, 21, 22].

We embrace the notation from the SI epidemic model, and call the vertex in state 0 susceptible and the vertex in state 1 infected. At time $t$, the sets of susceptible or infected vertices, that have the degree $k$, and the number of infected neighbours $m$, are denoted with $S_{k,m}(t)$ and $I_{k,m}(t)$. For the calculations, it is necessary to define the densities, $s_{k,m}(t)$ and $i_{k,m}(t)$ of the susceptible and infected vertices with degree $k$ and $m$ infected neighbours. The density of all the infected vertices with degree $k$ is then the sum of densities over all possible $m$, $\rho_k(t) = \sum_{m=0}^{k} i_{k,m}(t) = 1 - \sum_{m=0}^{k} s_{k,m}(t)$. The total density $\rho(t)$ can be obtained from Eq. (2.27).

The temporal evolution of $s_{k,m}(t)$ can be written as:

$$\begin{aligned} s_{k,m}(t+dt) = s_{k,m}(t) &- W(S_{k,m} \to I_{k,m}) s_{k,m} dt + W(I_{k,m} \to S_{k,m}) i_{k,m} dt \\ &- W(S_{k,m} \to S_{k,m+1}) s_{k,m} dt + W(S_{k,m-1} \to S_{k,m}) s_{k,m-1} dt \\ &- W(S_{k,m} \to S_{k,m-1}) s_{k,m} dt + W(S_{k,m+1} \to S_{k,m}) s_{k,m+1} dt, \end{aligned} \tag{2.28}$$

where only transitions whose rate is linear in $dt$ are taken into account. The transition rates between the $S_{k,m}$ and $I_{k,m}$ classes are already defined:

$$W(S_{k,m} \to I_{k,m}) = F_{k,m}, \quad W(I_{k,m} \to S_{k,m}) = R_{k,m}. \tag{2.29}$$

The remaining rates concern the transitions of the neighbours, which can in turn be caused by their own neighbours. We can factor them as the number of neighbours in the state of interest that multiplies the probability of the their transitions:

$$W(S_{k,m} \to S_{k,m+1}) = (k-m)\beta^s, \quad W(S_{k,m} \to S_{k,m-1}) = m\gamma^s. \tag{2.30}$$

Since we are not tracking the state of their neighbours, we introduce the mean-field approximation at this level. We look for the probability an SI pair will arise from an SS pair, and vice versa. We assume that there are no correlations present between pairs of vertices and average over the entire network:

$$\beta^s = \frac{\sum_k P_k \sum_{m=0}^{k} (k-m) F_{k,m} s_{k,m}}{\sum_k P_k \sum_{m=0}^{k} (k-m) s_{k,m}}, \quad \gamma^s = \frac{\sum_k P_k \sum_{m=0}^{k} (k-m) R_{k,m} i_{k,m}}{\sum_k P_k \sum_{m=0}^{k} (k-m) i_{k,m}}. \tag{2.31}$$

After inserting the obtained expressions back into Equation (2.28) and taking the limit $dt \to 0$, we get the approximate master equation for the evolution of $s_{k,m}(t)$:

$$\begin{aligned} \frac{d}{dt} s_{k,m} = &- F_{k,m} s_{k,m} + R_{k,m} i_{k,m} - \beta^s (k-m) s_{k,m} + \beta^s (k-m+1) s_{k,m-1} \\ &- \gamma^s m s_{k,m} + \gamma^s (m+1) s_{k,m+1}. \end{aligned} \tag{2.32}$$

Analogously, the equation for the evolution of $i_{k,m}(t)$ follows:

$$\begin{aligned} \frac{d}{dt} i_{k,m} = &- R_{k,m} i_{k,m} + F_{k,m} s_{k,m} - \beta^i (k-m) i_{k,m} + \beta^i (k-m+1) s_{k,m-1} \\ &- \gamma^i m i_{k,m} + \gamma^i (m+1) i_{k,m+1}, \end{aligned} \tag{2.33}$$

with the rates $\beta^i$ and $\gamma^i$ defined as:

$$\beta^i = \frac{\sum_k P_k \sum_{m=0}^{k} m F_{k,m} s_{k,m}}{\sum_k P_k \sum_{m=0}^{k} m s_{k,m}}, \quad \gamma^i = \frac{\sum_k P_k \sum_{m=0}^{k} m R_{k,m} i_{k,m}}{\sum_k P_k \sum_{m=0}^{k} m i_{k,m}}. \tag{2.34}$$

The initial fraction of infected vertices, $\rho(0)$ is turned into initial conditions using the binomial factor, $s_{k,m}(0) = [1 - \rho(0)]B[\rho(0)]$ and $i_{k,m}(0) = \rho(0)B[\rho(0)]$.

We will extend this framework for the class of directed networks with temporal ordering, that we use in Chapter 3.

### 2.2.4 Detection of contagion processes on networks

Dynamical processes on networks can have two types of driving: either the vertices' states change due to some external influence over them, or some propagation mechanism causes the state changes to spread through the network. An important line of research in complex systems builds on the relationship between these two possible natures of a dynamical process. Distinguishing between the social and external influences is key in studying information spreading, epidemics, social system or economic defaults.

In [23] the authors investigate the peaks in the time series of book sales, and classify them according to their shapes into two types, endogenous and exogenous. The spreading of the influence between buyers is modeled on a network, using an epidemic model with long-range memory. The functional form of the solution the expected number of sales turns out to be different depending on whether the system received an exogenous shock, or it was simply in a critical state which allowed small fluctuations to cause endogenous cascades. The model shows a good fit to the peaks observed in the empirical data.

The diffusion of information on social networks is modeled in [24, 25, 26]. Authors of [24] use a hazard function for the internal propagation of information, and an event profile function to describe the external source of information. They derive a distribution of the number of exposures a vertex receives over time, and a an exposure curve, that describes the probability a vertex is infected as a function of the number of exposures. They validate it on synthetic data, and then infer the parameters for the real data, showing that external information reaches the Twitter users more often than expected. In [25], the authors propose a model for the probability that a vertex was influenced through peer influence propagation. They distinguish the externally influenced vertices as the ones that were influenced, although their probability determined by the model was lower than the average of the entire network. After optimizing the parameters of the function, they give

an estimate of the external influence in the information propagation on Facebook. Another functional form for the propagation of information is given in [26], and the existence of social influence is determined by estimating the coefficient governing the diffusion and comparing it to coefficients for randomized data.

The present research focuses on using parametric models to determine the existence of endogenous and endogenous components. The exogenous component is either parametrized, or set as a null hypothesis, while the endogenous component is always represented parametrically. In the Chapter 3 we propose a methodology that requires no parameters for the process, and does not assume the existence of the endogenous process a priori.

## 2.3 Systemic risk

Systemic risk is the risk that a large part of the system, consisting of interacting agents, will collapse. The interconnectedness of the system lies at the core of the systemic risk, as a cascading event, which leads the entire system into failure, can be triggered by a single individual failure. Systemic risk thus emerges as a macroscopic property of a system, that can be technical, social, or financial. Considering this, it is clear that complex networks arise as a natural framework to study systemic risk [10].

The authors in [88] provide a general framework that encompasses contagion processes with both deterministic and stochastic failure dynamics and homogeneous and heterogeneous thresholds. Although there is no formal definition for the systemic risk, the macroscopic variable they use as a proxy is the fraction of defaulted vertices [88]:

$$X(t) = \frac{1}{N} \sum_i I_i(t). \tag{2.35}$$

They obtain a mean-field limit of this fraction of failed vertices for all the processes in their framework, and observe how the type of microscopic dynamics affects the behaviour of the total fraction of failures.

In financial systems, systematically important companies, whose default can have a detrimental effect on the entire system, are deemed "too big to fail". The problem of determining how systemically important a company in the financial system is, does not have

an universal solution. A methodology that takes a network perspective on the problem was introduced in [27]. The authors employ a metric similar to a centrality measure from network science to quantify the importance of the vertices in a network. They call it DebtRank, after the PageRank centrality metric [72] that measures the importance of a vertex by the number of important neighbours it has, while normalizing the influence of each neighbour with its own out-degree. However, in DebtRank, the walks in which edges are repeated are excluded, to avoid counting the impact of a vertex more than once. Quantitatively, it represents the total loss incurred to the system from a certain individual defaulted institution, or a small group of institutions. They stress the importance of the network in the amplification of a small shock, and suggest broadening the concept of "too big to fail" into "too central to fail".

Beside researchers quantifying the systemic importance of institutions, a large body of research is concerned with mitigating the systemic risk. There are several perspectives on systemic risk reduction: acknowledging the ambiguous impact of diversification [28, 29] and increasing the information available to the market participants [35], introducing Pigouvian taxation [30, 31] and assessing optimal network topologies and levels of assortativity [32, 33, 34].

The belief that diversification always leads to a reduction of systemic risk, formed before the Great Financial Crisis is challenged in [28, 29, 35]. Two scenarios are observed, one in which the complete information about the system is known to all the participants, thus making the losses conservative, and the other, in which there exist levels of uncertainty about the market participants, causing loss amplification [28]. For example, after an agent receives an external shock, or some of its counterparties default, its fragility increases. Consequentially, its lenders are left with an uncertainty about the agent's state, which can cause them either to claim their funds back and make the agent to go into fire sales, increasing its fragility even further, or increase their external risk premium, which in turn also causes the agent's fragility to increase [35]. It is shown that even for conservative losses diversification can increase systemic risk, if the financial fragility of the institutions is heterogeneous and low on average [28]. On the other hand, in the scenario with loss amplification, with an increase in connectivity there is an interplay between the benefits of diversifying the losses, and increasing the amplification of shocks. For low levels of

connectivity risk diversification is beneficial, while for connectivity that is already high, increasing it even further increases the systemic risk [29].

Apart from the propagation of actual losses, an important part in the recent crisis was played by the propagation of distress in the shape of mark-to-market reevaluations of interbank claims, occurring due to an increased likelihood of default of the counterparties [35]. This type of reevaluation of assets, called Credit Valuation Adjustment [36], is usually calculated by taking only the creditors of the institution into account. In Chapter 4 we show the importance of considering the entire financial network when calculating CVA, and elaborate for different scenarios, with and without correlation on the external assets.

Pigouvian tax is a tax imposed on institutions or individuals that engage in activities that create adverse effects on the society. In the context of systemic risk, it could be a type of tax that would disincentivize actions that increase the systemic risk [31], or charge institutions proportionally to their systemic importance [30]. The first type was proposed in [31], as a tax on transactions that is proportional to their marginal increase of the systemic risk. The authors use an agent-base model to show that the proposed tax gives rise to network restructuring and reduces the systemic risk without reducing the transaction volume. Another approach is given in [30], taxing the institutions in an amount proportional to their equity, and then applying a strategy for utilizing the rescue fund. To avoid a moral hazard, rescue funds are given to the lenders of a defaulted institution, to stop a potential cascading effect. The authors show that the cascade risk, a proxy they use for systemic risk, is greatly reduced with their taxation scheme.

The question of the optimal network architecture is tackled by [32]. They show that there is no optimal network topology in general, however, when the market is illiquid (due to fire-sales causing additional losses to agents) scale-free networks show different stability behaviour from Erdős-Rényi and regular graphs. More particularly, their stability behaviour is not universal, compared to these two other topologies. Whether it will be more stabile or more fragile, depends on the initial endowments of institutions and the correlation between their debtors and creditors. The authors of [33] study the effects of the network assortativity in scale-free networs on the systemic risk. They show that disassortative networks reduce the systemic risk. However, although assortativity increases

the systemic risk, it slows down the contagion process, giving the regulators more time to react [34].

# Chapter 3

# Inference of contagion processes using causal motifs

## 3.1 Introduction

The mathematical description of dynamical processes in complex systems most often belongs either to the class of stochastic cellular automata, or their continuous time analogue, interacting particle systems. As in this Chapter we focus on the latter, we shortly introduced them in Subsection 2.2. Dynamical processes in systems that are modeled with this type of stochastic models are epidemics [89, 90, 86], default propagation in financial industry [27, 91], propagation of information in society [92], queueing [93], etc. We discussed in Chapter 2 how macroscopic effects often occur from the microscopic mechanics of interactions in complex systems. Phenomena such as cascades of default or wide-spread disease contagion can be detrimental to the system. Detecting the cascade of events in time can be of utmost importance for the functioning of the underlying system, and it can provide us with some new insights in the modeling of their dynamics.

However, when creating a model for a specific dynamical process from temporal data, it is not enough to assume from the nature of the system that it is convenient to use one of the aforementioned processes. Perhaps it could be that there is not enough contact data available to successfully reconstruct the network on which the contagion occurred [94, 95, 96], or the multiplex network representation of the systems is lacking the very layers that

conducted the spreading process [97, 98, 99]. Finally, it is also possible that the network did not contribute to the events in the process at all. The dynamics in all of these cases would be then best described as a field-effect.

This leads us naturally to a question that we are addressing in this Chapter, that is, for some temporal data that we have, whether interacting particle systems should be employed for the description of the dynamics, or should we simply resort to field-effect modeling. Attempts made so far to answer this question are mentioned in the Section 2.2. However, the distinguishing factor of our research is that we do not make any assumption about the nature of the contagion, that is, our method is non-parametric.

In order to formalize our question, we proceed with some definitions. With respect to the available data, an endogenous event is defined as a change of state of an agent that was caused by a change of state of some other agent through its directed relation (represented as a directed edge in a network) to the first agent. We loosely define a cascade as a series of connected endogenous events.

On the other hand, an exogenous event is defined to be a change of state of an agent that happened with no connection to the change of state of any agents that share a directed relation to the first agent.

Although we cannot infer causal relations from the data, we can use event times from the data to distinguish two types of binary relationships between events; those that are for certain not related by causality, as the time $t_1$ on the beginning point of the edge is larger than time $t_2$ on the ending point $t_1 > t_2$, and those whose temporal ordering does not exclude a causal relationship, i.e. $t_1 < t_2$. Since we cannot confirm causality for each relation individually, we check whether a global effect can be observed statistically in the entire system. We define causal motifs containing only potentially causal edges below in Table 3.1 and compare their frequency of occurrence in the process to their null-distribution, which we obtain from an ensemble of networks with randomized event times. Then, if there were events that were caused endogenously, the frequency of causal motifs is expected to differ from its null distribution, which conserves no temporal ordering. On the other hand, if the process is completely exogenous, there should be no statistically significant difference.

Leaning on this rationale, we use the statistically significant difference between the original motif count and its null distribution as an indicator of the presence of an endogenous component in the dynamical process recorded in the data.

This work was motivated by a phenomenon that appears in economical systems - default cascades [28, 100]. It is manifested as a series of mutually caused defaults and its appearance threatens the stability of the system. In poorly regulated financial systems cascades of default are common, and they are believed to largely increase systemic risk even in well regulated systems [28, 100, 30]. Therefore, it is crucial for regulators to know what is causing the companies to go into default. They have to distinguish between exogenous causes such as loss of market access, loss of access to finance, monetary shocks, etc. and endogenous propagation, i.e. when different agents fail to fulfill their obligations to their creditors, thereby increasing the risk for a larger systemic event.

In modeling financial contagion, the network is usually created using balance sheets of nodes [91], which consists of different categories that insulate the node from default. This approach is equivalent to the threshold model presented in Section 2.2, as demonstrated in [101]. The interplay between the network architecture and financial contagion between banks is studied in [102]. Credit chains are mostly used to study contagion, being firstly outlined in [103] as a principal way to propagate financial distress. In addition, defaults on trade credit have also been recognized as an important cause of distress, as shown on US data [104] and Swedish data [105]. An attempt to meaningfully model links between firms in a production network has been taken by the authors in [106, 107, 108]. A business network has been constructed in [109] from supply chain data. Two types of bankruptcies were distinguished in the contagion within production network in [110]; one is a random shock to revenues or costs and the other a creditor not being paid by the debtor. However, no method exists still that identifies these types of bankruptcies from the data. As already mentioned, the need to be able to differ between the shocks that affect the node defaults externally, and the shocks transmitted within the network was the motivation behind this Chapter.

This work is the continuation of the effort taken by [111] to develop a methodology for detection of endogenous propagation in financial networks to be used on Croatian company defaults. There the authors first constructed a directed temporal network with

the companies and their times of default on vertices, and their mutual debts on edges, pointed from debtor to creditor. They used single causal edges (with the time on the tail preceding the time on the head of the edge) and tested their frequency in the original network against their frequency in a randomized reference model (RRM) of the same network [37]. They refined the method by including the values of the debt as edge weights. We extend this research by making the choice of the test statistic more general, as we include *causal motifs* with two and three edges and the largest component of causal edges, in addition to single causal edges. Unlike the authors, we omit any other data than the data that is necessary to build a directed temporal network, and provide some theoretical insights into the methodology. Moreover, this methodology can have a more general application on any spreading phenomena with two states on a directed network.

The use of the RRM in the literature goes from temporal embedding [112], inference of structures in communication networks [113] to analysing collective behaviour in social networks [114]. In the last-mentioned work the RRM was used to benchmark the count of network motifs as well, however the two-event motifs they employed had no causal ordering.

We introduced the network motifs and their application in the inference in complex networks in Section 2.1. We can additionally mention their use in understanding metabolic and other biological networks [12, 115, 116], the properties of ecological system through food webs [117], in economical setting to understand corporate governance [118, 119], and organization of knowledge in Wikipedia [120]. Their extension to temporal networks has been made by [121].

Temporally ordered motifs [122] were introduced for food webs, but unlike our definition of causal motifs, they allow edges to point opposite of the temporal ordering on the vertices. Process motifs were defined in [123] as small subgraphs composed of walks on them, that represent building blocks of dynamical process. Authors in [124] use a similar concept of temporal motifs to ours to detect anomalies in time series on networks.

Before using our methodology on the real data of company defaults, we validate it on artificial data where we control the exogenous and endogenous process components. From the vast space of all possible processes and networks, we choose two specific types of processes, that we consider to best represent financial contagion, and simulate them on

Erdős-Rényi networks. For computational reasons we use relatively small networks. Since we expect the results on such networks to be the least conclusive, they provide a lower bound for the result significance. We test the significance of our results and the robustness of test statistics using Kolmogorov-Smirnov test, z-score and Mahalanobis distance. All the scripts used for simulations and the company default data are available in our Git-Hub repository [125, 126].

## 3.2 Data

The data was gathered from the website of the Croatian Financial Agency. There, all the documents connected to "Zakon o financijskom poslovanju i predstečajnoj nagodbi", that involved debt renegotiation and restructuring in the Republic of Croatia, are published. The criteria for filing for this type of bankruptcy are defined in a law that was passed at the end of 2012. That is, a company, that has failed to attain liquidity over a period of 60 days, can file to a pre-bankruptcy settlements if at most 21 days passed since its insolvency onset. On the day it files for a pre-bankruptcy settlements, it is obliged by law to deliver a list of its creditors. For brevity, from now on we will call pre-bankruptcy settlements - defaults, which for all practical purposes of this research, they are. Due to the format of the published data, extensive data mining and cleaning had to be done in order to convert it into a format suitable for network formation.

The final form of the data is a table consisting of a list of debtors, a list of all of their creditors, the amount of debt per creditor, the starting time and the duration of the pre-bankruptcy settlement. A temporal network is created with all the creditors that are also debtors, represented as vertices. The directed edges represent debts, pointing from the debtors to the creditors, in the direction of the possible default propagation. The debtors are attributed with the starting times of the pre-bankruptcy settlements, which are regarded as times of default. We do not include the duration of the pre-bankruptcy settlements or the values of debt, as we devise our methodology on the minimal amount of information possible. In Fig. 3.1 we depict the scheme of the network creation. The data consists of 25469 creditors and debtors and 52507 debts, with creditors being banks, private and public firms, government and individuals and debtors being exclusively firms.

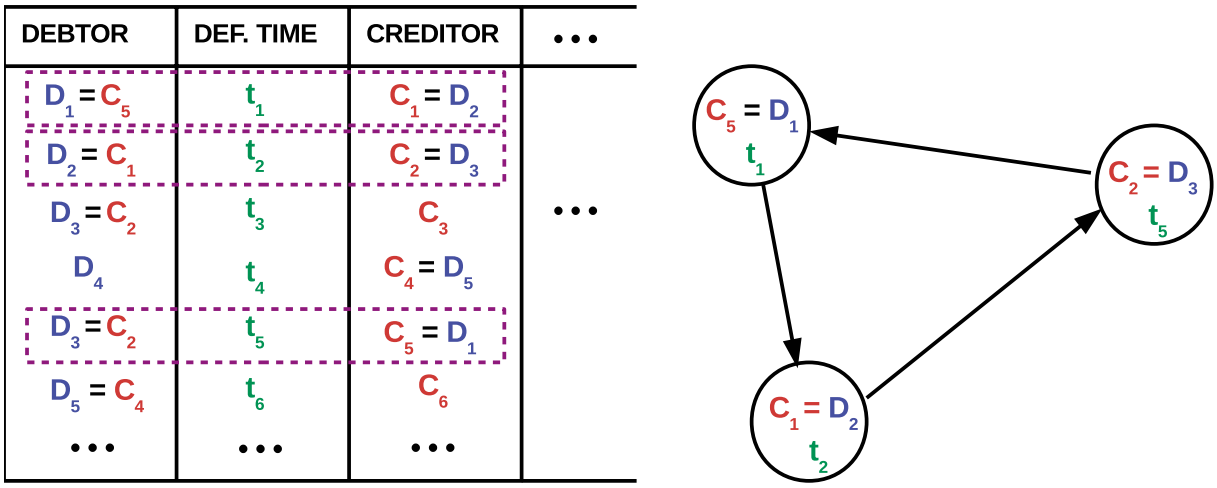



Figure 3.1: **The scheme of the network construction from the data.** The simplified table on the left side of the figure shows the only three columns we employ from the original data; the identities of debtors, the start of the pre-bankruptcy settlement (default time) and the identities of creditors. We use only the firms that are both debtors and creditors, and rows that satisfy that requirement are marked with a purple rectangle in the table. On the right side, the network representation of the selected rows is depicted. The relationship of debt is represented with a directed edge, pointing from the vertex that entered the settlement as a debtor, to the vertex that entered the settlement as a creditor, The time of default is assigned to the vertex that represents the debtor. The direction of the edge is opposite to the flow of the money, from debtor to creditor, as it is chosen to correspond to the direction of the possible default propagation.

The debts amount in total to 5.97 billion euros, corresponding to 13.6 percent of Croatia's GDP in 2014. The dataset encompasses the period between December 19th 2012 and February 26th 2014. It starts right after the law for pre-bankruptcy settlements was introduced, since the largest and most interconnected debtors filed for the procedure right away, and it ends as the number and size of firms diminishes in the following two years. From the economical perspective, in this time frame cascades are expected to be observed, which is the reason we choose it to apply our methodology to.

## 3.3 Methodology

### Rationale

The main contribution of this Chapter is a methodology devised to infer the existence of a contagion component in a dynamical process on a network. Even though our range of application extends to any process with two state unidirectional dynamics, considering

that our data is from the economy, we immerse entire setting in the economy-related terminology. We call the initial state of the vertex *non-defaulted* and the changed state *defaulted*.

The data that we employ needs to consist of institutions that are both debtors and creditors over the time frame, in order for us to build a connected network on which propagation can occur. In such a setting two possible scenarios can come about. The first possibility is that the default of the vertex is exogenous, following the definition from Section 3.1 . In practice this means that all the debts between this vertex and its already defaulted debtors were either too small to cause significant distress to the vertex, or some other institution intervened and repaid the debtors obligations. The second possibility is that the vertex defaults from an endogenous cause, in the manner generally defined in Section 3.1. Translated to the economy, it means that a default of the debtor at $t_1$ caused the default of its creditor at $t_2$. It is clear that in this case the relation between the two times will be $t_1 < t_2$, unlike the first scenario, when the relation is not determined.

With our methodology, we do not need to go into the mechanisms behind the occurrence of the defaults, as they will not affect this signature in the timestamps. Therefore, there is no need to assume a functional form of the mechanics of contagion on the network. Our main observation is that the pattern of timestamps produced by the process will be very different depending on whether that process is purely exogenous, or it has an endogenous component as well. More precisely, as we stated that the exogenous scenario does not determine the relation between the creditor's and debtor's default times, a permutation of all the timestamps in the data should result in a realization of the process that is equally likely to occur as the original one.

## Counting causal motifs

Using this reasoning, we count the causal motifs, already mentioned in Section 3.1 and check if their frequency corresponds to a frequency expected from a purely exogenous process. To reiterate again, a *causal edge* is an edge with default times $t_1$ and $t_2$ at the tail and the head vertices respectively, that obey a temporal ordering $t_1 < t_2$. A *causal motif* is any motif consisting exclusively of causal edges. Unlike the motifs from the

previous works that we referred to in Section 3.1 that classify the motifs by the number of their vertices, we order them by the number of edges they contain. We restrict our investigation to the order of three, and depict all the motifs in the Table 3.1. The larger orders of motifs are skipped, since finding them in a network is computationally expensive, and we do not expect a change in the trend with further increase. However, in addition to causal motifs, we also include the *largest causal component* in our set of test statistics. We define it as a largest weakly connected component [72] of a network from which all the edges except causal ones are removed. The largest causal component is computationally easy to find, and it can be thought of as a largest causal motif that can be found in given instance of the network.

We assume that the frequency of the causal motifs and the size of the largest causal component will be larger for a process that is composed of both endogenous and exogenous components than for a purely exogenous process. Moreover, we expect the test statistics from a purely exogenous process not to significantly differ from the statistics calculated when the temporal ordering in the process is destroyed. Therefore, our null-hypothesis suggests that the causal motif count and the size of the largest component will correspond to the values obtained in a process that has only exogenous influence.

To obtain a null-distribution of the motif counts and the largest component sizes, we use a randomized reference model [37]. A RRM is a microcanonical ensemble of networks created by randomizing one property of the original network while keeping all the other constrained. We create an ensemble with randomized timestamps on the vertices and calculate all the test statistics for every realization, thus obtaining a null-distribution for motif counts and largest component sizes.

## Measures used to compare data and RRM

We start the analysis of the results by verifying that the entire statistic distributions of the simulated data and their RRM ensembles differ. We do so by performing a Kolmogorov-Smirnov test, that shows whether two distribution samples were drawn from the same underlying probability distribution. Finding an overall difference between the two samples then implies that we could further expect a motif frequency in an individual simulated

process to be different from the distribution of frequencies in its RRM.

The first measure we use to compare individual processes to the RRM is a z-score. It is a test statistic for the z-test, that requires the null-distribution to be Gaussian. We use the right-tailed z-test in accordance with our alternative hypothesis, which states that the presence of an endogenuos component drives the test statistic counts to values higher than expected from a purely exogenous process.

As the distribution of motif counts resembles the normal distribution, we estimated that it justified to use z-score to make the comparison of the motif counts. In order to verify that, we performed a Shapiro-Wilk test on all null-distributions of the motif counts and the largest component sizes. In case of causal motifs for almost all parameters used to simulate processes the assumption of normality is not rejected, with the exception of the early stages of the process, when a small percentage of the vertices are defaulted (5%-10%). This is in accordance with the KS test results for this phase of the process. However, the distributions of largest component sizes obtained from the RRM are often non-normal, with the statistical significance $p < 0.01$.

Further on, we use the generalization of the z-score, the Mahalanobis distance [38] to exploit the fact that every class of motifs contains a set of different "submotifs". We notice, as it is shown in the Table 3.1, that in the motif classes with two and three edges, the motif that is simply composed of a sequence of causal edges is the least probable to be created by an exogenous process (by chance), compared to the other submotifs in the same class. Therefore, we expect that individuating the submotifs will increase the sensitivity of our methodology.

Out of each count of causal motifs in the class $i$, we create a vector $C^{(i)}$ that has individual submotif counts as components. For the class of one-edge motifs, the dimension of the vector is 1, for two-edge motifs it is 3, and for three-edge motifs 9. The Mahalanobis distance $D^{(i)}$ between the vector $C^{(i)}$ and the null-distribution created by shuffling the timestamps is given as:

$$D^{(i)} = \sqrt{(\mathbf{C}^{(i)} - \mu^{(i)})^T \Sigma_i^{-1} (\mathbf{C}^{(i)} - \mu^{(i)})}, \tag{3.1}$$

where $\mu^{(i)}$ is vector of average submotif counts, and $\Sigma_i$ is a covariance matrix of the

different submotif counts obtained in temporally randomized networks. For a diagonal covariance matrix, we can think of the Mahalanobis distance as the square root of the sum of squares of z-scores of each vector component, i.e. submotif. From this it follows that the Mahalanobis distance is always greater or equal to each of the individual z-scores, which leads us to assume it will increase the robustness of our methodology. The statistical significance for the Mahalanobis distance is presented as the p-value, which is computed according to [127].

## Simulating stochastic processes of defaults

To validate our methodology, we create artificial data by simulating defaults with processes composed from both endogenous and exogenous components on random directed networks. We use directed Erdős-Rényi graphs as the underlying random networks. For $N$ vertices and $\langle k_{in} \rangle$ and $\langle k_{out} \rangle$ as expected in- and out-degrees, we generate an ensemble with $N_{graph}$ networks. Then, on each network from the ensemble, we simulate $N_{process}$ different realizations of a default process, determined by parameters $\alpha$ and $\beta$, that we define below. For each realization of a process, a RRM ensemble with $N_{random}$ networks is created.

We compose the process for simulation from an exogenous and an endogenous component. Each component is modeled with a Poisson process with respective rates $\alpha$ and $\beta_i = x_i\beta$. The endogenous rate $\beta_i$ for a vertex $i$ is defined as a rate $\beta$ that is common to all the vertices, weighted by some internal vertex property $x_i$. We start by drawing exogenous and endogenous event times from the cumulative distribution function of the exponential distribution:

$$F(t;\lambda) = 1 - e^{-\lambda t}, \quad t \geq 0, \quad \lambda = \alpha, x_i\beta. \tag{3.2}$$

The type of the contagion process that we employ defines the weights $x_i$; for an SI type of process the weights are equal for all the vertices $x_i = 1$, whereas in a voter model type of process the total rate depends on the in-degree of every node $x_i = 1/k_i^{in}$.

Finally, the default times resulting from the contagion, for an SI and voter model type of

contagion process are:

$$t^{\beta}_{j(SI)} = t^{\alpha}_{i} + \Delta t^{\beta}_{(i,j)}, \quad \text{or}$$
$$t^{\beta/k^{in}_j}_{j(VM)} = t^{\alpha}_{i} + k^{in}_{j}\Delta t^{\beta}_{(i,j)}, \tag{3.3}$$

and we use the event-driven algorithm to perform the simulation [86].

Since only the relative difference between the rates of the two components is what matters, we can define a ratio $\zeta := \alpha/\beta$ as a parameter that controls the endogenous component contribution to the process we simulate. Throughout the simulation, the data on test statistics is collected after every 5 % of the new vertices default, both from the ongoing process and the RRM of the network part that has been defaulted so far.

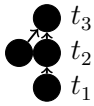
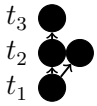
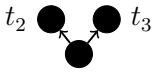
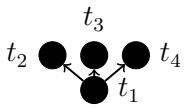
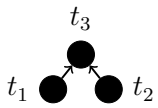
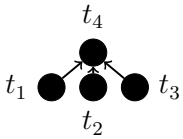
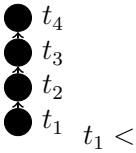
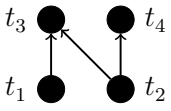
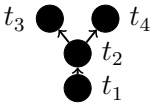
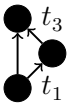
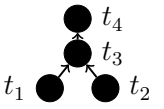
## 3.4 Analytical results

In order to provide a better understanding of the results we obtain with our methodology, we conducted some analytical calculations. In this section we first present the frequencies of submotifs of each class that we would expect in a network with randomly distributed default times, along with a Table 3.1 showing all the submotifs we use. We follow up with calculating a ratio between exogenously and endogenously defaulted vertices depending on the total number of defaults, to get a clear picture of how the simulated process unfolds on the network. Finally, we present an extension of the approximate master equations [22] that encompasses directed networks with temporal ordering, which we then use to calculate the temporal evolution of the expected frequencies of the causal motifs.

### 3.4.1 Expected frequency of submotifs in an exogenous process

Since our methodology relies on the expectation that the frequency of the causal motifs changes when an endogenous contribution is included in some exogenous default process, we find it useful to first find the frequency that is expected when only the exogenous process is present. In general, the probability of finding a causal submotif $\mathcal{CM}$ in a network is proportional to *(i)* the probability $p(\mathcal{M})$ of finding a structural motif $\mathcal{M}$

Table 3.1: **All possible acyclic causal motifs up to order 3.** The probability represents the probability of occurrence for every a motif in the case when only the exogenous influence exists.

| Causal Motif | Probability | Causal Motif | Probability |
|---|---|---|---|
|  $t_1 < t_2$ | $p(\mathcal{M})\frac{1}{2}\pi(T)^2$ |  $t_1 < t_2 < t_3;\ t_4 < t_3$ | $p(\mathcal{M})\frac{1}{8}\pi(T)^4$ |
|  $t_1 < t_2 < t_3$ | $p(\mathcal{M})\frac{1}{6}\pi(T)^3$ |  $t_1 < t_2 < t_3;\ t_1 < t_4$ | $p(\mathcal{M})\frac{1}{8}\pi(T)^4$ |
|  $t_1 < t_2, t_3$ | $p(\mathcal{M})\frac{1}{3}\pi(T)^3$ |  $t_1 < t_2, t_3, t_4$ | $p(\mathcal{M})\frac{1}{4}\pi(T)^4$ |
|  $t_1, t_2 < t_3$ | $p(\mathcal{M})\frac{1}{3}\pi(T)^3$ |  $t_1, t_2, t_3 < t_4$ | $p(\mathcal{M})\frac{1}{4}\pi(T)^4$ |
|  $t_1 < t_2 < t_3 < t_4$ | $p(\mathcal{M})\frac{1}{24}\pi(T)^4$ |  $t_1, t_2 < t_3;\ t_2 < t_4$ | $p(\mathcal{M})\frac{5}{24}\pi(T)^4$ |
|  $t_1 < t_2 < t_3, t_4$ | $p(\mathcal{M})\frac{1}{12}\pi(T)^4$ |  $t_1 < t_2 < t_3$ | $p(\mathcal{M})\frac{1}{6}\pi(T)^3$ |
|  $t_1, t_2 < t_3 < t_4$ | $p(\mathcal{M})\frac{1}{12}\pi(T)^4$ | | |

on the underlying network, that could serve as the basis of the causal motif, *(ii)* the probability $\mathcal{P}(T)$ that all the vertices in the structural motif have defaulted by the time $T$, and *(iii)* $p(\mathcal{CM})$ the probability that the default times on its vertices obey the time ordering necessary to make it a causal motif.

The probability $p(\mathcal{M})$ depends exclusively on the network topology, and for any existing structural motif we can choose the time $T$ when all the vertices have already defaulted. Therefore, the probability that is of interest in this section is the third one, $p(\mathcal{CM})$. For an exogenous process, which produces the timestamps in a completely random way, it can be calculated simply using combinatorics.

An alternative approach that proved to be simpler in our case consists of integrating the

exogenous rate of default. Let us denote the exogenous rate of default per unit time as $\phi(t)$. Now we can take the submotif shaped like letter N as an example and write the total probability of the appearance of a submotif in an exogenous process as:

$$\begin{aligned} P(\mathcal{CM}) &= p(\mathcal{M}) \int_0^T \int_0^{t_3} \left( \int_0^{t_3} \phi(t_1) dt_1 \int_{t_2}^T \phi(t_4) dt_4 \right) \phi(t_2)\phi(t_3) dt_2 dt_3 \\ &= p(\mathcal{M}) \int_0^T \int_0^{t_3} \pi(t_3)[\pi(T) - \pi(t_2)]\phi(t_2) dt_2 \phi(t_3) dt_3 \\ &= p(\mathcal{M}) \frac{5}{24} \pi(T)^4, \end{aligned} \tag{3.4}$$

with $\pi(\tau)$ defined as $\pi(\tau) := \int_0^\tau \phi(t)dt$.

We can see from Table 3.1 that increasing the order of the motifs reduces its probability to occur in an exogenous process, and that this probability varies within a certain order of motifs. Therefore using the higher order motif frequency, and distinguishing between the submotifs in the same order is a justified choice for our methodology.

### 3.4.2 Ratio of the exogenously and endogenously defaulted vertices

In the previous section, we defined the parameter $\zeta$ to quantify the endogeneity of the process that we simulate. However, as it is a macroscopic parameter, we also wish to inspect the microscopical implications it has. We want to calculate the numbers of exogenously and endogenously defaulted vertices, $n_\alpha$ and $n_\beta$, respectively, and to do so we use the mean field approximation.

Let the process take place on an Erdős-Rényi network with $N$ vertices, and average in- and out-degrees $\langle k_i \rangle$ and $\langle k_o \rangle$. In the time increment $dt$, the number of exogenously defaulted vertices, $dn_\alpha$, will increase proportionally to the number of the non-defaulted vertices $N - n_\alpha - n_\beta$, rate $\alpha$ and that time increment $dt$:

$$dn_\alpha = (N - n_\alpha - n_\beta)\alpha dt. \tag{3.5}$$

For endogenously defaulted vertices in an SI process, the increase $dn_\beta$ is proportional to

the number of defaulted vertices $n = n_\alpha + n_\beta$, and their average out degree $\langle k_o \rangle$, which gives the number of edges through which the default can potentially propagate. The propagation, however, depends on the vertices on the other side of those edges, which are non-defaulted with the mean field probability $1 - \frac{n_\alpha + n_\beta}{N}$. Altogether with the endogenous rate of default $\beta$, it leads to:

$$dn_\beta = (n_\beta + n_\alpha)\langle k_o \rangle (1 - \frac{n_\alpha + n_\beta}{N})\beta dt. \tag{3.6}$$

From the form of Equations (3.5) and (3.6) and the fact that both $n_\alpha$ and $n_\beta$ are monotonously increasing functions of time, we see that we can remove the time dependence, and get a "phase space" differential equation

$$\frac{dn_\alpha}{dn_\beta} = \frac{N\zeta}{\langle k_o \rangle (n_\alpha + n_\beta)}, \tag{3.7}$$

$$dn_\beta = \frac{\langle k_o \rangle}{N\zeta}(n_\alpha dn_\alpha + n_\beta dn_\alpha). \tag{3.8}$$

We can integrate Equation (3.8) to obtain:

$$n = \frac{N\zeta}{\langle k_o \rangle}\left(e^{\frac{\langle k_o \rangle n_\alpha}{N\zeta}} - 1\right), \tag{3.9}$$

$$\frac{n_\alpha}{n_\beta} = \frac{\frac{\zeta}{\langle k_o \rangle} ln(\frac{d\langle k_o \rangle}{\zeta} + 1)}{d - \frac{\zeta}{\langle k_o \rangle} ln(\frac{d\langle k_o \rangle}{\zeta} + 1)}, \tag{3.10}$$

with $n = n_\alpha + n_\beta$ as the total number of defaulted vertices in the network and $d = n/N$ as the total default percentage in the network with $N$ vertices.

If we want to obtain the same expression in the case of a voter model process, we simply substitute $\zeta \to \zeta\langle k_i \rangle$. Considering also $\langle k_i \rangle = \langle k_o \rangle$, we get:

$$n = N\zeta\left(e^{\frac{n_\alpha}{N\zeta}} - 1\right), \tag{3.11}$$

$$\frac{n_\alpha}{n_\beta} = \frac{\zeta ln(\frac{d}{\zeta} + 1)}{d - \zeta ln(\frac{d}{\zeta} + 1)}. \tag{3.12}$$

If we compare Equations (3.10) and (3.12), we see that for a given $\zeta$ and with $\langle k_o \rangle > 1$, the voter model process will always have a higher ratio $n_\alpha/n_\beta$ than the SI process. We also notice that for the voter model process the ratio does not depend on the degree of the

network. For both processes the ratio decreases with the increase of the total percentage of defaulted vertices. This all means that, for the same parameters, we expect to be able to detect endogeneity more easily in the simulated processes that are of SI type and in later stages of the contagion.

### 3.4.3 Approximate master equations for motif frequencies

Considering the results from the first subsection, we can easily calculate the expected frequencies for all the motif types in the case of a completely exogenous process on the network. However, when there is a contagion process in the interplay with the exogenous one, things get more complicated, since the vertices in the emerging causal motifs no longer default independently from one another. At this point we resort to Approximate Master Equations (AME), [21], already mentioned in Section 2.2 and generalize them to directed networks with temporal ordering.

Although we are focused on our immediate application to the network of defaults in the economy, here we resort to the epidemic spreading terminology, to build on the terminology in [21]. We take the reasoning from the Section 2.2 and simply extend classifying the vertices by their degree $k$ and infected neighbours number $m$, into the in-degree $k_i$, out-degree $k_o$, infected in-neighbors $m_i$ and out-neighbors $m_o$. In addition to that, we introduce $z_i$ and $z_o$ to be the numbers of in- and out-neighbors whose default can be in a possibly causal relationship with the default of the central vertex. By definition, we see that the subsetting into $z_i$ and $z_o$ makes sense only for the infected compartment.

We start by assuming that the network is completely connected in a weak sense (some vertices can have only in or only out edges attached to them) and that it can have correlations on vertices i.e.

$$P(k_i, k_o) \neq P(k_i)P(k_o). \tag{3.13}$$

Let $S_{k_i,k_o,m_i,m_o}$ (respectively $I_{k_i,k_o,m_i,m_o,z_i,z_o}$) be the set of vertices which are susceptible (respectively infected), have in-degree $k_i$, out-degree $k_o$ and have $m_i$ and $m_o$ infected neighbors on their incoming and outgoing edges, which we can further subset to $z_i$ and $z_o$ neighbours that comprise a potentially causal edge together with the central node ($z_i =$

$z_o = 0$ by definition for susceptible vertices). In economic terms, susceptible vertices are those which are financially healthy, while infected vertices are those which have financial problems or have already defaulted. On these sets we define measures $s_{k_i,k_o,m_i,m_o}$ and $i_{k_i,k_o,m_i,m_o,z_i,z_o}$ which represent the fraction of vertices that are susceptible or infected among the vertices that have their $\{k_i, k_o, m_i, m_o, z_i, z_o\}$ parameters. We define $i^z_{k_i,k_o,m_i,m_o}$ and $i^m_{k_i,k_o,z_i,z_o}$ as the fractions for which we stop keeping track of $z$ and $m$ subsets of neighbours, respectively, by simply summing them up:

$$i^z_{k_i,k_o,m_i,m_o} \equiv \sum_{z_i=0}^{k_i} \sum_{z_o=0}^{k_o} i_{k_i,k_o,m_i,m_o,z_i,z_o}, \tag{3.14}$$

$$i^m_{k_i,k_o,z_i,z_o} \equiv \sum_{m_i=0}^{k_i} \sum_{m_o=0}^{k_o} i_{k_i,k_o,m_i,m_o,z_i,z_o}. \tag{3.15}$$

Clearly, the fractions $s_{k_i,k_o,m_i,m_o}$ and $i^z_{k_i,k_o,m_i,m_o}$ need to add up to the entire system:

$$1 = s_{k_i,k_o,m_i,m_o} + i^z_{k_i,k_o,m_i,m_o}. \tag{3.16}$$

Also the total fraction of infected (defaulted) vertices in the network, with in-degree $k_i$ and out-degree $k_o$, is defined as

$$\begin{aligned} \rho_{k_i k_o}(t) &= \sum_{m_i=0}^{k_i} \sum_{m_o=0}^{k_o} i^z_{k_i,k_o,m_i,m_o} \\ &= 1 - \sum_{m_i=0}^{k_i} \sum_{m_o=0}^{k_o} s_{k_i,k_o,m_i,m_o}. \end{aligned} \tag{3.17}$$

Then, the total fraction of infected vertices, in a network with the degree distribution $P(k_i, k_o)$, is found by:

$$\rho(t) = \sum_{k_i,k_o} P(k_i, k_o)\, \rho_{k_i,k_o}(t). \tag{3.18}$$

To be able to follow this temporal evolution, we need to examine how the sizes of the $S_{k_i,k_o,m_i,m_o}$ and $I_{k_i,k_o,m_i,m_o,z_i,z_o}$ sets change in time. We write the general expression for

the fractions of susceptible and infected vertices:

$$\begin{aligned}
s_{k_i,k_o,m_i,m_o}(t+dt) = s_{k_i,k_o,m_i,m_o}(t) &- W(S_{k_i,k_o,m_i,m_o} \to I_{k_i,k_o,m_i,m_o,z_i=m_i,z_o=0})\, s_{k_i,k_o,m_i,m_o}\, dt \\
&+ W(S_{k_i,k_o,m_i\ \ 1,m_o} \to S_{k_i,k_o,m_i,m_o})\, s_{k_i,k_o,m_i\ \ 1,m_o}\, dt \\
&+ W(S_{k_i,k_o,m_i,m_o\ \ 1} \to S_{k_i,k_o,m_i,m_o})\, s_{k_i,k_o,m_i,m_o\ \ 1}\, dt \\
&- W(S_{k_i,k_o,m_i,m_o} \to S_{k_i,k_o,m_i+1,m_o})\, s_{k_i,k_o,m_i,m_o}\, dt \\
&- W(S_{k_i,k_o,m_i,m_o} \to S_{k_i,k_o,m_i,m_o+1})\, s_{k_i,k_o,m_i,m_o}\, dt,
\end{aligned}$$

$$\begin{aligned}
i_{k_i,k_o,m_i,m_o,z_i,z_o}(t+dt) = i_{k_i,k_o,m_i,m_o,z_i,z_o}(t) &+ W(S_{k_i,k_o,m_i,m_o} \to I_{k_i,k_o,m_i,m_o,z_i=m_i,z_o=0})\, s_{k_i,k_o,m_i,m_o}\, dt \\
&+ W(I_{k_i,k_o,m_i\ \ 1,m_o,z_i,z_o} \to I_{k_i,k_o,m_i,m_o,z_i,z_o})\, i_{k_i,k_o,m_i\ \ 1,m_o,z_i,z_o}\, dt \\
&+ W(I_{k_i,k_o,m_i,m_o\ \ 1,z_i,z_o\ \ 1} \to I_{k_i,k_o,m_i,m_o,z_i,z_o})\, i_{k_i,k_o,m_i,m_o\ \ 1,z_i,z_o\ \ 1}\, dt \\
&- W(I_{k_i,k_o,m_i,m_o,z_i,z_o} \to I_{k_i,k_o,m_i+1,m_o,z_i,z_o})\, i_{k_i,k_o,m_i,m_o,z_i,z_o}\, dt \\
&- W(I_{k_i,k_o,m_i,m_o,z_i,z_o} \to I_{k_i,k_o,m_i,m_o+1,z_i,z_o+1})\, i_{k_i,k_o,m_i,m_o,z_i,z_o}\, dt, \qquad (3.19)
\end{aligned}$$

and represent it graphically in Fig. 3.2. Equation (3.19) accounts for the transitions which are linear in $dt$, as all the other transitions vanish in the limit of $dt \to 0$. The fundamental transition probability describes the transition from the set $S_{k_i,k_o,m_i,m_o}$ to $I_{k_i,k_o,m_i,m_o,z_i,z_o}$ and is denoted $W(S_{k_i,k_o,m_i,m_o} \to I_{k_i,k_o,m_i,m_o,z_i=m_i,z_o=0})$. We introduce the nature of the dynamical process into the framework through the transition probability. For the particular processes that we used in our simulations, containing an exogenous and an endogenous component (SI or the voter model), with the individual rates $\alpha$ and $\beta$ already defined in Section 3.3, the transition rates are:

$$\begin{aligned}
W(S_{k_i,k_o,m_i,m_o} \to I_{k_i,k_o,m_i,m_o,z_i=m_i,z_o=0}) &= F_{k_i,m_i}, \\
F^{SI}_{k_i,m_i} &= \alpha + m_i\beta, \\
F^{VM}_{k_i,m_i} &= \alpha + \frac{m_i}{k_i}\beta. \qquad (3.20)
\end{aligned}$$

The transition probability of the central vertex is described exactly by Eq. (3.20). As we have limited our sets to central vertices and their first neighbours, we still need to obtain the probabilities of the neighbours' transitions. In order to do that, we approximate the effect the rest of the network has on them to be a mean field. Therefore, rates concerning defaults of neighbouring vertices in Equation (3.19) are factored into the mean field rate of change of the edge type and the number of susceptible neighbours of the central

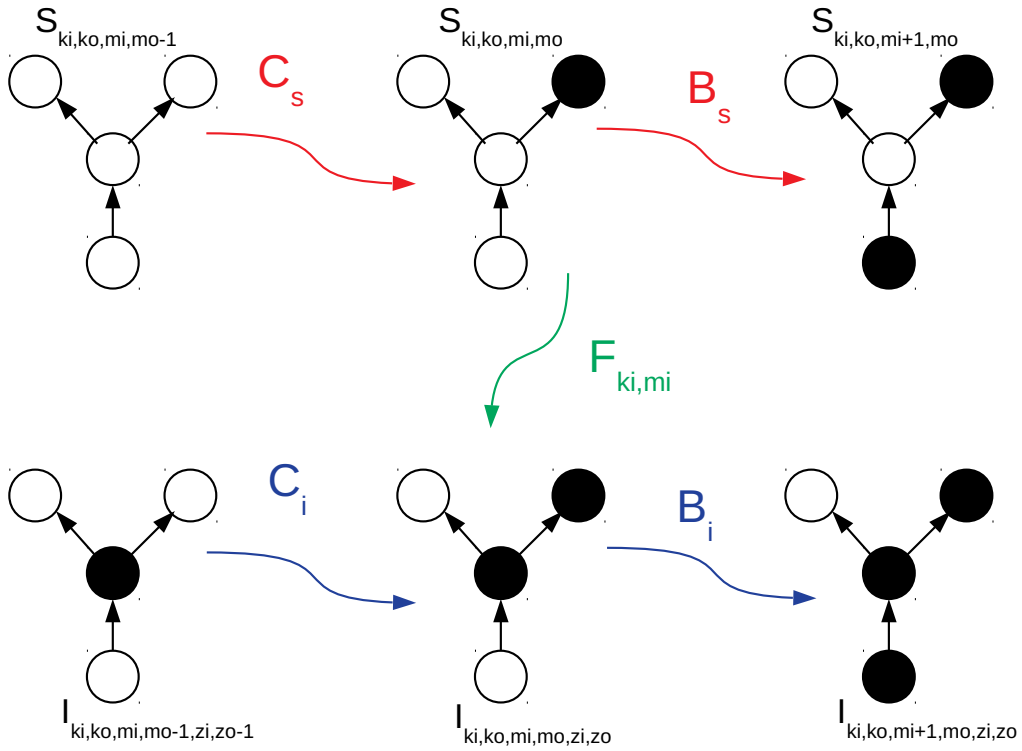



Figure 3.2: **The scheme of transitions between different sets.** A set consisting of a central vertex and its in- and out- neighbours is depicted. We show the susceptible vertices in white and the infected in black. All the possible transitions in our process are represented together with their respective rates.

vertex. In Fig. 3.2, the rates of change of the edge types are visualized and denoted with $B_s, C_s, B_i$ and $C_i$. We will use the following abbreviations in the further calculations: $k_i, k_o, m_i, m_o \equiv \vec{K}_s$, $k_i, k_o, m_i, m_o, z_i, z_o \equiv \vec{K}_i$, $k_i, k_o, m_i + 1, m_o, z_i, z_o \equiv \vec{K}_i + \vec{e}_{m_i}$. For the first equation in (3.19) we have:

$$\begin{aligned} W(S_{\vec{K}_s} \to S_{\vec{K}_s + \vec{e}_{m_i}}) &= B_s(k_i - m_i), \\ W(S_{\vec{K}_s - \vec{e}_{m_i}} \to S_{\vec{K}_s}) &= B_s(k_i - m_i + 1), \\ W(S_{\vec{K}_s} \to S_{\vec{K}_s + \vec{e}_{m_o}}) &= C_s(k_o - m_o), \\ W(S_{\vec{K}_s - \vec{e}_{m_o}} \to S_{\vec{K}_s}) &= C_s(k_o - m_o + 1). \end{aligned} \tag{3.21}$$

The term $B_s$ represents the rate of change of edges directed from a susceptible vertex into a susceptible vertex ($S \to S$) into edges directed from an infected vertex into a susceptible

vertex $(I \to S)$:

$$B_s\,dt = \frac{\sum_{k_i,k_o} P(k_i,k_o) \sum_{m_i}^{k_i} \sum_{m_o}^{k_o} (k_o - m_o) F_{k_i,m_i}\, s_{\vec{K}_s}\, dt}{\sum_{k_i,k_o} P(k_i,k_o) \sum_{m_i}^{k_i} \sum_{m_o}^{k_o} (k_o - m_o)\, s_{\vec{K}_s}}. \tag{3.22}$$

The denominator counts the number of $S \to S$ edges, while the numerator counts how many of them change to $I \to S$ in a time interval $dt$. $C_s$ is defined for the transition from $S \to S$ to $S \to I$, and computed similarly.

$$C_s\,dt = \frac{\sum_{k_i,k_o} P(k_i,k_o) \sum_{m_i}^{k_i} \sum_{m_o}^{k_o} (k_i - m_i) F_{k_i,m_i}\, s_{\vec{K}_s}\, dt}{\sum_{k_i,k_o} P(k_i,k_o) \sum_{m_i}^{k_i} \sum_{m_o}^{k_o} (k_i - m_i)\, s_{\vec{K}_s}}. \tag{3.23}$$

For the second equation in (3.19), rates of neighbour infections, $B_i$ and $C_i$, are defined analogously.

$$\begin{aligned}
W(I_{\vec{K}_i} \to I_{\vec{K}_i + \vec{e}_{m_i}}) &= B_i(k_i - m_i), \\
W(I_{\vec{K}_i - \vec{e}_{m_i}} \to I_{\vec{K}_i}) &= B_i(k_i - m_i + 1), \\
W(I_{\vec{K}_i} \to I_{\vec{K}_i + \vec{e}_{m_o} + \vec{e}_{z_o}}) &= C_i(k_o - m_o), \\
W(I_{\vec{K}_i - \vec{e}_{m_o} - \vec{e}_{z_o}} \to I_{\vec{K}_i}) &= C_i(k_o - m_o + 1).
\end{aligned} \tag{3.24}$$

After inserting the calculated rates into Eq. (3.19), dividing by $dt$ and taking the limit $dt \to 0$, we get the differential equation for the evolution of the fraction of susceptible vertices:

$$\begin{aligned}
\frac{d}{dt} s_{\vec{K}_s} = {} & B_s \left( (k_i - m_i + 1) s_{\vec{K}_s - \vec{e}_{m_i}} - (k_i - m_i) s_{\vec{K}_s} \right) \\
& + C_s \left( (k_o - m_o + 1) s_{\vec{K}_s - \vec{e}_{m_o}} - (k_o - m_o) s_{\vec{K}_s} \right) \\
& - F_{k,m} s_{\vec{K}_s},
\end{aligned} \tag{3.25}$$

and of the fraction of infected vertices:

$$\begin{aligned}\frac{d}{dt}i_{\vec{K}_i} &= B_i\left((k_i - m_i + 1)i_{\vec{K}_i - \vec{e}_{m_i}} - (k_i - m_i)i_{\vec{K}_i}\right) \\ &+C_i\left((k_o - m_o + 1)i_{\vec{K}_i - \vec{e}_{m_o} - \vec{e}_{z_o}} - (k_o - m_o)i_{\vec{K}_i}\right) \\ &+ F_{k,m}s_{\vec{K}_s, z_i = m_i, z_o = 0}.\end{aligned} \tag{3.26}$$

The solutions $s_{\vec{K}_i}$ and $i_{\vec{K}_i}$ of the Eq. (3.25) and (3.26) are obtained numerically. Using $i_{\vec{K}_i}$, we can easily calculate the temporal evolution of the fraction of causal motifs with one and two edges, for the central vertex with $k_i$ and $k_o$. We denote them $c^{(1)}_{k_i,k_o}(t)$ and $c^{(2)}_{k_i,k_o}(t)$, respectively.

$$\begin{aligned}c^{(1)}_{k_i,k_o}(t) &= \sum_{z_i=0}^{k_i}\sum_{z_o=0}^{k_o} z_o i^m_{k_i,k_o,z_i,z_o}(t), \\ c^{(2)}_{k_i,k_o}(t) &= \sum_{z_i=0}^{k_i}\sum_{z_o=0}^{k_o}\left(\binom{z_i}{2} + z_i z_o + \binom{z_o}{2}\right) \times i^m_{k_i,k_o,z_i,z_o}(t).\end{aligned} \tag{3.27}$$

The temporal evolution of the fraction of causal motifs with one and two edges, for the central vertex with $k_i$ and $k_o$, when the time ordering of the defaulted vertices is randomly distributed is denoted as $\tilde{c}^{(1)}_{k_i,k_o}(t)$ and $\tilde{c}^{(2)}_{k_i,k_o}(t)$ and equals :

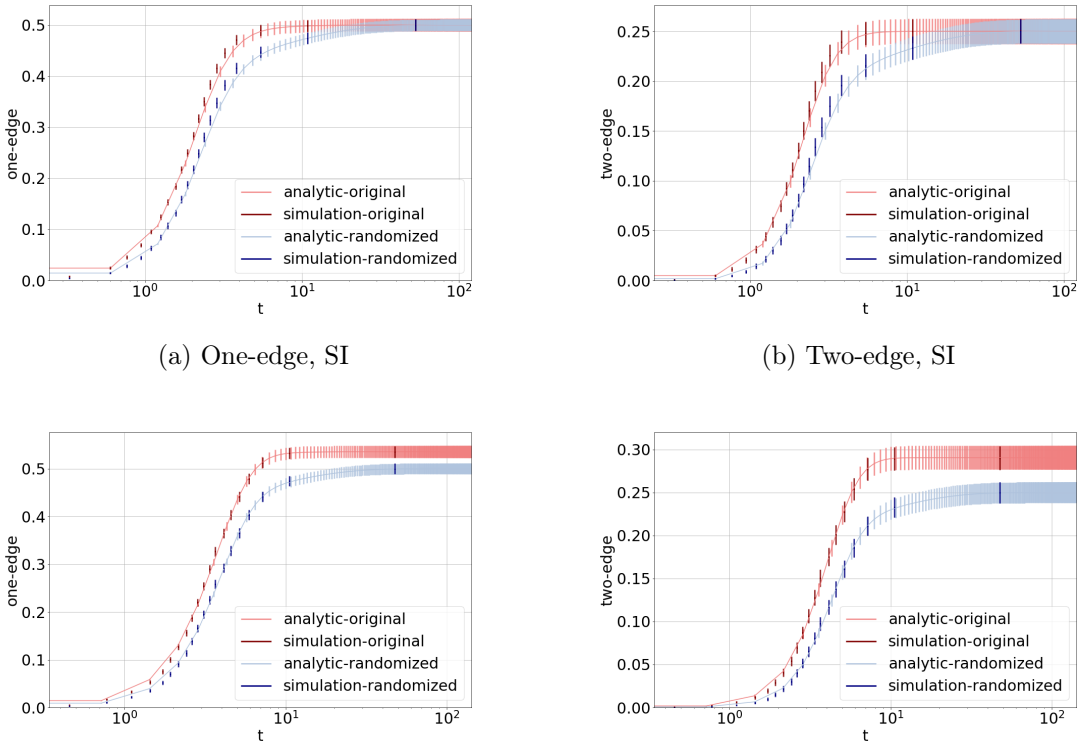
$$\begin{aligned}\tilde{c}^{(1)}_{k_i,k_o}(t) &= \frac{1}{2}\sum_{m_i=0}^{k_i}\sum_{m_o=0}^{k_o} m_o i^z_{k_i,k_o,m_i,m_o}(t), \\ \tilde{c}^{(2)}_{k_i,k_o}(t) &= \sum_{m_i=0}^{k_i}\sum_{m_o=0}^{k_o}\left(\frac{1}{3}\binom{m_i}{2} + \frac{1}{6}m_i m_o + \frac{1}{3}\binom{m_o}{2}\right) \times i^z_{k_i,k_o,m_i,m_o}(t).\end{aligned} \tag{3.28}$$

If we want to calculate the expected frequency of causal motifs on the entire network we only need to specify the distribution $P(k_i, k_o)$. The first moments for one and two edges are:

$$E[\mathcal{C}^{(1,2)}] = N\sum_{k_i,k_o} P(k_i,k_o)c^{(1,2)}_{k_i,k_o}(t), \tag{3.29}$$

$$E[\tilde{\mathcal{C}}^{(1,2)}(t)] = N\sum_{k_i,k_o} P(k_i,k_o).\tilde{c}^{(1,2)}_{k_i,k_o}(t). \tag{3.30}$$

Calculating the second moments is a bit more complicated, as the overlaps of motifs in the




(a) One-edge, SI

(b) Two-edge, SI

(c) One-edge, VM

(d) Two-edge, VM

Figure 3.3: **Temporal evolution of causal motifs.** We plot the comparison of the analytical results with the simulations. The figure shows one-edge and two-edge statistics for $\zeta = 0.1$ SI and VM process on an Erdős-Rényi graph with $N = 1000, \langle k \rangle = 4$. On x-axis the time of the process evolution is shown, while y-axis represents the relative frequency of a test statistic. The standard deviation, shown by the error bars, is calculated in an exact way for the one-edge statistic, while for the two-edge statistic it is only approximated.

network make the sum contain no longer independent variables, and introduce covariant terms. One can use the approach developed in [81] to obtain the correct variances but only for the purely exogenous case. In that case expected square of the number of one edge causal motifs on Erdős-Rényi network can be computed as:

$$\begin{aligned} E[\mathcal{C}^{(1)^2}] =& E[\sum_{a=k_i,k_o} \sum_{b=q_i,q_o} Y_a(\mathcal{C}^{(1)}|t) Y_b(\mathcal{C}^{(1)}|t)] \\ =& \sum_{a=k_i,k_o} \sum_{b=q_i,q_o} E[Y_a(\mathcal{C}^{(1)}|t) Y_b(\mathcal{C}^{(1)}|t)] \\ =& N(N-1)(N-2)(N-3)(p\frac{1}{2}\pi(t)^2)^2 + N\binom{N-1}{2} p^2 \frac{1}{3}\pi(t)^3 \\ &+ N(N-1)(p\frac{1}{2}\pi(t))^2, \end{aligned} \tag{3.31}$$

where $Y_\alpha(\mathcal{C}^{(1)}|t)$ similarly to [81] represents the indicator random variable of finding the causal motif of order one (causal edge) between two random vertices. Clearly the random indicator variable is connected to probabilities of occurrence of one and two edge causal motifs computed in Table 3.1, and $p$ is just the probability of edge existing in the directed version of Erdős-Rényi graph.

With similar equations one can compute all the motif variances in a purely exogenous process. However, in a process with an endogenous component, we can no longer calculate the probabilities that causal motifs appear in this way, and obtaining an estimate for the motif variances in a dynamical process is the next step in our research.

Finally, we present some plots showing a good fit of our analytical and simulation results in Fig. 3.3. The figures depict the temporal evolution of one-edge and two-edge motifs for $\zeta = 0.1$ SI and VM process on an Erdős-Rényi graph with $N = 1000$, $\langle k \rangle = 4$.

## 3.5 Results

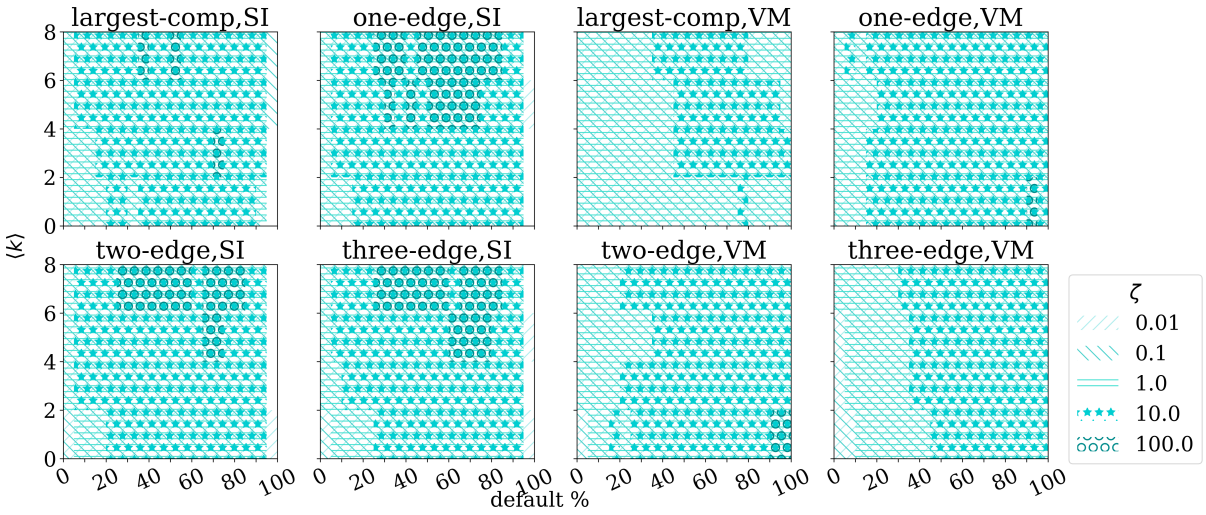



Figure 3.4: **KS-test results.** We show the results for SI and VM processes on a network of $N = 1000$ vertices. The x-axis depicts the percentage of defaults in the network and the y-axis depicts the average degree of the network. The significant statistic ($p < 0.01$) for a given percentage of default, $\zeta$ and mean degree $\langle k \rangle$ is marked with the corresponding pattern.

We get the results from $N_{process} = 10$ processes simulated on $N_{graph} = 100$ different graphs, with $N_{shuffle} = 100$ randomized reference models created for each process, in the

form of two distributions of test statistics. One is the distribution of the test statistics obtained from the simulated processes, and the other is the distribution of the test statistics recorded on the randomized reference models.

We first employ Kolmogorov-Smirnov test to check whether the two sample distributions belong to the same probability distribution, and if not, to quantify their difference. We follow up by using z-score to compare each instance of a simulated process with its own randomized version. Finally, we utilize the difference between frequencies of submotifs expected in a randomized model, as explained in the previous Section 3.4, and generalize from the z-score to employing Mahalanobis distance.

The simulated process is composed of an exogenous and an endogenous component, quantified by the parameters $\alpha$ and $\beta$, respectively. Parameter $\alpha$ represents the rate at which a vertex defaults due to the influence of the factors external to the network, and parameter $\beta$ is related to the rate that the default is propagated from immediate neighbours. The precise rate depends on the type of the endogenous process. For the SI process it is simply the product of the number of defaulted debtor vertices (in-neighbours) and the rate $\beta$, while for the voter model this product is also weighted by the total number of debtor vertices. Since the absolute flow of time is not relevant, we can set the ratio of the component parameters $\zeta = \frac{\alpha}{\beta}$ as a control parameter that quantifies the strengths of the endogenous component . We use $\zeta$ to present and interpret the results. If we have a single endogenous and a single exogenous process, $\zeta$ also represents the ratio of the increments in the numbers of exogenously and endogenously defaulted vertices. However, on the network, the exogenous default of every vertex is a separate Poisson process with rate $\alpha$, and the endogenous default depends on the number of defaulted in-neighbours as each neighbour propagates the default at rate $\beta$ in an SI process (or $\beta/k_{in}$ in a voter model process). The ratio $\Delta n_\alpha/\Delta n_\beta$ will then in general be different from $\zeta$ and the exact expression is given by Equation (3.8) in the Section 3.4. In addition to $\zeta$, it depends on the total number of vertices, the mean degree and the percentage of vertices that defaulted.

We simulate on networks of size $N = 1000$ nodes, which is the typical size of financial networks, and they are small enough to explore the parameter space with no computational limitations. Later, we also show that the significance of detection scales with the size of the network.

We apply the methodology to the real data of Croatian companies' defaults and show the results in the last section.

## SI model

First, we will have a look at the results obtained from the simulations in which the endogenous component of the process is modeled as an SI process. We employ the two sample Kolmogorov-Smirnov test (Fig. 3.4), the z-score and Mahalanobis distance (Fig. 3.5 and 3.6) to interpret the results.

We start the analysis of the results by using the KS test to compare the distributions obtained from the original processes with their respective randomized distributions. As we already stated, we do this as a pre-screening of the results, since we cannot expect an individual process realization to significantly differ from its randomized ensemble if their respective distributions are shown to be the same by the KS test. We depict the KS test results in the four left panels of Fig. 3.4. We can conclude that, for $\zeta$ up to $\zeta = 10$, the process distributions of all the test statistics are significantly different from the null distributions, except when the process is only starting, so the defaulted portion of the network is small, and at the very end of the default process, which is in line with the clarification further below. For the processes with the largest values of $\zeta$ ($\zeta = 100$), the significant difference is observed only for larger mean degrees, and the intermediate fraction of the defaulted vertices. However, the significant difference provided by the causal motifs covers a larger part of the parameter space in comparison with the causal largest component.

Since the one-tailed z-test only gives the information whether a single process is significantly different from its randomized reference model or not, and we wish to make the conclusion more robust, we present the results in the following way; we test the z-scores of all processes for the statistical significance $p < 0.1$, and we show the parts of the parameter space ($\langle k \rangle$, $\zeta$, $n$) where the percentage of different process realizations with a statistically significant z-score is larger than 70% in a darker shade of blue (Figure 3.5). We can conclude that any of the causal motifs are better test statistics than the largest component, which is in line with the fact that the distribution of the size of the largest

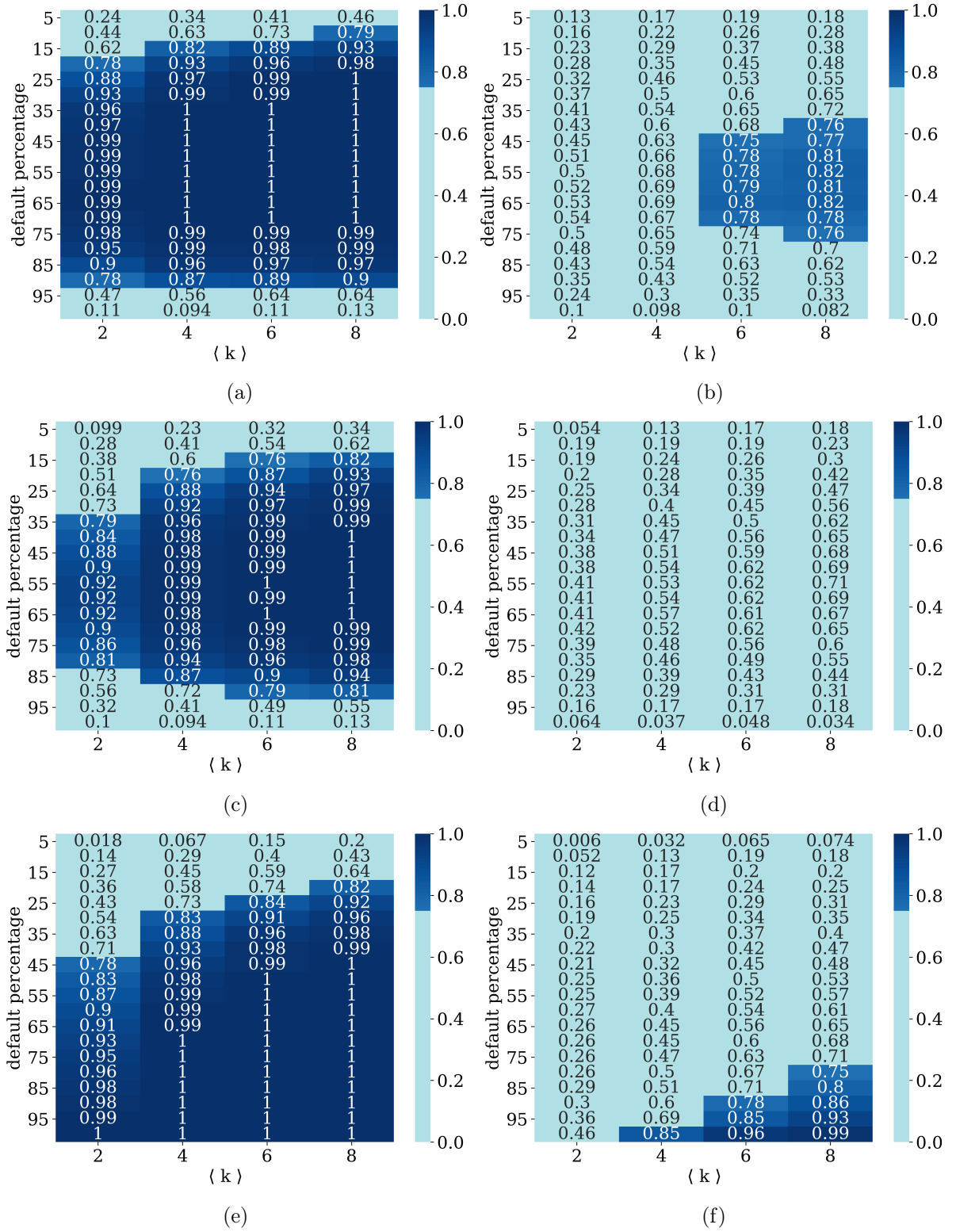



Figure 3.5: **SI process results.** In panels a) and b) we present the one-tailed z-score results, depending on the percentage of network that defaulted (y-axis) and average network degree (x-axis), for an SI process with $\zeta = 1$ and $\zeta = 4$, respectively. The results that are significant with $p < 0.1$ are marked in dark blue. In panels c) and d) we present in the same form the results of the one-tailed z-score of two-edge motifs, and in panels e) and f) we present the results of the Mahalanobis distance of two-edge motifs. For these figures, the number of vertices equals $N = 1000$, on each of 100 realized networks, 10 processes are simulated and 100 random shuffles are created.

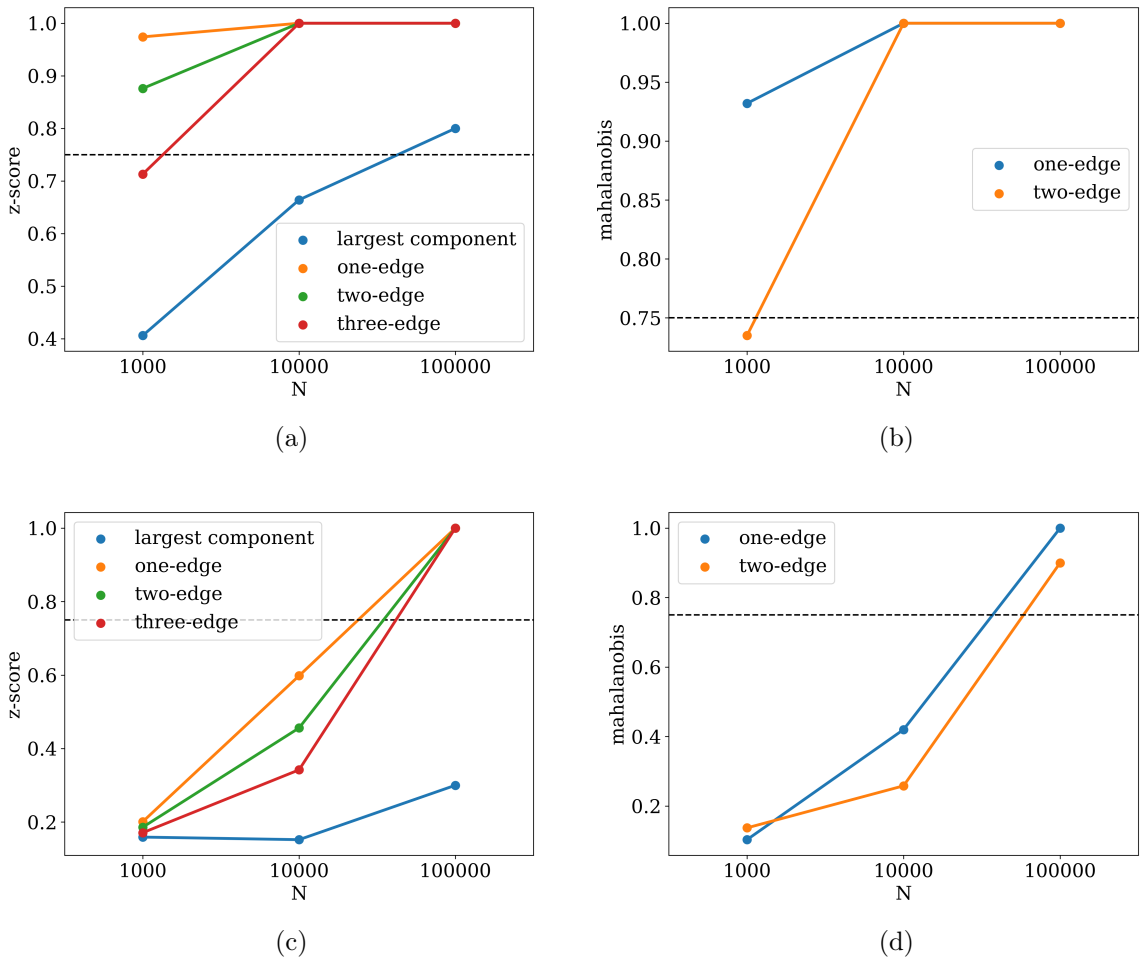



Figure 3.6: **SI process results - scaled.** We present the dependence of the percentage of significant ($p < 0.1$) results for an SI process with $\langle k \rangle = 4$, fixed at 25% of the network default, on the size of the network. In panels a) and b) we show the one-tailed z-score and Mahalanobis distance results, respectively, for an SI process with $\zeta = 1$. In panels c) and d) we show the results when the parameter of the SI process is increased to $\zeta = 10$. Horizontal lines are drawn at the limit in which 75% of simulations exhibit significant difference from the RRM-model.

component is not Gaussian. However, it is surprising that the area of the parameter space that is considered significant shrinks with the increasing order of the causal motif. Even though the distance between the first moments of the original and randomized distribution increases with the order of the causal motifs, there is also an increase in the variance of the distributions, which shields the successfulness of the higher order causal motif statistics. The comparison between one-edge and two-edge statistics can be seen in Subfigures 3.5a and 3.5c, and 3.5b and 3.5d. In these Subfigures we can also observe that the area of significant results is shrunk with the increase of the parameter $\zeta$. Subfigures 3.6a and 3.6c show how an increase of the system size increases the significant part of the parameter space.

We can notice in the Figure 3.3 that for the SI process, the expected frequency of causal motifs from both the original process and randomized reference model converges in time to a constant fraction of the frequency of all possible motifs. For the one-edge and two-edge motifs the fraction is 1/2 and 1/4, respectively. This value is clearly expected for causal motifs on randomized networks, where the time ordering of the process is destroyed. However, it is observed regardless of the endogenous strength of the process, so we present a short calculation for the expected number of causal one-edges to elucidate the effect. We assume that we have a completely endogenous process, that is, the defaults can only happen by contagion, and in each time step $t$ a defaulted vertex spreads default. The probability that an out-edge of a defaulted vertex reaches a non-defaulted vertex equals the total fraction of the remaining non-defaulted vertices $1 - t/N$, since SI process does not discriminate between neighbours, and the Erdős-Rényi network is homogenenous. Considering also that the number of outgoing edges per vertex is $\langle k_{out} \rangle$ on average, we can write the expected sum of causal one-edges as:

$$E[\mathcal{C}^{(1)}] = \langle k_{out} \rangle \sum_{t=1}^{N} \frac{N-t}{N} = \frac{\langle k_{out} \rangle}{N} \frac{N^2 - N}{2} \approx \frac{N\langle k \rangle}{4} = \frac{L}{2}. \tag{3.32}$$

This provides an explanation on why causal motifs start failing as test statistics when an SI contagion process approaches its end.

We extend the analysis by addressing the contributions of the individual submotifs within a motif order with the use of Mahalanobis distance. Mahalanobis distance is tested for

significance using the fact that its square follows a chi-squared distribution [38], where the degree of freedom is represented with the number of submotifs. As we already stated, it is shown in Figure 3.5 how the increase of the causal motif order from one-edge to two-edge shrinks the area of significant results for z-score. However, in Figures 3.5e and 3.5f, we can see that using Mahalanobis distance on the two-edge statistics not only expands the area, but also shifts it towards the larger percentages of default, thus making the result complementary to one-edge z-score result. As it is clear from the expression (3.1) that defines the Mahalanobis distance, its value is always greater or equal to the absolute value of the z-score of a given motif. However, due to the setting of our investigation, we are using the one-tailed z-score to test for significance, while the Mahalanobis distance is the absolute distance of the statistic from the null distribution and disregards the direction. Therefore, the Mahalanobis distance for one-edge statistic equals the absolute value of the two-sided z-score, which has to attain a higher value than the one-sided version for the same level of statistical significance. This downplays the beneficial effect that we get from engaging all the submotifs in the analysis. Nevertheless, Mahalanobis distance exploits the fact that, although the causal motifs converge to the same limit regardless of the endogeneity of the process, their submotifs do not. As expected from this observation, the Mahalanobis distance outperforms z-score in the later stages of the default process. As well as for the z-score, we see in Figs. 3.6b and 3.6d that the with increase of the network size, the Mahalanobis distance shows significant results for larger $\zeta$ values.

## Voter model

Since the properties of the SI process, combined with properties of the Erdős-Rényi network, lead motif test statistics from both the original and randomized process to converge to the same limit, we consider another type of process to simulate the endogenous component. In order to obtain a separation between the final fractions of the motif test statistics, we employ the voter model process, which scales the effect of the incoming defaulted neighbours by the total number of the incoming neighbours per vertex. Thus, for the same number of defaulted incoming neighbours $m_{in}$, the voter model will ascribe a larger probability of default to the vertex that has less incoming neighbours $k_{in}$ in total, unlike the SI process, which would not differ between these two cases. Due to this inho-

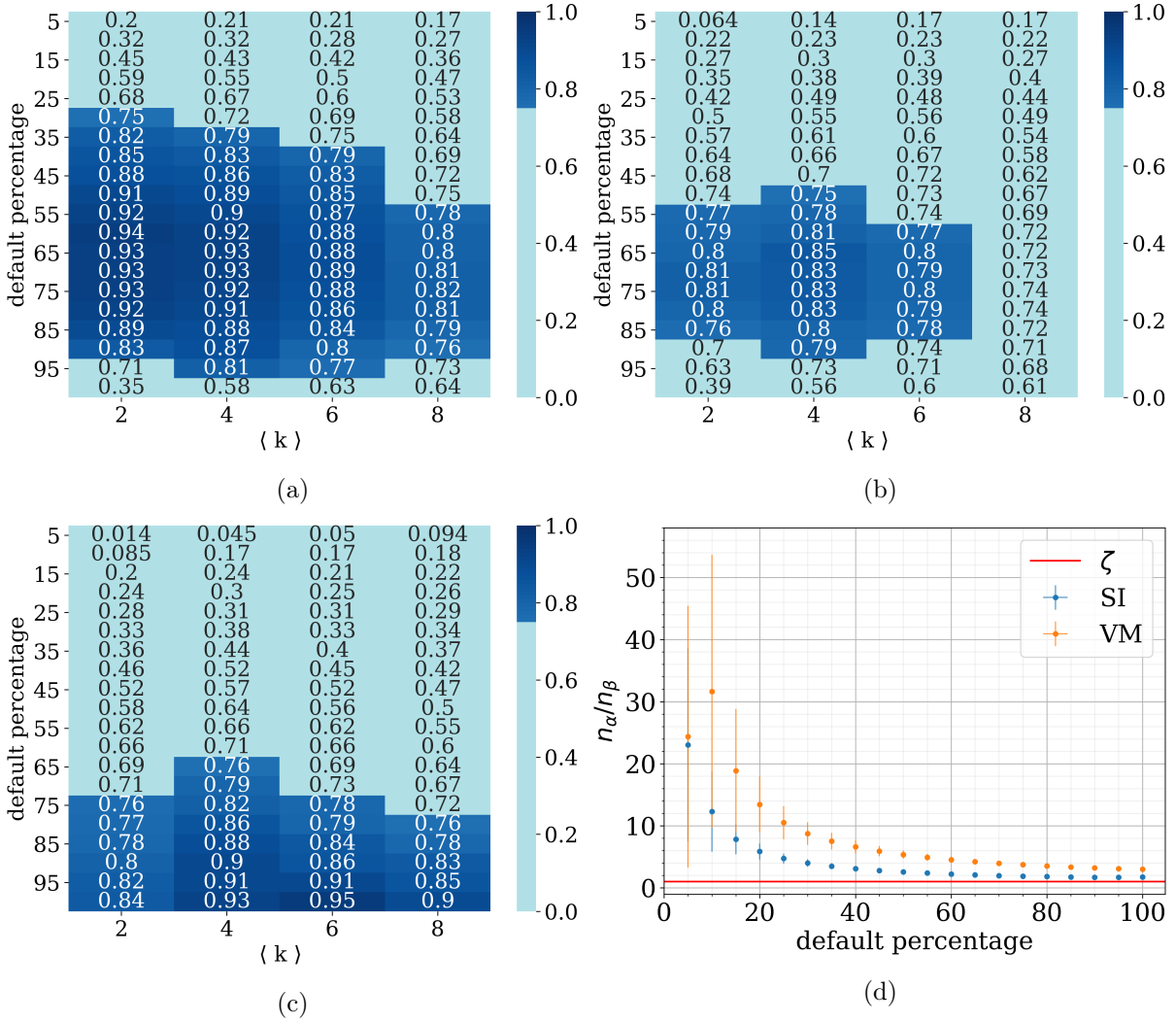



Figure 3.7: **VM process results.** In panel a) we present the one-tailed z-score results, depending on the percentage of network that defaulted (y-axis) and average network degree (x-axis), for a voter model process with $\zeta = 1$. The results that are significant with $p < 0.1$ are marked in dark blue. In panel b) we present in the same form the results of the one-tailed z-score of two-edge motifs, and in panel c) we present the results of the Mahalanobis distance of two-edge motifs. For these figures, the number of vertices equals $N = 1000$, on each of 100 realized networks, 10 processes are simulated and 100 random shuffles are created. In panel d) we show the ratio of numbers of exogenously and endogenously defaulted vertices, $n_\alpha$ and $n_\beta$, both for the SI and the voter model variants of the endogenous process, depending on different default percentages (with parameters $\langle k \rangle = 4$ and , $\zeta = 1$).

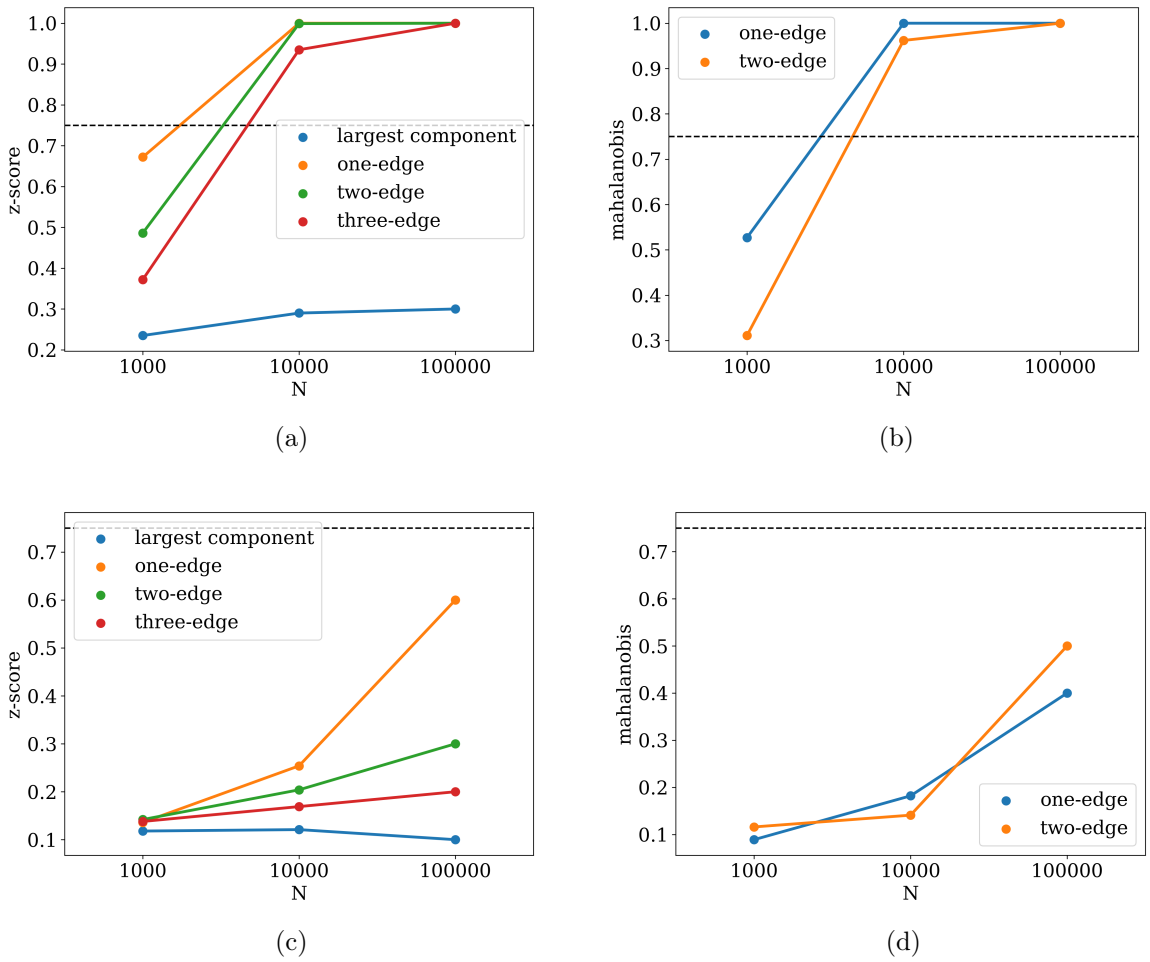



Figure 3.8: **VM process results - scaled.** We present the dependence of the percentage of significant ($p < 0.1$) results for a voter model process with $\langle k \rangle = 4$, fixed at 25% of the network default, on the size of the network. In panels a) and b) we show the one-tailed z-score and Mahalanobis distance results, respectively, for a voter model process with $\zeta = 1$. In panels c) and d) we show the results when the parameter of the SI process is increased to $\zeta = 10$. Horizontal lines are drawn at the limit in which 75% of simulations exhibit significant difference from the RRM-model.

mogeneity, we expect that the final fractions of the motif test statistics will not coincide and the robustness of our method will be increased at the end of the default process.

From the KS-results in Fig. 3.4, we can draw the conclusion that the test statistic with a distribution significantly different from the randomized one in the largest area of the parameter space is one-edge motif. We also note that, unlike for the SI process, for $\zeta = 100$ none of the test statistics show significant results, which we address in more detail further below. For values of $\zeta$ below 100, and for all mean degrees $\langle k \rangle$, one-edge motif dominates the early stages of the default process. In the final stages, all causal motif statistics show significant distinguishing, since the voter model does not drive both the original and the randomized process into the same limit of convergence.

Z-score results for one-edge and two-edge test statistics are shown in Fig. 3.7a and Fig. 3.7b. The takeaway is similar to the case with the SI endogenous component, as the significant part of the parameters space shrinks with the increase of the motif order. However, as expected from the KS results, the voter model results lose significance at a lower value of $\zeta$, that is, already at $\zeta > 1$. Although the three-edge statistic obtains the largest z-scores on average at the low percentages of default in highly endogenous processes, also the other causal motifs show significant results there. In the proximity of the detectability limit, it is the one-edge motif that is the most successful test statistic. As we already mentioned, the reason for this is the fact that the variance of the higher order motifs grows faster than the separation between the mean values.

The results for the Mahalanobis distance (Fig. 3.7c) show that, for $\zeta = 1$, it complements the one-edge z-score results on the $\langle k \rangle - \zeta$ parameter space. Just like with the SI process, one-edge z-score prevails for the intermediate stage of the default process, and Mahalanobis distance takes over in the final stages. For larger values of $\zeta$ we do not get significant results.

With the increase of the system, both z-score and Mahalanobis distance detect the propagation closer to its onset for $\zeta = 1$ (Fig. 3.8a and 3.8b). However, as we increase $\zeta$ to 10, neither of the measures are able to detect the endogenous process (Fig. 3.8c and 3.8d).

In Fig.3.7d, we provide some insight into why the voter model is harder to detect than the SI. We compare the temporal evolution of the ratio of exogenously and endogenously

defaulted vertices $n_\alpha/n_\beta$ for both types of process for the same parameters ($\langle k \rangle = 4$, $\zeta = 1$). It is clear from the plot, that for the same $\zeta$, less vertices default endogenously for the voter model than for the SI model, which explains the faster decline in the significance of detection with the increase of $\zeta$ for the voter model. Furthermore, the ratio $n_\alpha/n_\beta$ is in general different to the corresponding parameter $\zeta$, as it is only one of the parameters it depends on, which is outlined in the Equations (3.10) and (3.12). From these equations we also conclude that the voter model always has a higher ratio than the SI process. More specifically, in the plot 3.7d we see that, towards the end of the process, the ratio goes to approximately $n_\alpha/n_\beta \approx 2$ for the SI component, and to $n_\alpha/n_\beta \approx 3$ for the voter model component. In conclusion, we observe that for the example of $\zeta = 1$, the actual number of defaults that occur by contagion is 2-3 time lower than the number of defaults with an external cause.

## Defaults of Croatian companies

From the data presented in the Section 3.2, we filter only the companies that can both default from an endogenous cause and spread their default further into the network, since this structure is necessary for the cascade formation. After imposing the condition that all institutions have to be both creditors and debtors, we are left with 549 firms, connected together with 1198 edges (debts), which gives a mean degree of 4.36 (mean out and in degree are 2.18), and we denote that the maximal degree is 60. We proceed to apply our method to this network and show the results in Fig. 3.9.

In Figure 3.9a, we plot the one-edge statistic at the end of the process with a red vertical line, together with its null distribution created by RRM as a blue histogram. We show the temporal evolution of the default process using three-edge motifs in Figure 3.9b. Z-score and Mahalanobis distance are calculated on all causal motif statistics, and the results are depicted in Figs 3.9c,3.9d,3.9e and 3.9f. Possible default cascades are detected with our method in the early interval around when around $10 - 15\%$ companies defaulted, and in the later period, when the default process was at $70 - 85\%$ of completion.

From this analysis, we can conclude that a detailed analysis of balance sheets as well as interviews with executives of companies that defaulted could reveal that a default

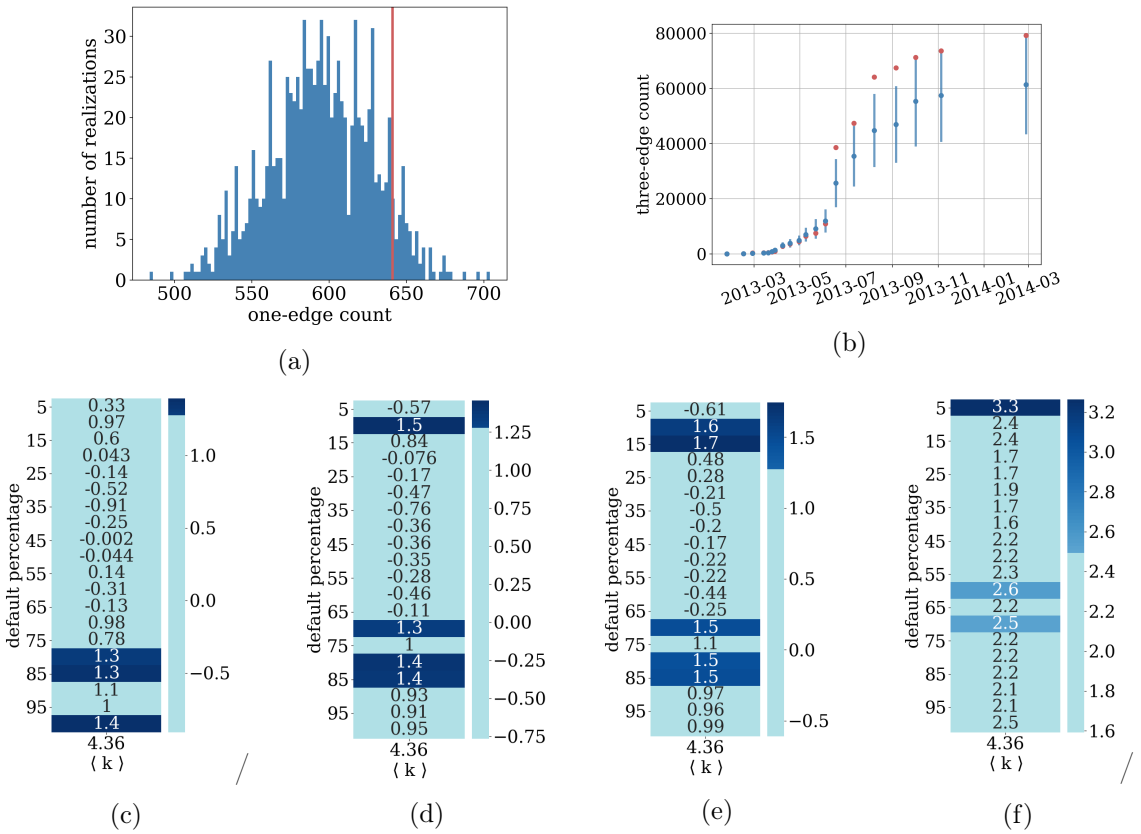



Figure 3.9: **Results for defaults of Croatian companies.** In figure a) we present with a red line the count of one-edge motifs in a completely defaulted network, compared to a histogram of the same statistic obtained from a RRM (1000 shuffled instances). In panel b) we present the count of three-edge motifs (red points) compared to RRM realizations represented with blue points (mean) and a blue line (standard deviation) for different different stages of the default process on the Croatian companies. In panels c) (one-edge) d) (two edge) and e) (three edge) we quantify the results with the one-tailed z-scores of the test statistics and show the statistically significant results ($p < 0.1$) in dark blue. In panel f) we use the Mahalanobis distance on the two-edge test statistic and show the results in the same manner.

contagion occurred in the periods marked as significant by our method.

# Chapter 4

# Modeling of contagion processes on financial networks

## 4.1 Introduction

The propagation of default cascades is one of the ways cascading behaviour can lead to an increase of systemic risk in a financial system. However, for the cascade to have a systemic effect, the state being transmitted does not necessarily need to be a default of an institution. When entering a credit contract, creditors are faced with the risk that the counterparty will not fulfill its obligations, so they take actions such as hedging or buying insurance to offset that risk. After being settled, these contracts can again be traded on a market, with their value on that market being affected by, amongst other factors, the risk of the debtor going into default. Thus, merely the increase in the probability that an institution will default, lowers the market value of all the credit relationships it entered as a debtor. In a network of creditors and debtors connected with debts, this re-evaluation of the debt values can trigger a cascade of further debt devaluation, which can in turn have a destabilizing effect on the entire system [35].

During the last financial crisis, two thirds of losses attributed to counterparty credit risk were due to such asset devaluations and not to actual defaults [39]. This difference between the risk-free value of a debt and the value that takes into account the counterparty credit risk is termed Credit Valuation Adjustment (CVA). In response to the crisis, in 2010 the

Basel II regulatory framework, a set of international banking regulations, was updated to Basel III, and CVA risk has become the core component of counterparty credit risk [40].

Formally, the Credit Valuation Adjustment is defined as the difference between the face value of a debt (value of the contract at the time of its conclusion) and its expected value at the time of maturity, when risks of the counterparty default are incorporated into the calculation. The standard formula for the calculation follows [36]:

$$CVA = (1 - R) \sum_{i=1}^{m} DF(t_i) EE(t_i) PD(t_{i-1}, t_i), \tag{4.1}$$

where $R$ is the amount of the claim that can be recovered and $(1 - R)$ is then loss given default. $DF(t_i)$ discounts the value of the future loss to its present value, $EE(t_i)$ is the expected credit exposure at the future time $t_i$, $PD(t_{i-1}, t_i)$ is the probability of the counterparty's default between time $t_{i-1}$ and $t_i$ and we assume that the credit exposure and default probability are independent. The first two factors in the product under summation are considered to be market risks, while the loss given default and probability of default are credit risks.

The part of the formula (4.1), that we will be focusing our research on, is the calculation of default probability. In practice, the frameworks used to calculate the default probability are called structural models and intensity models [41]. Structural models are based on the Merton's model [128], which treats the assets of the counterparty as a stochastic process, and considers the counterparty to be defaulted if the assets go below a certain level. The name comes from the fact that the capital structure of the firm is used to calculate its probability of default. A widely used structural model that is based on Merton's model is the KMV [129] model, that introduces the calculation of the probability of default based on historical patterns. On the other hand, intensity models disregard the internal mechanics of defaults and simply model default times using Poisson processes, either with a homogeneous [130], deterministic or stochastic (Cox process) rate of default. The rates of default are also called default intensities or hazard rates, and they can be derived through calibration to market instruments, such as credit default swaps.

However, all the aforementioned models take only the inherent risks of the counterparty into account, while completely neglecting the potential default contagion that can spread

from the counterparties of the counterparties. Therefore, our aim is to address the contribution of the entire network of credit exposures in the estimation of the default probability of the counterparty. We will analyze the structure of the default probability to quantify the contribution of the default contagion on the network, compared to the probability of the initial shocks that trigger the contagion.

We approach the problem stated above by simulating exogenous shocks and the contagion process that follows on interbank networks. Using the simulations, we are able to tackle realistic sizes of interbank networks, for any value of correlation, to obtain a distribution of default probabilities. In parallel with the simulations, we build on the results of [101, 91], that showed that this type of financial contagion models have an analytical representation in the form of the threshold model [19], to provide an analytical framework that includes initial correlated shocks. We obtain the results that show under which circumstances the contribution of the contagion phenomena to the total probability of default cannot be neglected. Therefore, we conclude that, in such cases, the contribution of the financial system should not be left out of the calculation of the CVA.

Our work continues on the strand of literature concerning the network valuation frameworks. The authors of [42] build on the seminal work of [46], where an existence of a clearing vector for a system of financial institutions constructed from balance sheets is proven, to analyse and distinguish exogenous shocks and the contagion that follows them. As the system of equations they use is exact, with a growing number of banks $N$ their computations start suffering from the curse of dimensionality. In [43], formulas for debt and equity valuation of firms in a financial network under comonotonic endowments are presented. The comonotonicity framework enables the authors to circumvent the curse of dimensionality, and, with a theoretically justified constraint, obtain analytical results for the CVA that include the contribution of the entire network. The work of [44] provides a network valuation framework for interbank claims. Under the assumption about limited information in the system, the authors obtain an iterative solution to the valuation problem, and show that their framework is a generalization of some widely used financial clearing models [46, 131, 132].

In our framework, we use the clearing model of Eisenberg and Noe [46], who were one of the first to model the financial dependencies as a directed network. In their work, the

problem of the interdependency of payments is turned into a fixed point problem, and the authors show that a unique clearing vector always exists. A more realistic setting with non-zero bankruptcy costs is considered in [132]. The author of [133] deals with the effect of cross-holdings of equities, while [134] introduce the time dynamics of the interbank liabilities in the Eisenberg and Noe model and then study the default risk dynamically.

## 4.2 Methodology

### 4.2.1 Outline of the framework

We start by outlining the framework that we use. We first describe the way we set up the financial system from the balance sheets, and the way external shocks are introduced in the system. We provide detail on the models that we chose for the financial shocks. We employ the Gaussian copula [45] to introduce correlations on the external shocks, due to its analytical tractability. Further on, we turn the balance sheet representation of the financial system into a network. For the network topology, we choose random k-regular graph since it fits the constraints we have for the balance sheets, while introducing randomness at the same time. In this system setup, we assume that we know all the information about the banks, unlike the authors of [44]. We model the dynamics of the shock propagation process through simulations, and obtain the results after the system reaches an equilibrium. Since we simulate on systems of realistic sizes ($N = 10000$), monitoring the state of each institution becomes unfeasible, so we represent the results as a distribution. This way we avoid the curse of dimensionality encountered in [42], where the authors solve the exact system of equations to obtain final states for individual institutions. Compared to the approach of the authors of [43], where they consider a comonotonic setting to reduce the number of all possible default scenarios for a system with $N$ banks, we obtain our results without having to impose restrictions on the shock correlations. In addition to the simulations, we make use of the mapping of this type of clearing processes to the threshold model [91, 101]. We combine the analytical form of the shock distributions with the threshold model to obtain the expected values of the defaulted institutions. The analytical approach enables us to study the system behaviour

in the infinite network limit. Finally, we precisely outline how the results of our research are reflected on the calculation of the Credit Valuation Adjustment.

### 4.2.2 Model of the financial system

We investigate a system of $N$ banks with a financial network of credit contracts. In addition to contracts among banks, we consider contracts with assets external to the system of banks. At time $t$ all the investments are made, and for each bank $i$ they are represented as a balance sheet consisting of interbank assets $A_i^b = \sum_j A_{ij}^b$, interbank liabilities $L_i^b = \sum_j L_{ij}^b = \sum_j {A_{ij}^b}^T$, external assets $A_i^e$, and external liabilities $L_i^e$. In general, the difference between the assets and liabilities, the equity $E_i(t)$, is defined at time $t$ as:

$$\begin{aligned} E_i(t) &= A_i^e(t) + A_i^b(t) - L_i^e(t) - L_i^b(t) \\ &= A_i^e(t) + \Lambda^b \sum_j B_{ij} x_j^b(t) - L_i^e - L_i^b. \end{aligned} \tag{4.2}$$

The decomposition of interbank assets is the following. The interbank leverage, $\Lambda_i^b = A_i^b/E_i$ is set to be the same for every bank: $\Lambda_i^b = \Lambda^b$. The matrix $B_{ij}$ represents the adjacency matrix of the underlying interbank network. The elements of $B_{ij}$ are normalized as fractions of the total interbank assets. At the initial time $t = 0$, unitary values of investments are equal to $x_j^b(t = 0) = 1$.

First, we explain the time dependence of the unitary values of interbank investments $x_j^b(t)$. We define the indicator variable $\chi_i(t)$, that indicates the state of default from the equity level: $\chi_j(t) = \Theta(-E_j(t))$, and can be either 0 or 1. The values of interbank investments depend on the counterparty's default state, and can, accordingly, take two possible values $x_i^b(\chi_i = 0) = 1$ and $x_i^b(\chi_i = 1) = R_i$. In the case the counterparty has not defaulted yet, the unitary value stays equal to 1 as initially set, and in the case of default, it gets reduced to $R_i \in [0, 1\rangle$. $R_i$ represents the recovery rate and accounts for the possibility that a bank that defaulted can still partially repay its debts.

For the purpose of this paper, we disregard the structure of the external asset investments, and combine the external factors in the Equation (4.2) into the quantity we define and

name the external equity, $E_i^e(t) := A_i^e(t) - L_i^e$. We have then as the balance sheet equation:

$$E_i(t) = E_i^e(t) + \Lambda^b \sum_j B_{ij} x_j^b(t) - L_i^b. \tag{4.3}$$

Since we set the total interbank assets and liabilities to be the same for every bank $i$, they also have to be equal to each other in amount, due to the relationship $L_{ij}^b = A_{ij}^{b\,T}$. We set the total equity of each bank $i$ to be equal to one at the initial time, $E_i(t=0) = 1$, by setting every external equity to be equal to one, $E_i^e(t=0) = 1$. The default process starts at time $T > 0$, when we shock the external equity $E_i^e$ of a bank $i$ with a shock $s_i < 0$, $E_i^e(T) = E_i^e(0)(1 + s_i)$. After the process has started, a bank $i$ is considered to default when the equity value goes below zero $E_i < 0$. The default propagation occurs immediately on the interbank network and produces the following fixed point equation:

$$E_i(T) = E_i^e(0)(1 + s_i) + \Lambda^b \sum_j B_{ij} x_j^b(T) - L_i^b. \tag{4.4}$$

The propagation follows the Eisenberg-Noe clearing process, which will be outlined in the subsection on the process simulation.

### 4.2.3 Model of the shocks and shock correlations

For a set of $N$ banks, we sample a shock vector $\vec{s}$ of dimension $N$ whose entry $s_i$ is a value of the shock that bank $i$ has received. We model the possible shock values in a discrete manner, so every shock $s_i$ can take one of the $n_\sigma$ possible values $\sigma_\mu, \mu = 1, ..., n_\sigma$, each with probability $p_\mu$. The value of $\sigma_1$ is set to be enough to cause a default of the institution it hits right away, that is, it has to be $\sigma_1 < -1$. The values of other shocks $\sigma_\mu, \mu = 2, ..., n_\sigma$ implement the levels of damage on the equities of institutions without causing their default, $\sigma_\mu \in [-1, 0\rangle$. The last value, $\sigma_{n_\sigma} = 0$ is set to be zero, as it represents the case when no damage happens to the equity. For the purpose of clearer demonstration, in this thesis we choose $n_\sigma$ to be 3, since that is the minimal choice that covers all three different types of shocks: immediate default, damage, and no damage at all. We group the affected nodes into compartment $\mu$, depending on the shock $\sigma_\mu$ they received, and denote the number of nodes in compartment $\mu$ as $N_\mu$.

We define the random variable of the shock value as $\mathcal{S}$, which realizes shock values $\sigma_\mu$ with the previously defined respective probabilities $p_\mu$. The cumulative distribution function of random variable $\mathcal{S}$ will have the form:

$$F_{\mathcal{S}}(\mathcal{S} = \sigma_x) = \sum_{\mu=1}^{x} p_\mu, \quad x = \sigma_1, \sigma_2, \ldots, \sigma_{n_\sigma}. \tag{4.5}$$

The sampling of values of the shock random variable $S$ is then performed using the inverse of this CDF on uniform random variables, $F_{\mathcal{S}}^{-1}(U)$.

A number of examples from the financial systems point to the existence of correlation between the external shocks. Companies that belong to the same industry, or the same geographic region will be affected by similar exogenous events. In some time periods, adverse general economic conditions can cause a higher correlation of defaults [41]. Therefore, to obtain a more realistic picture of the shock structure, we need to introduce correlations on them by correlating the uniform random variables. For that purpose, we use the copula [135, 45], a function that isolates the dependency structure of a multivariate distribution. In particular, out of the several families of copulas, we choose the Gaussian copula for this analysis, since it is analytically most tractable. In addition to that, for a single factor Gaussian copula [136], i.e. a copula that has all the correlations set to the same value, the Gaussian random variable can be decomposed into a common and an idiosyncratic random variable, that are mutually independent, which is a property we will use in the next Section 4.3. The Gaussian copula is represented as a cumulative distribution function of the multivariate Gaussian distribution $F_{\mathbf{Z}}$, with marginal distributions transformed to the uniform distribution using the uniform Gaussian quantile functions $F_{Z_i}$:

$$C_\Sigma^{Ga}(u_1, u_2, \ldots, u_N) = F_{\mathbf{Z}}(F_{Z_1}^{-1}(u_1), F_{Z_2}^{-1}(u_2), \ldots, F_{Z_N}^{-1}(u_N)). \tag{4.6}$$

The sampling of the uniform variables all correlated with $\rho$ then goes in the following way: we take an n-variate Gaussian probability density function $f_{\mathbf{Z}}$, with the covariance matrix $\Sigma$ that has diagonal terms equal to 1 and all off diagonal terms equal to the same chosen level of correlation, $\rho_{ij} = \rho$. We sample an n-dimensional vector $\vec{z} = (z_1, \ldots, z_N)$ from the multivariate Gaussian distribution, and transform it into an n-dimensional sample

$\vec{u} = (u_1, \ldots, u_N)$ from the uniform distribution, using the CDF of the Gaussian univariate distribution.

The sample of uniform random variables that we obtained is now correlated, and we can proceed in transforming it to a sample of shock values $\vec{s} = (s_1, \ldots, s_N)$, using the inverse of a discrete CDF $F_S^{-1}(U)$.

### 4.2.4 Network topology

In order to model the propagation of the shocks we defined in the previous section, we employ the random k-regular graph. This topology introduces randomness in the pairs of vertices, while keeping the vertex degree fixed. Under our constraints for the equity and the interbank leverage, randomness in the vertex degree would greatly extend the time needed for the simulations. In addition to reducing the length of the simulations, a simple network topology enables us to have more control over the parameters.

Equal interbank leverage is assigned to each bank $\Lambda_b$ and divided over its interbank assets. The condition that the equities and interbank leverage are uniform over all the banks ($E_i = E$, $\Lambda_i^b = \Lambda^b$) translates into all the assets and liabilities having the same value, $A_i^b = A^b$, $L_i^b = L^b$. Since the assets and liabilities can be represented as a network, from $L_{ij}^b = A_{ij}^{b\,T}$ we recover $A^b = L^b$.

On a random k-regular network, where all the nodes have the same degree $k$, these constraints lead to a solution where all the individual assets and liabilities are equal in value, $A_{ij}^b = L_{ij}^b = \frac{A^b}{k/2}$. In comparison, using an Erdős-Rényi graph would require iteratively looking for a solution for $A_{ij}^b$ and $L_{ij}^b$, which is not even guaranteed to exist for every instance of an Erdős-Rényi graph. Therefore, our choice of the network topology is justified.

### 4.2.5 Process simulation

In the first step, the external shock is applied to the external assets. If it triggers a bank to default, the default is further propagated with the Eisenberg-Noe [46] clearing algorithm

until the condition, that all the equity changes are less than 3% of the original equity, is fulfilled. In case the condition cannot be fulfilled, there is a cutoff set after $c$ steps. The recovery rate is set to $R_i = 0$.

Shocks are sampled 1000 times and imposed on the external assets of the nodes in a network. For every network degree $k$, 5 different instances of the random network are generated. On every network instance, 1000 process realizations $\xi$ are started using the sample of shocks, resulting in a total of $N_\xi = 5000$ process realizations. For each realization, a vector of default indicators for banks is obtained. The fraction of defaulted vertices within the realization $\xi$, $q(\xi)$, is calculated as the expected value of the default indicator vector. We then obtain a distribution of default fractions from entire sample with the same network degree $k$, and use its expected value $\langle q \rangle = \frac{1}{N_\xi} \sum_\xi q(\xi)$ as the probability of default.

### 4.2.6 Threshold model

It has been shown in [101, 91], that the financial contagion model, based on balance sheets and with variable external assets, is equivalent to the threshold model [19] already outlined in Chapter 2. We use the mean field approximation to obtain the solution for the threshold model. We obtain the distribution of the banks shocked by each value of shock $\sigma_\mu$ for both the correlated and uncorrelated case, and use it to calculate the expected fraction of defaults $\langle q \rangle$, in addition to obtaining $\langle q \rangle$ from the simulations. The analytical approach enables us to study the behaviour of the expected fraction of defaults in the infinite system size limit.

### 4.2.7 CVA computation

We reformulate the Equation (4.1) within our framework. We state once again that the CVA is defined as the difference between the face value of an asset $A_{ij}$ and its expected value at maturity. If we set the recovery rate $R$ to zero, the expectation is simply obtained as:

$$CVA(A_{ij}) = A_{ij} - [(1 - \langle q \rangle) \cdot A_{ij} + \langle q \rangle \cdot 0] = \langle q \rangle A_{ij}. \tag{4.7}$$

That is, for a fixed value of $A_{ij}$, the CVA depends only on the probability of default. Therefore, to show whether the network contribution needs to be taken into account in the CVA, we need to obtain the structure of the probability of default. We will thus compare the probability of default obtained from the simulations of the contagion as the expected fraction of defaulted banks $\langle q \rangle$ with the initial probability of default $p_0$.

## 4.3 Theoretical description

In this section we provide the analytical solution of the presented problem, using the mean-field approximation to the threshold model. We first derive the probability density functions for the non-correlated and correlated cases, and then use these results to proceed to the final formulas for the expected fraction of defaults.

### 4.3.1 Probability density functions for the distributions of shocks

#### Non-correlated case

For the case $\rho = 0$ (independent shocks), when there is no correlation on the shock levels, we can show that the distribution of the counts of banks $N_\mu$ hit with each shock $\sigma_\mu$ equals the multinomial probability distribution.

If we set the probabilities of the shock values $\vec{\sigma} = (\sigma_1, \sigma_2, \sigma_3)$, that result in equity values $\vec{\epsilon} = 1 - \vec{\sigma}$, to be $\vec{p} = (p_1, p_2, p_3)$, the cumulative distribution function of the random variable of the shock $\mathcal{S}$ will be, as in 4.5:

$$F_{\mathcal{S}}(x) = \begin{cases} p_1 \coloneqq p_I, & x = \sigma_1 \\ p_1 + p_2 \coloneqq p_{II}, & x = \sigma_2 \\ 1, & x = \sigma_3. \end{cases} \tag{4.8}$$

We use the inverse transform sampling to sample the shock random variable $\mathcal{S}$, i.e. we transform the uniform random variable $U$ using the inverse CDF $F_{\mathcal{S}}^{-1}(U) = \mathcal{S}$.

For $N$ banks, we write the total probability that $N_1$ banks end up being shocked by $\sigma_1$ with probability $p_1$, $N_2$ banks get shocked by $\sigma_2$ with probability $p_2$ and $N_3$ banks get shocked by $\sigma_3$ with probability $p_3$ as:

$$\begin{aligned}
&\mathcal{P}(\vec{N};N;\vec{p}) = \\
&= \binom{N}{N_1,N_2,N_3} P(U_1 \leq p_I,\ldots,p_I < U_i \leq p_{II},\ldots,U_N > p_{II}) \\
&= \binom{N}{N_1,N_2,N_3} P(U_1 \leq p_I,\ldots,p_I < U_i \leq p_{II},\ldots,U_N \leq 1-p_{II}) \\
&= \binom{N}{N_1,N_2,N_3} \left[C(p_I,\ldots,p_{II},\ldots,1-p_{II}) - C(p_I,\ldots,p_I,\ldots,1-p_{II})\right] \\
&= \binom{N}{N_1,N_2,N_3} \left[p_I\cdots p_{II}\cdots(1-p_{II}) - p_I\cdots p_I\cdots(1-p_{II})\right] \\
&= \binom{N}{N_1,N_2,N_3} p_I\cdots(p_{II}-p_I)\cdots(1-p_{II}) \\
&= \binom{N}{N_1,N_2,N_3} \underbrace{p_1\cdots p_1}_{N_1}\cdot\underbrace{p_2\cdots p_2}_{N_2}\cdot\underbrace{p_3\cdots p_3}_{N_3} \\
&= \binom{N}{N_1,N_2,N_3} \quad p_1^{N_1}\cdot p_2^{N_2}\cdot p_3^{N_3}. \qquad (4.9)
\end{aligned}$$

We expressed the probability using the uniform variables that we use for sampling. The multinomial coefficient counts the number of ways we can get the same counts, $N_1$, $N_2$ and $N_3$, from a set of $N$ banks. In the third row we utilize the definition of the copula, and switch to the fourth row using the fact that the independence of all the variables in the copula leads to the explicit form of the copula to be simply a product of all those variables [45]. From there on, with a bit of algebra, and by returning to the originally defined probabilities $p_1, p_2, p_3$, we recover a multinomial probability mass function.

### Correlated case

In the previous paragraph, we obtained the distribution of the numbers of banks $\vec{N} = (N_1, N_2, N_3)$ affected by the set of shocks $\vec{\sigma} = (\sigma_1, \sigma_2, \sigma_3)$ that occurred with probabilities $\vec{p} = (p_1, p_2, p_3)$, in the case when all the shocks were independent. Now we will look more closely into the case when correlation between the shocks is introduced. We wish

to obtain the probability mass function $\mathcal{P}(N_1, N_2, N_3; N; p_1, p_2, p_3; \rho)$ for the numbers of banks $\vec{N} = (N_1, N_2, N_3)$ in compartments $\mu = 1, 2, 3$ when correlation $\rho$ is present.

In the model, we introduced the correlations between shocks that reduce equity values by employing the Gaussian copula $C_{\Sigma}^{Ga}(\mathbf{u})$ to correlate our uniform variables $U_i$ before we used them for sampling shocks with Equation (4.5). However, the approach used to derive the distribution in the uncorrelated case (4.9) relies on the independence of the uniform variables, so we need to take a different perspective. Therefore, instead of using the uniform random variables $U_i$ as in (4.9), we base this derivation on Gaussian random variables, $Z_i \sim \mathcal{N}(0, 1)$, and use the property of a single factor copula that enables the decomposition of the correlated variables $Z_i$ into the sum of independent Gaussian variables $X \sim \mathcal{N}(0, \rho)$ and $Y_i \sim \mathcal{N}(0, 1 - \rho)$. The random variable $X$ is the part common to all the sampled variables, while $Y_i$ is the idiosyncratic part, specific for each bank $i$.

$$Z_i = X + Y_i, \quad Z_i \sim \mathcal{N}(0, 1), \quad X \sim \mathcal{N}(0, \rho), \quad Y_i \sim \mathcal{N}(0, 1 - \rho). \tag{4.10}$$

Next, we need to transform the subintervals for sampling with limits $p_I, p_{II}$, from the $[0, 1]$ segment, into the real line $\langle -\infty, \infty \rangle$ with limits $z_I, z_{II}$. Respective limits need to divide both distributions into the same percentiles:

$$\begin{aligned} p_1 &= p_I = F_Z(z_I), \\ p_2 &= p_{II} - p_I = F_Z(z_{II}) - F_Z(z_I), \\ p_3 &= 1 - p_{II} = 1 - F_Z(z_{II}). \end{aligned} \tag{4.11}$$

From this we can get $z_I, z_{II}$ easily:

$$\begin{aligned} z_I &= F_Z^{-1}(p_I), \\ z_{II} &= F_Z^{-1}(p_{II}). \end{aligned} \tag{4.12}$$

We start the derivation of the probability mass function in the same way as in (4.9), by counting all the combinations of shocks on banks that can occur for the same counts $N_1$,

$N_2$ and $N_3$:

$$\begin{aligned}\mathcal{P}(\vec{N}; N; \vec{p}) &= \\ &= \binom{N}{N_1, N_2, N_3} P(\underbrace{Z_1 \leq z_I, \ldots}_{N_1}, \underbrace{z_I < Z_i \leq z_{II}, \ldots}_{N_2}, \underbrace{Z_{i+k+1} > z_{II}, \ldots, Z_N > z_{II}}_{N_3}).\end{aligned} \tag{4.13}$$

Then we look separately into the joint probability and utilize the decomposition of the random variables we defined earlier, $Z_i = X + Y_i$, so that we can express probability distribution function variables $\gamma_i$ as a sum $\gamma_i = \alpha + \beta_i$. The corresponding probability distribution functions for the new variables are the univariate PDF $f_X$ and the multivariate PDF $f_{\mathbf{Y}}$. We first express the probability that Gaussian variables will take the values that correspond to bank counts $N_1$, $N_2$ and $N_3$ as the multidimensional integral of the the Gaussian multivariate PDF $f_{\mathbf{Z}}$, and use the abbreviation $d\vec{\gamma} = d\gamma_1 \cdots d\gamma_i \cdots d\gamma_N$:

$$\begin{aligned}&P(Z_1 \leq z_I, \ldots, z_I < Z_i \leq z_{II}, \ldots, Z_N > z_{II}) = \\ &= \int_{-\infty}^{z_I} \cdots \int_{z_I}^{z_{II}} \cdots \int_{z_{II}}^{\infty} f_{\mathbf{Z}}(\gamma_1, \ldots, \gamma_i, \ldots, \gamma_N)\, d\vec{\gamma} \\ &= \int_{-\infty}^{z_I} \cdots \int_{z_I}^{z_{II}} \cdots \int_{z_{II}}^{\infty} f_{\mathbf{Z}}(\alpha + \beta_1, \ldots, \alpha + \beta_i, \ldots, \alpha + \beta_N)\, d\vec{\gamma} \\ &= \int_{-\infty}^{z_I} \cdots \int_{z_I}^{z_{II}} \cdots \int_{z_{II}}^{\infty} \left( \int_{-\infty}^{\infty} f_X(\alpha) f_{\mathbf{Y}}(\gamma_1 - \alpha \ldots, \gamma_i - \alpha, \ldots, \gamma_N - \alpha)\, d\alpha \right) d\vec{\gamma} \\ &= \int_{-\infty}^{z_I} \cdots \int_{z_I}^{z_{II}} \cdots \int_{z_{II}}^{\infty} \left( \int_{-\infty}^{\infty} f_X(\alpha) f_Y(\gamma_1 - \alpha) \cdots f_Y(\gamma_i - \alpha) \cdots f_Y(\gamma_N - \alpha)\, d\alpha \right) d\vec{\gamma} \\ &= \int_{-\infty}^{\infty} f_X(\alpha) \left( \int_{-\infty}^{z_I} f_Y(\gamma_1 - \alpha)\, d\gamma_1 \right) \cdots \left( \int_{z_I}^{z_{II}} f_Y(\gamma_i - \alpha)\, d\gamma_i \right) \cdots \left( \int_{z_{II}}^{\infty} f_Y(\gamma_N - \alpha)\, d\gamma_N \right) d\alpha \\ &= \int_{-\infty}^{\infty} f_X(\alpha) \left( \int_{-\infty}^{z_I} f_Y(\gamma_1 - \alpha)\, d\gamma_1 \right) \cdots \left( \int_{-\infty}^{z_{II}} f_Y(\gamma_i - \alpha)\, d\gamma_i - \int_{-\infty}^{z_I} f_Y(\gamma_i - \alpha)\, d\gamma_i \right) \cdots \\ &\cdots \left( 1 - \int_{-\infty}^{z_{II}} f_Y(\gamma_N - \alpha)\, d\gamma_N \right) d\alpha \\ &= \int_{-\infty}^{\infty} f_X(\alpha) F_Y(z_I - \alpha) \cdots \big( F_Y(z_{II} - \alpha) - F_Y(z_I - \alpha) \big) \cdots \big( 1 - F_Y(z_{II} - \alpha) \big)\, d\alpha \\ &= \int_{-\infty}^{\infty} f_X(\alpha) (F_Y(z_I - \alpha))^{N_1} \cdot (F_Y(z_{II} - \alpha) - F_Y(z_I - \alpha))^{N_2} \cdot (1 - F_Y(z_{II} - \alpha))^{N_3}\, d\alpha.\end{aligned} \tag{4.14}$$

In the third row, we replace the $\gamma_i$ variables with the sum $\alpha + \beta_i$, and proceed to express $f_{\mathbf{Z}}$ as a convolution of $f_X$ and $f_{\mathbf{Y}}$. In the fifth row we use the independence of the idiosyncratic

uniform random variables $Y_i$ to separate the $f_{\mathbf{Y}}$ PDF into univariate $f_{Y_i} = f_Y$ PDFs. After integrating all the integrals depending on $\gamma_i$ in the seventh row, we get a product of $N$ CDFs of the random variable $Y$ that depend on the limits $z_I$ and $z_{II}$ according to the class $\mu$ they belong to ($N_\mu$ of them in each class $\mu$). The final result is an integral, going over the real axis, of the product of a PDF of the random variable $X$ and each CDF raised to its respective power $N_\mu$. To make sense of the expression, we define new shock probabilities that depend on the variable $\alpha$:

$$\begin{aligned}
\pi_1(\alpha) &= F_Y(z_I - \alpha) = F_Y(F_Z^{-1}(p_I) - \alpha), \\
\pi_2(\alpha) &= F_Y(z_{II} - \alpha) - F_Y(z_I - \alpha) = F_Y(F_Z^{-1}(p_{II}) - \alpha) - F_Y(F_Z^{-1}(p_I) - \alpha), \\
\pi_3(\alpha) &= 1 - F_Y(z_{II} - \alpha) = 1 - F_Y(F_Z^{-1}(p_{II}) - \alpha).
\end{aligned} \tag{4.15}$$

Finally, the complete expression for the probability mass function is structured as an integral of the multinomial distribution with varying probabilities weighted by the Gaussian distribution of the common random variable $X$:

$$\mathcal{P}(\vec{N}; N; \vec{p}) = \int_{-\infty}^{\infty} f_X(\alpha) \binom{N}{N_1, N_2, N_3} \pi_1(\alpha)^{N_1} \cdot \pi_2(\alpha)^{N_2} \cdot \pi_3(\alpha)^{N_3} \, d\alpha. \tag{4.16}$$

### 4.3.2 Threshold model - a mean field approximation

The threshold model [19] is specified in the Section 2.2 about dynamical processes. The mapping between the financial contagion an the threshold model has been shown by [91, 101]. We will shortly explain the reasoning behind this. Since our default condition is $E_i < 0$, bank $i$ defaults either by being hit with the initial default shock $p_1$, or by having enough neighbours default, according to the level of equity $\epsilon_\mu$ it has after the shock . We can easily calculate the required number of neighbours $m_\mu$ that need to default to transmit it:

$$m_\mu = \left\lceil \frac{k}{2} \frac{\epsilon_\mu}{\Lambda_b} \right\rceil \leq \frac{k}{2}. \tag{4.17}$$

The number of neighbours $m_\mu$ then represents the threshold from the threshold model.

In our setup, the external shocks, sampled from $\vec{\sigma} = (\sigma_1, \sigma_2, \sigma_3)$ and delivered to the system, are then simply reflected in the initial distribution of different thresholds of equity

values $\vec{\epsilon} = (\epsilon_1, \epsilon_2, \epsilon_3)$. We group the nodes depending on the type of the shock they received into compartments $\vec{N} = (N_1, N_2, N_3)$. The underlying network is assumed to be a random regular network, in which every node has the same degree $k$, which is the sum of in- and out-degrees $k = k_{in} + k_{out}, k_{in} = k_{out}$.

We can write an iterative equation for the expected fraction of default $q_n$ by using the mean field approximation. Assuming the lack of correlations between the neighbours allows us to use the binomial distribution to calculate the probability that at least $m_\mu$ neighbours defaulted for a central node that was shocked with $\sigma_\mu$. Summing over all possibilities for shocks $\mu$, we get for $q_n$:

$$q_n = \sum_\mu \frac{N_\mu}{N} \sum_{m=m_\mu}^{k/2} \binom{k/2}{m} q_{n-1}^m (1 - q_{n-1})^{k/2-m} = f(q_{n-1}). \tag{4.18}$$

The solution to this equation can be obtained by looking for a fixed point for $q^* = f(q^*) = \lim_{n\to\infty} f(q_n)$.

For any possible realization of the vector of nodes $\vec{N}$ resulting from some shock, if we know the degree of the network $k$, the interbank leverage $\Lambda_b$ and the equity sizes after the shock, $\vec{\epsilon} = \vec{1} - \vec{\sigma}$, the fixed point solution provides us with the expected fraction of defaulted vertices:

$$q(\vec{N}) = \sum_\mu \Pi_\mu(N_\mu) \left[1 - F_{CDF}^{BINOM}\left(m_\mu, \frac{k}{2}, q(\vec{N})\right)\right], \quad \Pi_\mu = \frac{N_\mu}{N}, \quad m_\mu = \left\lceil \frac{k}{2} \frac{\epsilon_\mu}{\Lambda_b} \right\rceil \leq \frac{k}{2}. \tag{4.19}$$

To calculate the expected fraction of defaulted vertices $\langle q \rangle$ over all the possible shock realizations, i.e. represented in this calculation with all the possible values of the vector $\vec{N}$, we need the probability distribution $\mathcal{P}(\vec{N}; N; \vec{p})$ of $\vec{N}$, given the individual initial shock probabilities $\vec{p} = (p_1, p_2, p_3)$:

$$\langle q \rangle = \sum_{N_1, N_2, N_3} \mathcal{P}(\vec{N}; N; \vec{p}) \cdot q(\vec{N}). \tag{4.20}$$

In the previous subsection we derived the distribution for the non-correlated case 4.9,

which we plug into the previous expression to get:

$$\langle q \rangle = \sum_{N_1,N_2,N_3} \binom{N}{N_1, N_2, N_3} \quad p_1^{N_1} \cdot p_2^{N_2} \cdot p_3^{N_3} \cdot q(\vec{N}). \tag{4.21}$$

We show the comparison of the analytical calculations with the simulation results in Fig. 4.1a.

For the correlated case, we derived the probability mass function (4.16), so the expected fraction of defaults when correlations on shocks exist follows:

$$\langle q \rangle = \int_{-\infty}^{\infty} f_{X(\rho)}(\alpha) \left( \sum_{N_1,N_2,N_3} \binom{N}{N_1, N_2, N_3} \pi_1(\alpha)^{N_1} \pi_2(\alpha)^{N_2} \pi_3(\alpha)^{N_3} \right) \times q(\vec{N})\, d\alpha, \tag{4.22}$$

with $\pi_1(\alpha), \pi_2(\alpha), \pi_3(\alpha)$ as:

$$\pi_1(\alpha) \coloneqq F_Y(z_I - \alpha) = F_Y(F_Z^{-1}(p_I) - \alpha), \tag{4.23}$$

$$\begin{aligned} \pi_2(\alpha) &\coloneqq F_Y(z_{II} - \alpha) - F_Y(z_I - \alpha) = F_Y(F_Z^{-1}(p_{II}) - \alpha) - F_Y(F_Z^{-1}(p_I) - \alpha), \\ \pi_3(\alpha) &\coloneqq 1 - F_Y(z_{II} - \alpha) = 1 - F_Y(F_Z^{-1}(p_{II}) - \alpha). \end{aligned} \tag{4.24}$$

and $p_I \coloneqq p_1$ and $p_{II} \coloneqq p_1 + p_2$, $z_I \coloneqq F_Z^{-1}(p_I)$, $z_{II} \coloneqq F_Z^{-1}(p_{II})$.

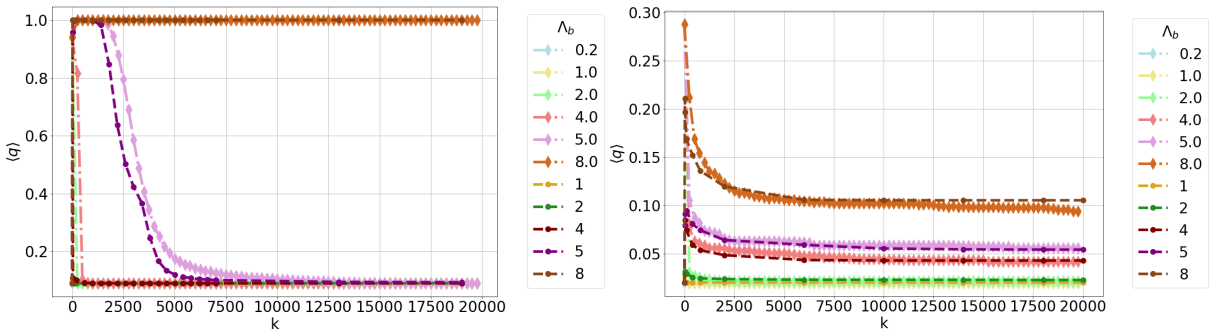
To shorten the time necessary for the numerical evaluation of Eq. (4.26), we approximate the varying multinomial distribution by using only the expected values, $\mathbb{E}[N_\mu(\alpha)] = N\pi_\mu(\alpha)$ which then replace the factor $\Pi_\mu = N_\mu/N$ with $\mathbb{E}[\Pi_\mu] = \mathbb{E}[N_\mu(\alpha)]/N = \pi_\mu(\alpha)$ in the Equation (4.19):

$$q(\alpha) = \sum_\mu \pi_\mu(\alpha) \left[ 1 - F_{CDF}^{BINOM}\left( m_\mu, \frac{k}{2}, q(\alpha) \right) \right], m_\mu = \left\lceil \frac{k}{2} \frac{\epsilon_\mu}{\Lambda_b} \right\rceil \leq \frac{k}{2}. \tag{4.25}$$

The expected fraction of defaults then becomes:

$$\langle q \rangle = \int_{-\infty}^{\infty} f_X(\alpha) \cdot q(\alpha)\, d\alpha. \tag{4.26}$$

The comparison of the result of Eqn. (4.26) and the simulation results is depicted in Fig. 4.1b.




(a) $N = 10000$, $\rho = 0$ (b) $N = 10000$, $\rho = 0.3$

Figure 4.1: **Comparison of the theoretical calculation and the simulation results.** Both panels show results on the network of size $N = 10000$. Results obtained from simulations are presented in a lighter shade, and the analytical results are depicted with the darker shade of the respective color. Panel **a)** shows the expected fractions of default depending on the network degree for the shock with parameters $\vec{p} = (0.09, 0.083, 0.827)$, $\vec{\sigma} = (-1.1, -0.5, 0.0)$ and no correlation between shocks, $\rho = 0$. Panel **b)** shows the expected fractions of default depending on the network degree for the shock with parameters $\vec{p} = (0.02, 0.09, 0.89)$, $\vec{\sigma} = (-1.1, -0.5, 0.0)$ and with correlation $\rho = 0.3$ between shocks.

### 4.3.3 Infinite network limit

We study the behaviour of the default probability in the limit of an infinitely large network ($N \to \infty$) that is complete ($k \to \infty$). To start, we consider the Equation (4.19) in that light. We wish to obtain the form of the cumulative distribution function of the binomial distribution in the limit of $k \to \infty$. We observe that the standard deviation grows with the square root of the in-degree $k/2$, $\sigma = \sqrt{k/2 \cdot q(1-q)}$, while the sample space grows linearly, $\sim k/2$. Therefore, if we look instead at a defaulted fraction of total in-neighbours $k/2$, its standard deviation behaves like $\sqrt{\frac{q(1-q)}{k/2}} \underset{k\to\infty}{\longrightarrow} 0$. As the network degree goes to infinity, the probability distribution of the fraction of defaulted neighbours gets more and more localized around its expected value. Thus, for a large enough $k/2$, we can approximate the CDF of a binomial function with the Heaviside function $\Theta(x)$, and we use the half maximum convention, $\Theta(0) = 1/2$:

$$\left[1 - F_{CDF}^{BINOM}\left(m_\mu, \frac{k}{2}, q\right)\right] \approx \Theta\left(\frac{k}{2} \cdot q - \left\lceil \frac{k}{2}\frac{\epsilon_\mu}{\Lambda_b} \right\rceil\right) \approx \Theta\left(\frac{k}{2}\left(q - \frac{\epsilon_\mu}{\Lambda_b}\right)\right). \tag{4.27}$$

For the factor $\Pi_\mu$ in the Equation (4.19), setting $N \to \infty$ with similar arguments as above can be shown to reflect in $\Pi_\mu \to E(\Pi_\mu) = E(N_\mu)/N = Np_\mu/N = p_\mu$. Thus, the limiting

value of the average default probability in the case of an infinite network, is simply:

$$\langle q_{lim} \rangle = p_1 + \sum_{\mu=2} p_\mu \cdot \prod_{\phi=2}^{\mu} \Theta \left( \sum_{\nu=1}^{\phi-1} p_\nu - \frac{\epsilon_\phi}{\Lambda_b} \right). \tag{4.28}$$

In case the correlation is present, we insert the approximation (4.27) into Equation (4.26):

$$\begin{aligned} \langle q_{lim} \rangle &= \int_{-\infty}^{\infty} f_X(\alpha) \cdot \left[ \pi_1(\alpha) + \sum_{\mu=2} \pi_\mu(\alpha) \cdot \prod_{\phi=2}^{\mu} \Theta \left( \sum_{\nu=1}^{\phi-1} \pi_\nu(\alpha) - \frac{\epsilon_\phi}{\Lambda_b} \right) \right] d\alpha \\ &= \int_{-\infty}^{\infty} f_X(\alpha)\pi_1(\alpha)\, d\alpha + \sum_{\mu=2} \int_{-\infty}^{\infty} f_X(\alpha) \cdot \pi_\mu(\alpha) \cdot \prod_{\phi=2}^{\mu} \Theta \left( \sum_{\nu=1}^{\phi-1} \pi_\nu(\alpha) - \frac{\epsilon_\phi}{\Lambda_b} \right) d\alpha \\ &= \int_{-\infty}^{\infty} f_X(\alpha)\pi_1(\alpha)\, d\alpha + \sum_{\mu=2} \int_{-\infty}^{l_\mu} f_X(\alpha) \cdot \pi_\mu(\alpha)\, d\alpha. \end{aligned} \tag{4.29}$$

The finite upper limits $l_\mu$ in the integrals result from the theta function. For the case of $\mu = 2$ we get from the condition:

$$\begin{aligned} \pi_1(\alpha) &> \frac{\epsilon_2}{\Lambda_b}, \\ F_Y(F_Z^{-1}(p_I) - \alpha) &> \frac{\epsilon_2}{\Lambda_b}, \\ \alpha &< F_Z^{-1}(p_I) - F_Y^{-1}\left(\frac{\epsilon_2}{\Lambda_b}\right) \coloneqq l_{II}. \end{aligned} \tag{4.30}$$

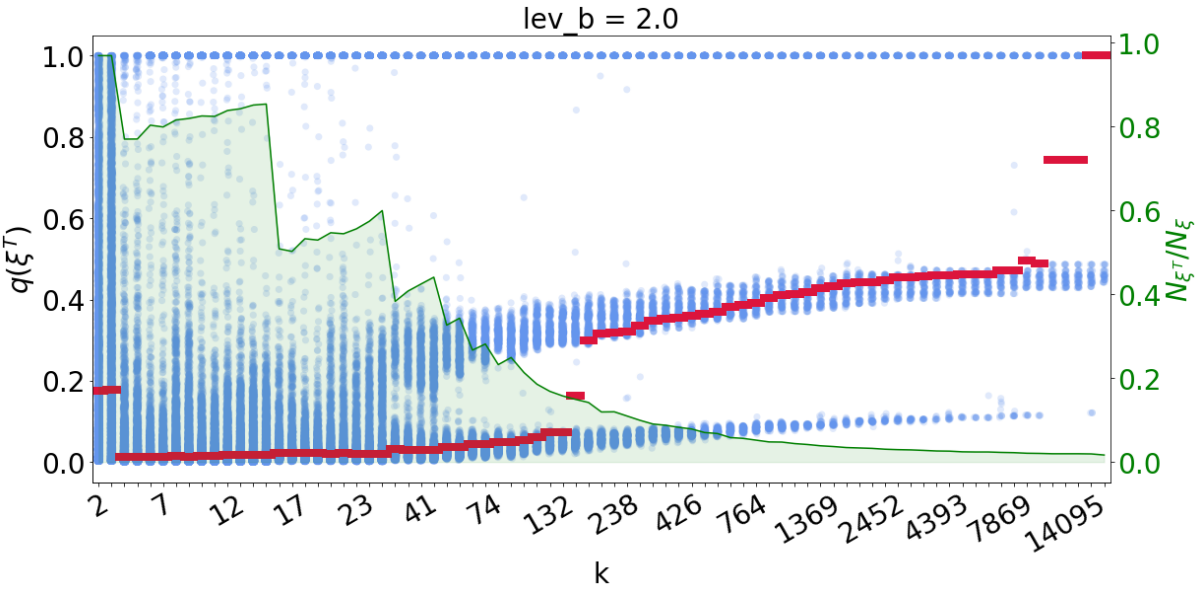
The term for $\mu = 3$, consists of a product of two theta functions, with the first resulting in the limit $l_{II}$ and the second being the solution of $\pi_2(\alpha) + \pi_3(\alpha) > \epsilon_3/\Lambda_b$:

$$\alpha < F_Z^{-1}(p_{II}) - F_Y^{-1}\left(\frac{\epsilon_3}{\Lambda_b}\right) \coloneqq l_{III}. \tag{4.31}$$

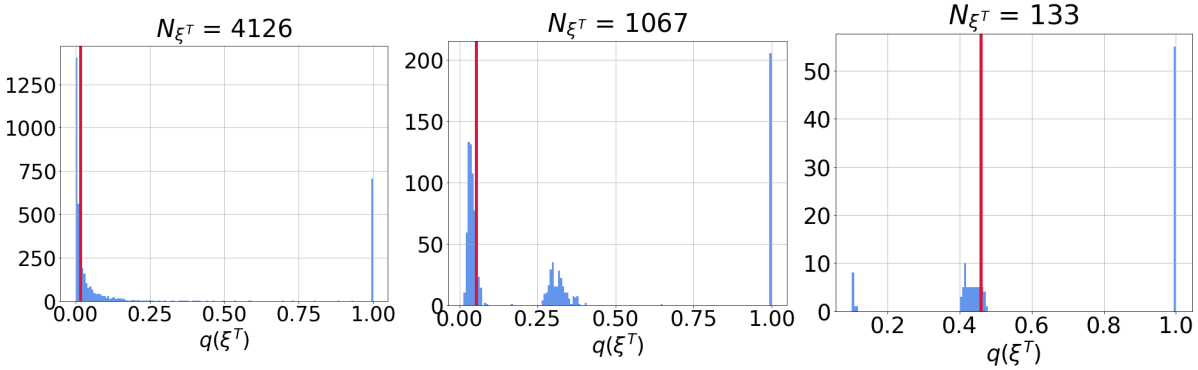
Due to the product of theta functions, the upper limit in the integral is determined by the smaller of the two values $l_{II}$ and $l_{III}$. The total expected fraction of default in the limit of an infinite network for $\mu = 1, 2, 3$ equals:

$$\begin{aligned} \langle q_{lim} \rangle &= \int_{-\infty}^{\infty} f_X(\alpha)\pi_1(\alpha)\, d\alpha + \int_{-\infty}^{l_{II}} f_X(\alpha)\pi_2(\alpha)\, d\alpha + \int_{-\infty}^{\min(l_{II}, l_{III})} f_X(\alpha)\pi_3(\alpha)\, d\alpha, \\ l_{II} &= F_Z^{-1}(p_I) - F_Y^{-1}\left(\frac{\epsilon_2}{\Lambda_b}\right), \quad l_{III} = F_Z^{-1}(p_{II}) - F_Y^{-1}\left(\frac{\epsilon_3}{\Lambda_b}\right). \end{aligned} \tag{4.32}$$

## 4.4 Results




(a) $\Lambda_b = 2$




(b) $k = 10$

(c) $k = 105$

(d) $k = 4393$

Figure 4.2: **Analysis of the simulation results.** We present results from a default process with parameters $\vec{p} = (0.02, 0.09, 0.89), \vec{\sigma} = (-1.1, -0.75, 0.0)$, $\Lambda_b = 2$, $\rho = 0.3$. In panel **a)** we present the fraction $N_{\xi^T}/N_\xi$ of the simulation realizations that contained network propagation with the green line and the shaded area. On top of that, we plot in blue the one-dimensional histograms containing individual fractions of defaults in realizations with propagation, for each network degree $k$. The red line represents the median value of those fractions of default, given propagation happened. In panel **b)**, **c)** and **d)**, for network degrees $k = 10, 105, 4393$ we depict the one-dimensional histograms from the panel **a)** in two dimensions, stating the total number of processes with propagation in the titles.

We simulated the process described in the Framework section on a network of $N = 10000$ banks. The network size is chosen to be the largest possible, in accordance with the computational constraints, to reduce the finite size effects. The default fraction is obtained as a result of each of the $N_\xi = 5000$ simulations performed for each network degree $k$.

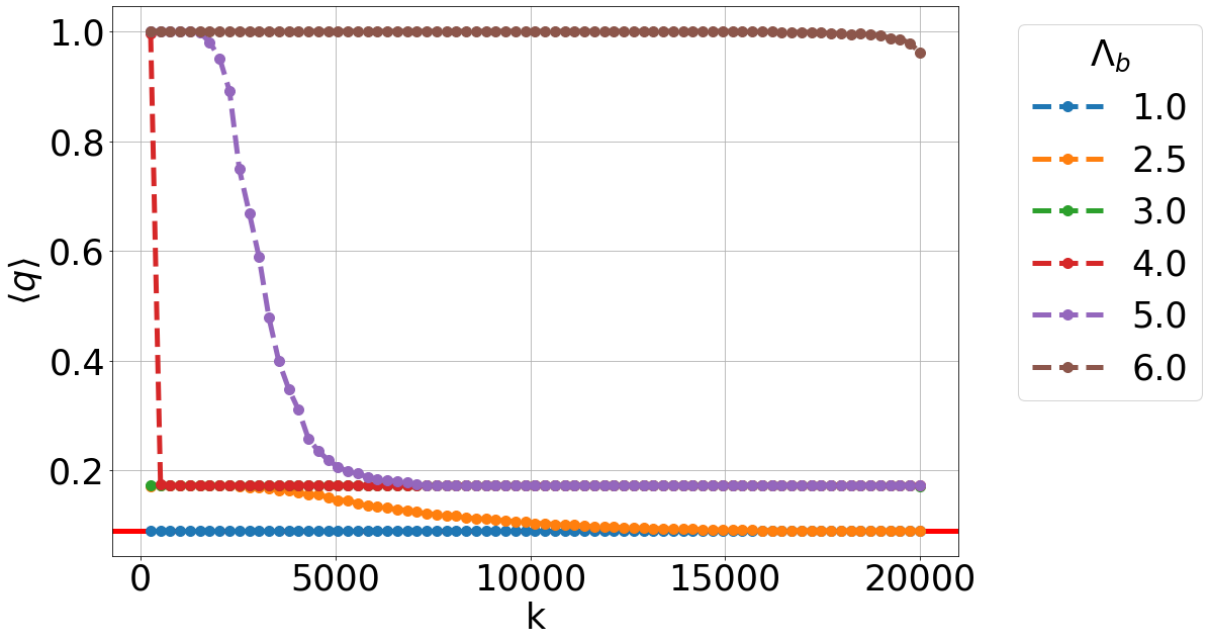



Figure 4.3: **Convergence limits.** We show in this figure that, for the set of parameters ($\vec{p} = (0.09, 0.083, 0.827)$, $\vec{\sigma} = (-1.1, -0.75, 0)$) and with no correlation, the system converges to three possible limits after diversification, depending on the interbank leverage $\Lambda_b$. On y-axis we plot the expected fraction of default, and on the x-axis the network degree, ranging from $k = 252$ to the complete network $k = 19998$.

The final result for each $k$ is a distribution of the default fractions.

We first present the result for $\vec{p} = (0.02, 0.09, 0.89), \vec{\sigma} = (-1.1, -0.75, 0.0)$, $\Lambda_b = 2$ and $\rho = 0.3$. To describe the shape of the results in Fig. 4.2a, we use only the subset of the results in which the propagation did happen on the network. We stress that in a sample of $N_\xi$ processes that start on a network, network propagation of defaults does not happen for every one of them, and we mark with $N_{\xi^T}$ the number of realizations $\xi^T$ in which it does happen. With the green shaded area we mark the fraction of the process realizations in which the propagation on the network was triggered, $N_{\xi^T}/N_\xi$. We denote the fraction of default for such process realizations $\xi^T$ with $q(\xi^T)$. The blue dots represent a one dimensional histogram of the individual default fractions resulting from propagation. The red line shows the median value of those fractions. The median is shown instead of the mean, since it is a more appropriate statistic due to the multimodal shape of the data. We stress that, due to the shape of the data, a large variance is present when we take the expected value of the sample in the later analysis [137]. In the Figures 4.2b,4.2c,4.2d, we show the histograms and the median value in two dimensions for three different values of network degree. We can conclude from this figure that, in

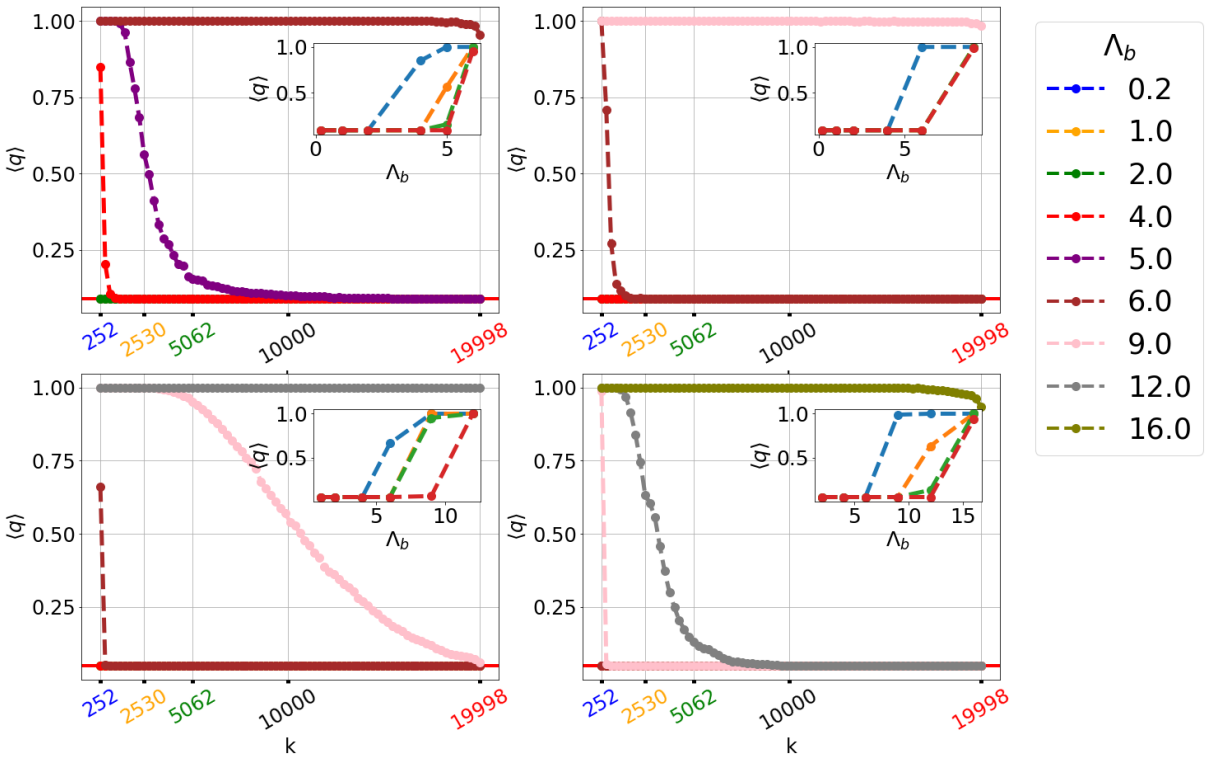



Figure 4.4: **Switching between sub- and supercriticality.** On x-axis the degree of the network starts from 252 and goes up to a complete network, $k = 19998$, and on the y-axis we plot the expected fraction of default, $\langle q \rangle$. In the first row the initial probability of default is fixed, $\vec{p}_{up} = (0.09, 0.083, 0.827)$ and we change the shock sizes from left $\vec{\sigma}_{left} = (-1.1, -0.5, 0)$ to right $\vec{\sigma}_{right} = (-1.1, -0.25, 0)$. In the second row we use a different initial probability of default $\vec{p}_{down} = (0.05, 0.086, 0.864)$ for the same pair shock sizes. Interbank leverage takes values from the range in $\Lambda_b = \{0.2, 1, 2, 4, 5, 6, 9, 12, 16\}$ and it is shown that with the decrease of the initial probability, or the shock size, the system behaviour stays the same if we increase $\Lambda_b$ appropriately. In the panel insets, we plot the dependence of $\langle q \rangle$ on the interbank leverage $\Lambda_b$, for the chosen values of $k$ that are marked on the x-axis of the plot containing the inset.

the case with correlation, diversification lowers the probability that propagation occurs in the system. Nevertheless, the probability of default conditioned on the occurrence of propagation increases significantly. If we apply this conclusion to a realistic setting, it is clear that these unlikely events, that can create significant damage to the system, will be underrepresented in the historical data, which will almost certainly lead to an underestimation of the expected fraction of defaults.

Further on, using all the results obtained from the simulations, we show that, for uncorrelated external shocks, the diversified systems end up in either a subcritical or a supercritical regime, i.e. either with diversification the expected fraction of default gets reduced to some limit, or the entire system defaults every time regardless of it. The regime the system chooses depends on the process and network parameters.

Next, when correlation is introduced to the external shocks, we demonstrate that, in a previously well diversified system, the expected fraction of default takes up a finite value regardless of the diversification.

Finally, we demonstrate that the effects do not depend on the size of the system $N$.

The process and network parameter ranges were chosen from the empirical values of the real world interbank and asset markets. The results outline the importance of the network contribution in the structure of the expected fraction of default, and accordingly in the calculation of the CVA 4.1

### 4.4.1 Limiting values of convergence for the probability of defaults

Depending on the parameters of the network (interbank leverage) and the process (shock sizes and probabilities), and if there are no correlations present, after the complete diversification, the probability of default will converge to one of the limiting values, as previously shown in the Equation (4.28). The maximal number of the possible limiting values of default probability is equal to the number of different shock values that we use in our model. For example, for the shock $\vec{p} = (p_1, p_2, p_3)$, the limiting values are equal to $p_1$, $p_1 + p_2$ and $p_1 + p_2 + p_3 = 1$. The interpretation of this phenomenon lies in the

fact that the initial shock with 3 possible values splits the $N$ banks into compartments with 3 different levels of vulnerability $(N_1, N_2, N_3)$. The compartment $N_1$ always defaults because the shock $p_1$ causes default right away. In addition to that, the compartment $N_2$ has the equity decreased to $\epsilon_2$, and will default if a fraction $\epsilon_2/\Lambda_b$ of its interbank assets is lost. If the probability of initial default $p_1$ is equal or larger than the fraction $\epsilon_2/\Lambda_b$ (in a large and complete network), every bank in $N_2$ will have a fraction of initially defaulted borrowers (i.e. $p_1$) large enough to drive it into default. Therefore, for some combinations of the shock probabilities and sizes, and the interbank leverage, beside the default of the compartment $N_1$, we can have either the compartment $N_2$, or both $N_2$ and $N_3$ (the whole system) going into default as shown in Fig. 4.3.

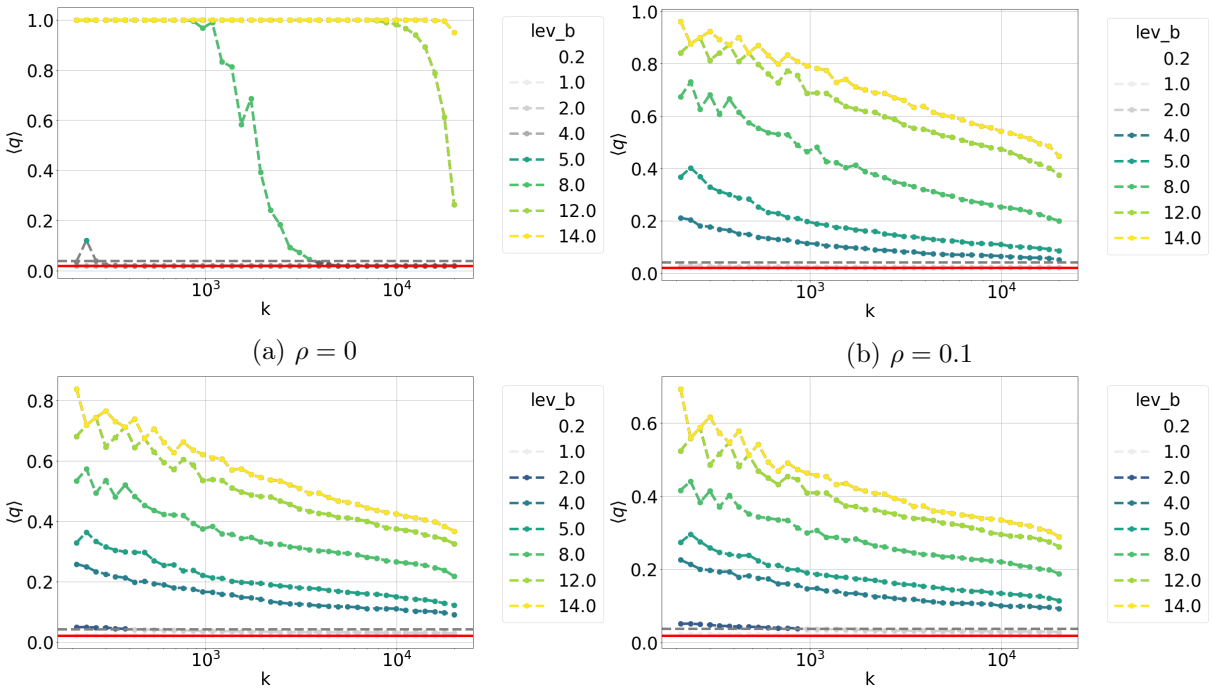
### 4.4.2 Subcritical and supercritical regime resulting from uncorrelated shocks

Depending on the parameters of the network (interbank leverage) and the process, diversification can lead the probability of default either to decrease to one of the limits mentioned above, or to remain at the highest level, equal to one. We refer to these two modes of behaviour as subcritical and supercritical regimes. For a fixed probability and size of the shocks, increasing the interbank leverage leads from the subcritical into the supercritical mode. The same effect, switching from a subcritical into supercritical regime, is obtained with increasing the probability of initial default $p_1$, and increasing the size of the shock $\epsilon_2$.

In Fig. 4.4 we present this effect on four panels by varying the initial probability in the vertical direction ($\vec{p}_{up} = (0.09, 0.083, 0.827)$ to $\vec{p}_{down} = (0.05, 0.086, 0.864)$) and the shock size in the horizontal direction ($\vec{\sigma}_{left} = (-1.1, -0.5, 0)$ to $\vec{\sigma}_{right} = (-1.1, -0.25, 0)$). We see that an appropriate choice of the interbank leverage $\Lambda_b$ recovers both regimes regardless of the intial probability of the shocks and their sizes.

Therefore, we see that maximal diversification can mitigate the effect of the network component in the expected fraction of default $\langle q \rangle$ only if the interbank leverage is in the subcritical range for the given parameters of the shock (probability and size).

### 4.4.3 Correlated shocks cause a non-vanishing probability of default




(a) $\rho = 0$ (b) $\rho = 0.1$

(c) $\rho = 0.2$ (d) $\rho = 0.3$

Figure 4.5: **Probability of defaults with correlation between shocks**. The degree of the nodes $\langle k \rangle$ starts from 211 and goes up to 19998 (complete network), for the number of nodes $N = 10000$. The plots do not show a smaller $\langle k \rangle$ than 211, as the transitional effects in that range are due to the choice of the network, and not relevant to real systems. At time $t = 0$ the shock values are $\vec{\sigma} = (-1.1, -0.75, 0)$ , with probabilities $\vec{p} = (0.02, 0.09, 0.89)$, respectively. We vary the interbank leverage $\Lambda_b = \{0.2, 1, 2, 4, 5, 8, 12, 14\}$. The red horizontal line represents the initial probability of default (without taking into account the network of liabilities). The grey dashed line represents double the values of the initial probability of default. Subfigures **a)-d)** show different correlation coefficients $\rho = \{0, 0.1, 0.2, 0.3\}$. Values where the introduced correlation increases the probability of default by 100% or more are shown in colour, values below that are shown in greyscale.

We model the correlation on the equity levels using the Gaussian copula, as described in the Section 4.2.

In a real market, banks invest in multiple assets outside the banking network. Their portfolios can overlap, and the values of the assets, and thus their default probabilities, can be correlated. However, simulating such a system for realistic parameters can be computationally too expensive. We argue that reducing the structure of the external assets and modeling the shocks on what we defined to be the external equity in Equation (4.3) will not change the qualitative form of the results, since correlations on the external

equity are a non-decreasing function of the correlations on the external assets.

For the shock levels that we use in simulations, we choose parameters realistic for markets, values $\vec{\sigma} = (-1.1, -0.75, 0)$, with the probabilities $\vec{p} = (0.02, 0.09, 0.89)$ and we simulate these shocks with 4 different levels of correlation $\rho = \{0, 0.1, 0.2, 0.3\}$.

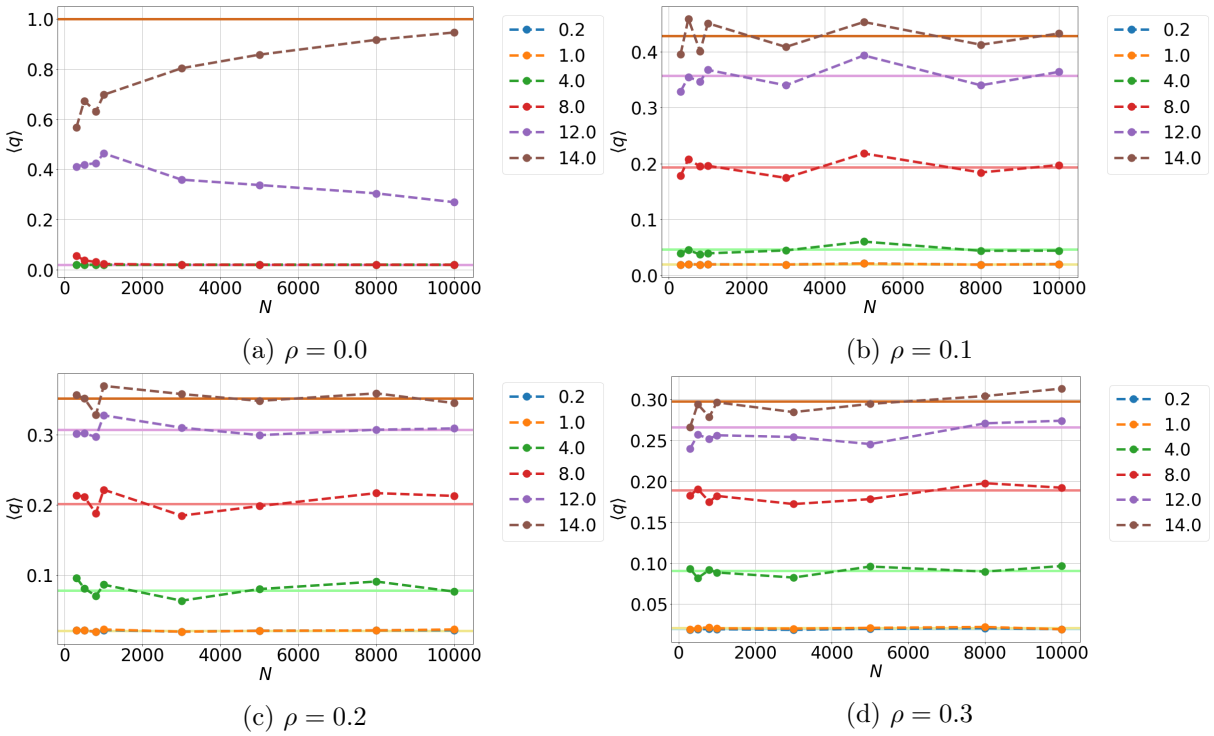
The results of the simulations are presented in Figure 4.5. We plot the dependence of expected fraction of default on the network degree. The horizontal red full line represents the fraction of banks defaulted by the initial external shocks. The horizontal grey dashed line represents the level of probability that is double than the fraction of banks defaulted from initial shocks. The lines with dots, showing the expected fraction of default for different interbank leverages, are presented in two color schemes, depending on whether their value exceeds double the initial expected fraction (the grey dashed line) or not. In the area where default fraction coming from the network is smaller than double the initial default fraction they are presented in greyscale, and in the area where it is larger than double, they are in colour. Figures 4.5b, 4.5c and 4.5d show how the introduction of correlation on the equity values raises the level of the expected fraction of network default for leverage values that were previously well diversified, lines $\Lambda_b = \{0.2, 1, 2, 4, 5\}$, on Fig. 4.5a.

On the other hand, if the interbank leverage $\Lambda_b$ was high enough to maximize the expected fraction of default without the correlations ($\Lambda_b = 14$ on Fig. 4.5a), the introduction of correlations will reduce it. In the case with no correlation, the fraction of initially defaulted banks comes from the binomial distribution, and for the large systems deviates very little from the expected value, which is in this case equal to $p_1$. If the expected fraction of initially defaulted banks is enough to default the entire system, and, as we concluded, the initial fraction of defaults does not vary much between realizations, the system will default in every realization and the expected total fraction of defaults will be equal to 1.

However, if the probabilities of initial defaults of banks are correlated, although the expected value of the fraction of defaults stays the same, the fractions of defaults are no longer drawn from a binomial, but from a distribution in the Equation (4.16), which has positive skewness. Therefore, realizations where the fractions of initially defaulted banks are very close to zero and thus not large enough to trigger the default of the entire system become very likely. Taking an expectation over the sample of all realizations necessarily

gives a value less than 1.

### 4.4.4 Finite probability of default is not system-size dependent




(a) $\rho = 0.0$ (b) $\rho = 0.1$

(c) $\rho = 0.2$ (d) $\rho = 0.3$

Figure 4.6: **Probability of default and the system scale** Probability of default is plotted on the y-axis for maximally diversified (complete) networks with the numbers of banks $N = \{300, 500, 800, 1000, 3000, 5000, 8000, 10000\}$ on the x-axis. At time $t = 0$ the shock values are $\vec{\sigma} = (-1.1, -0.75, 0)$ , with probabilities $\vec{p} = (0.02, 0.09, 0.89)$, respectively. In different subplots **(a)-d)**) we vary the correlation coefficient from 0 to 0.3. The subplots indicate that the correlation-induced probability of default does not depend on the system size, which is further supported with horizontal lines that represent the analytical result for the limit of an infinite network.

We show how the expected fraction of default scales with the size of the system with correlation (Fig. 4.6b, 4.6c, 4.6d) and without correlation (Fig. 4.6a). We plot expected fractions of default on complete networks for network sizes $N = \{300, 500, 800, 1000, 3000, 5000, 8000, 10000\}$. The shock values and probabilities used in simulations are $\vec{\sigma} = (-1.1, -0.75, 0)$ , and $\vec{p} = (0.02, 0.09, 0.89)$. We vary the correlation levels $\rho = \{0, 0.1, 0.2, 0.3\}$ and use the set of interbank leverages $\Lambda_b = \{0.2, 1, 4, 8, 12, 14\}$. The simulation results are plotted as points connected with dashed lines. In addition to that we depict theoretical results in the limit of an infinite network (Equations (4.28) and (4.32)) with horizontal lines.

It is clear from Fig. 4.6a that the system size plays a role when the correlations are absent. The theoretical limits show two limits for convergence, $\langle q_{lim} \rangle = \{0.02, 1.0\}$, which correspond to the aforementioned subcritical and supercritical regime. We can see the dashed lined approaching either of the two limits as the system size increases.

The introduction of correlations changes the picture, and in Fig. 4.6b, 4.6c, 4.6d we see that the expected fraction of default remains approximately constant for all the system sizes $N$, and that it corresponds to the theoretical limits. We can conclude that, in the case when the external shocks are correlated, the network contribution to the structure of the expected fraction of default cannot be neglected. Thus, a calculation of CVA without the information about the network contagion would be deficient.

# Chapter 5

# Discussion

## 5.1 Conclusion

In this thesis we studied the network effects in systemic risk propagation from two different perspectives, one being the detection of the default propagation from available data and the other the quantification of its effect on the value of debt.

The methodology that infers the presence of the endogenous propagation from data was developed in the Chapter 3. Our objective was, given a set of data, to be able to distinguish when it is necessary to include the endogenous, network structure into the description of the process, and when it is enough to model the dynamics as driven only by an external field.

For that purpose, we developed a methodology that can determine, for the data that can be turned into a temporal network, whether the endogenous propagation of default occurred or not. We required the results to be statistically significant to conclude that the contagion was present, and extensively tested the limits of the methodology.

We generated synthetic data to test our methodology, and it indicated the presence of an endogenous component in the simulated process with the ratio of the exogenous and endogenous rates up to $\zeta \approx 1$, for small networks with 1000 vertices. First, we used KS-test to verify if there was a significant difference between the original and randomized distributions of test statistic counts. Then we compared individual processes with their

respective RRM ensembles using z-score and Mahalanobis distance to test for the presence of endogenous spreading. For the cases of $\zeta$ that are borderline significant in distinguishing endogeneity, the success of the method depends also on the percentage of defaults, average degree and size of the network. The shape of the border between the part of the phase space of parameters where we are able to detect propagation and the part where we are not, is complex and differs for z-score and Mahalanobis distance. With the increase in network size, the range of $\zeta$ where the detection is possible also gets extended. When the endogenous component is modeled like an SI process, Mahalanobis distance performs either better or as good as z-score, whereas when it is modeled like a voter model process, z-score outperforms Mahalanobis distance. In addition to the simulated data, we provided an extension of a theoretical framework of Approximate master equations, which we then used to further our understanding of the temporal evolution of the test statistics. The equations that we derived provide a good match between the calculated expected counts of the causal motifs and the simulated results.

Ultimately, we applied our methodology to determine whether there was any endogenous spreading in the pre-bankruptcy settlement data of Croatian companies. From the results we obtained, we concluded that the defaults (pre-bankruptcy settlements) *probably* propagated endogenously. This conclusion is in agreement with the previous research [111] that was conducted on the same dataset, but employed more data, such as the values of debts, assets of companies etc. However, unlike other types of cascade processes, such as meme propagation, where it can be easily checked who copied the meme from whom, determining the ground truth in the default cascades is way more complicated. Determining the exact cause of default of a company involves checking the numerous court filings, financial reports of the companies and interviews with employees. Our method can be used to warn the regulators that the default process has obtained an endogenous component and point to the parts of the network of companies where such an investigation should be undertaken to prove the presence of the endogenous cascade.

In conclusion, we fulfilled our goal to distinguish potential endogenous spreading even when using minimal possible information.

In Chapter 4 we studied the structure of the probability of default when the institutions are connected with credit exposures, and the effect this has on the Credit Valuation Ad-

justment. We aimed to show that the propagation of default on the network of credit exposures can in some cases significantly increase the probability of default of the institutions, and that it should not be left out from the calculation of the CVA.

We approached this problem by setting up a simulation of the default propagation process that includes all the information about the credit network. For the external shocks that were not correlated, after diversification, the system occupied either the subcritical state, where the probability of default was reduced, or the supercritical state, which implied the default of the entire system. With the introduction of correlated external shocks, modeled with the Gaussian copula, we observed both qualitative and quantitative differences in the behaviour of the system, compared to the uncorrelated case. Systems that were previously well diversified obtain an increase in the probability of default to more than double of the value that disregards the network propagation. Although in the case where correlation exists, the risk cannot be diversified away, there is a beneficial effect on the systems that were previously in the supercritical state, since their probability of default gets reduced. This result is tested against the system size increase, and it is shown not to be a finite size effect. The simulated results are backed with an analytical treatment of the problem, obtained by combining the threshold model with the derived distributions of shocks. The advantage of the analytical approach is that we can take the limit of an infinite network, and observe the asymptotic results. Finally, a connection with the calculation of the Credit Valuation Adjustment is made, and the deficiency of the methodology that omits the effect of the default propagation is pointed out.

To conclude, the aim to find the contribution of the credit network in the structure of the probability of default was met, and its implication on the calculation of the CVA was highlighted.

## 5.2 Future work

The work on the detection of the economic cascades can be naturally extended by releasing out constraint about the minimal information. We could add the amounts of debt as weights on the edges, as it is quite straightforward that the size of debt should affect the possible propagation of defaults. Further on, we could introduce a limit on the time

that passes between the default of a debtor and its creditor's default for the default to be classified as potentially causal. It is reasonable to assume that the probability of a causal relationship between the debtor's and the creditor's default decays with time. The presented work can be thought of as a starting point for a systematic inclusion of more and more information in the detection of endogenous propagation, up to a point of purely data driven methods.

Another direction in the future work could be to use more realistic network classes, for example correlated or scale-free networks [138]. A study of the influence of the network topology on the robustness of the detection methodology could have interesting results, since its effects on contagion processes are well known [138, 139, 33].

We have seen in our results, that sometimes it is impossible to detect an endogenous process if the "noise" of the exogenous process is too strong. We could expect an existence of a phase transition between the detectable and undetectable phases of the contagion, similarly to the detectability limit of communities [140]. The understanding of the exogenous - endogenous interplay of the interacting particle systems on complex networks would be improved with a theoretical study of this limit.

The part of our research concerning the contribution of the network propagation to the probability of default can be easily extended by introducing additional external shocks that occur later in the process. They can be allowed to have a positive sign as well, to model the reactions of the regulators trying to mitigate the default cascade. In addition to having shocks that have a time dependence, a generalization could be made to shocks with a continuous distribution, which could be implemented both in the simulations and theoretically.

Another line of research could take the direction of making the structure of the external assets more complex. A different type of correlation on the shocks could arise if we allow the banks to have overlapping portfolios, and we could study the relationship between the two types of correlation.

As in the work we did on the detection of cascades, this work could too be expanded by considering networks that model the financial network more realistically. It has been empirically found that formation mechanisms lead the financial network to take up a core-

periphery structure, with the core forming a full clique and the peripheral banks being connected only to the core [47]. Such a structure can be modeled using the stochastic block model [141], which is a generative model that can reproduce the different edge densities in the core and from the core towards periphery.

We also could go in the direction of working on the actual implementation of the network contribution into the CVA policy framework. Since the real data needed to recreate the entire network of credit exposures is available to the regulators, we would not need to be concerned with modelling the underlying network. However, a more realistic model of the external shock would be necessary, in line with the propositions mentioned above. Also, the clearing framework would have to allow for more generality, such as possible non-zero bankruptcy costs [132] and recovery rates. We could calibrate the shock distributions on the historical shocks, and study how likely the actual defaults that occurred on the network would have been predicted by our model.

# Bibliography


[1] P. W. Anderson, "More Is Different", in: *Science* **177** (1972), pp. 393–396, DOI: 10.1126/science.177.4047.393.

[2] D. Sornette, *Critical Phenomena in Natural Sciences*, Berlin, Heidelberg: Springer, 2000, DOI: 10.1007/978-3-662-04174-1_15.

[3] P. Erdős and A. Rényi, "On the Evolution of Random Graphs", in: *Publication of the mathematical institute of the Hungarian Academy of Sciences*, 1960, pp. 17–61, DOI: 10.1515/9781400841356.38.

[4] E. N. Gilbert, "Random Graphs", in: *The Annals of Mathematical Statistics* **30** (1959), pp. 1141–1144, DOI: 10.1214/aoms/1177706098.

[5] R. Albert and A.-L. Barabási, "Statistical mechanics of complex networks", in: *Rev. Mod. Phys.* **74** (2002), pp. 47–97, DOI: 10.1103/RevModPhys.74.47.

[6] D. J. Watts and S. H. Strogatz, "Collective dynamics of 'small-world'networks", in: *Nature* **393** (1998), pp. 440–442, DOI: 10.1038/30918.

[7] A.-L. Barabasi and R. Albert, "Emergence of Scaling in Random Networks", in: *Science* **286** (1999), pp. 509–512, DOI: 10.1126/science.286.5439.509.

[8] T. Squartini, I. Van Lelyveld, and D. Garlaschelli, "Early-warning signals of topological collapse in interbank networks", in: *Scientific Reports* **3** (2013), pp. 1–9, DOI: 10.1038/srep03357.

[9] S. Battiston et al., "Complexity theory and financial regulation", in: *Science* **351** (2016), pp. 818–819, DOI: 10.1126/science.aad0299.

[10] F. Schweitzer et al., "Economic Networks: The New Challenges", in: *Science* **325** (2009), pp. 422–425, DOI: 10.1126/science.1173644.

[11] G. Caldarelli and A. Vespignani, *Large scale structure and dynamics of complex networks: from information technology to finance and natural science*, Complex Systems and Interdisciplinary Science, Singapore: World Scientific, 2007, DOI: 10.1142/6455.

[12] R. Milo et al., "Network Motifs: Simple Building Blocks of Complex Networks", in: *Science* **298** (2002), pp. 824–827, DOI: 10.1126/science.298.5594.824.

[13] A. Barrat, M. Barthélemy, and A. Vespignani, *Dynamical Processes on Complex Networks*, Cambridge University Press, 2008, DOI: 10.1017/CBO9780511791383.

[14] S. Grosskinsky, *Interacting stochastic processes*, Lecture notes, Warwick, 2009, URL: https://warwick.ac.uk/fac/sci/maths/people/staff/stefan_grosskinsky/ma4h3/notes_ips_10.pdf.

[15] T. Liggett, *Interacting Particle Systems*, Classics in mathematics, Springer New York, 1985, DOI: 10.1007/b138374.

[16] J. M. Swart, *A Course in Interacting Particle Systems*, 2022, arXiv: 1703.10007.

[17] G. Grimmett, *Probability on Graphs: Random Processes on Graphs and Lattices*, Institute of Mathematical Statistics Textbooks, Cambridge University Press, 2010, DOI: 10.1017/9781108528986.

[18] R. A. Holley and T. M. Liggett, "Ergodic Theorems for Weakly Interacting Infinite Systems and the Voter Model", in: *The Annals of Probability* **3** (1975), pp. 643–663, DOI: 10.1214/aop/1176996306.

[19] D. J. Watts, "A simple model of global cascades on random networks", in: *Proceedings of the National Academy of Sciences* **99** (2002), pp. 5766–5771, DOI: 10.1073/pnas.082090499.

[20] M. F. Weber and E. Frey, "Master equations and the theory of stochastic path integrals", in: *Reports on Progress in Physics* **80**, 046601 (2017), DOI: 10.1088/1361-6633/aa5ae2.

[21] J. P. Gleeson, "High-Accuracy Approximation of Binary-State Dynamics on Networks", in: *Phys. Rev. Lett.* **107**, 068701 (2011), DOI: 10.1103/PhysRevLett.107.068701.

[22] J. P. Gleeson, "Binary-State Dynamics on Complex Networks: Pair Approximation and Beyond", in: *Phys. Rev. X* **3**, 021004 (2013), DOI: 10.1103/PhysRevX.3.021004.

[23] F. Deschâtres and D. Sornette, "Dynamics of book sales: Endogenous versus exogenous shocks in complex networks", in: *Phys. Rev. E* **72**, 016112 (2005), DOI: 10.1103/PhysRevE.72.016112.

[24] S. A. Myers, C. Zhu, and J. Leskovec, "Information Diffusion and External Influence in Networks", in: *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '12, Beijing, China: Association for Computing Machinery, 2012, pp. 33–41, DOI: 10.1145/2339530.2339540.

[25] M. Piškorec et al., "Modeling Peer and External Influence in Online Social Networks: Case of 2013 Referendum in Croatia", in: *Complex Networks & Their Applications VI*, Cham: Springer International Publishing, 2018, pp. 1015–1027, DOI: 10.1007/978-3-319-72150-7_82.

[26] A. Anagnostopoulos, R. Kumar, and M. Mahdian, "Influence and Correlation in Social Networks", in: *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '08, Las Vegas, Nevada, USA: Association for Computing Machinery, 2008, pp. 7–15, DOI: 10.1145/1401890.1401897.

[27] S. Battiston et al., "Debtrank: Too central to fail? Financial networks, the FED and systemic risk", in: *Scientific Reports* **2**, 541 (2012), DOI: 10.1038/srep00541.

[28] S. Battiston et al., "Default cascades: When does risk diversification increase stability?", in: *Journal of Financial Stability* **8** (2012), pp. 138–149, DOI: 10.1016/j.jfs.2012.01.002.

[29] S. Battiston et al., "Liaisons dangereuses: Increasing connectivity, risk sharing, and systemic risk", in: *Journal of Economic Dynamics and Control* **36** (2012), pp. 1121–1141, DOI: 10.1016/j.jedc.2012.04.001.

[30] V. Zlatić, G. Gabbi, and H. Abraham, "Reduction of Systemic Risk by Means of Pigouvian Taxation", in: *PLOS ONE* **10** (2015), pp. 1–18, DOI: 10.1371/journal.pone.0114928.

[31] S. Poledna and S. Thurner, "Elimination of systemic risk in financial networks by means of a systemic risk transaction tax", in: *Quantitative Finance* **16** (2016), pp. 1599–1613, DOI: 10.1080/14697688.2016.1156146.

[32] T. Roukny et al., "Default cascades in complex networks: Topology and systemic risk", in: *Scientific Reports* **3**, 2759 (2013), pp. 1–8, DOI: 10.1038/srep02759.

[33] G. D'Agostino et al., "Robustness and assortativity for diffusion-like processes in scale-free networks", in: *EPL (Europhysics Letters)* **97**, 68006 (2012), DOI: 10.1209/0295-5075/97/68006.

[34] S. M. Krause et al., "Controlling systemic risk: Network structures that minimize it and node properties to calculate it", in: *Phys. Rev. E* **103**, 042304 (2021), DOI: 10.1103/PhysRevE.103.042304.

[35] G. Visentin, S. Battiston, and M. D'Errico, *Rethinking financial contagion*, 2016, arXiv: 1608.07831.

[36] J. Gregory, *Counterparty Credit Risk and Credit Value Adjustment: A Continuing Challenge for Global Financial Markets*, The Wiley Finance Series, Wiley, 2012, DOI: 10.1002/9781118673638.

[37] L. Gauvin et al., *Randomized reference models for temporal networks*, 2018, arXiv: 1806.04032.

[38] P. C. Mahalanobis, "On the generalized distance in statistics", in: National Institute of Science of India, 1936, URL: http://bayes.acs.unt.edu:8083/BayesContent/class/Jon/MiscDocs/1936_Mahalanobis.pdf.

[39] BCBS, *Capital treatment for bilateral counterparty credit risk finalised by the Basel Committee on Banking Supervision.* Tech. rep., 2011.

[40] BCBS, *Credit Valuation Adjustment risk: Targeted final revisions*, tech. rep., 2019.

[41] J. Hull, *Options, Futures, and Other Derivatives*, Prentice Hall, 2012, Google Books: LzBTpwAACAAJ.

[42] C. Gouriéroux, J.-C. Héam, and A. Monfort, "Bilateral exposures and systemic solvency risk", in: *Canadian Journal of Economics/Revue canadienne d'économique* **45** (2012), pp. 1273–1309, DOI: 10.1111/j.1540-5982.2012.01750.x.

[43] T. Banerjee and Z. Feinstein, "Pricing of Debt and Equity in a Financial Network with Comonotonic Endowments", in: *Operations Research* **0** (0), DOI: 10.1287/opre.2022.2275.

[44] P. Barucca et al., "Network valuation in financial systems", in: *Mathematical Finance* **30** (2020), pp. 1181–1204, DOI: 10.1111/mafi.12272.

[45] T. Schmidt, "Coping with copulas", in: J. Rank, *Copulas - From Theory to Application in Finance*, 2007, URL: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.139.888&rep=rep1&type=pdf.

[46] L. Eisenberg and T. H. Noe, "Systemic Risk in Financial Systems", in: *Management Science* **47** (2001), pp. 236–249, DOI: 10.1287/mnsc.47.2.236.9835.

[47] D. in 't Veld, M. van der Leij, and C. Hommes, "The formation of a core-periphery structure in heterogeneous financial networks", in: *Journal of Economic Dynamics and Control* **119**, 103972 (2020), DOI: 10.1016/j.jedc.2020.103972.

[48] J. C. Collins, *Renormalization: An Introduction to Renormalization, the Renormalization Group and the Operator-Product Expansion*, Cambridge Monographs on Mathematical Physics, Cambridge University Press, 1984, DOI: 10.1017/CBO9780511622656.

[49] M. Heimpel and P. Olson, "A seismodynamical model of lithosphere deformation: Development of continental and oceanic rift networks", in: *Journal of Geophysical Research: Solid Earth* **101** (1996), pp. 16155–16176, DOI: 10.1029/96JB00168.

[50] H. Saleur, C. G. Sammis, and D. Sornette, "Discrete scale invariance, complex fractal dimensions, and log-periodic fluctuations in seismicity", in: *Journal of Geophysical Research: Solid Earth* **101** (1996), pp. 17661–17677, DOI: 10.1029/96JB00876.

[51] H. Saleur, C. G. Sammis, and D. Sornette, "Renormalization group theory of earthquakes", in: *Nonlinear Processes in Geophysics* **3** (1996), pp. 102–109, DOI: 10.5194/npg-3-102-1996.

[52] J.-C. Anifrani et al., "Universal log-periodic correction to renormalization group scaling for rupture stress prediction from acoustic emissions", in: *J. Phys. I France* **5** (1995), pp. 631–638, DOI: 10.1051/jp1:1995156.

[53] J. A. Feigenbaum and P. G. Freund, "Discrete scale invariance in stock markets before crashes", in: *International Journal of Modern Physics B* **10** (1996), pp. 3737–3745, DOI: 10.1142/S021797929600204X.

[54] A. Johansen and D. Sornette, "Critical Crashes", in: *Risk* **12** (1999), pp. 91–94, arXiv: cond-mat/9901035.

[55] D. Sornette and A. Johansen, "Large financial crashes", in: *Physica A: Statistical Mechanics and its Applications* **245** (1997), pp. 411–422, DOI: 10.1016/S0378-4371(97)00318-X.

[56] D. Sornette, A. Johansen, and J.-P. Bouchaud, "Stock market crashes, precursors and replicas", in: *J. Phys. I France* **6** (1996), pp. 167–175, DOI: 10.1051/jp1:1996135.

[57] G. Grimmett et al., *Percolation*, Die Grundlehren der mathematischen Wissenschaften in Einzeldarstellungen, Springer, 1999, DOI: 10.1007/978-3-662-03981-6.

[58] I. Jensen, "Critical behavior of branching annihilating random walks with an odd number of offsprings", in: *Phys. Rev. E* **47** (1993), R1–R4, DOI: 10.1103/PhysRevE.47.R1.

[59] H. Takayasu and A. Y. Tretyakov, "Extinction, survival, and dynamical phase transition of branching annihilating random walk", in: *Phys. Rev. Lett.* **68** (1992), pp. 3060–3063, DOI: 10.1103/PhysRevLett.68.3060.

[60] P. Bak, C. Tang, and K. Wiesenfeld, "Self-organized criticality: An explanation of the 1/f noise", in: *Phys. Rev. Lett.* **59** (1987), pp. 381–384, DOI: 10.1103/PhysRevLett.59.381.

[61] P. Bak, *How nature works: the science of self-organized criticality*, Copernicus, 1996, DOI: 10.1007/978-1-4757-5426-1.

[62] H. J. Jensen, *Self-organized criticality: emergent complex behavior in physical and biological systems*, vol. 10, Cambridge university press, 1998, DOI: 10.1017/CBO9780511622717.

[63] T. C. Choy, *Effective medium theory: principles and applications*, vol. 165, Oxford University Press, 2015, DOI: 10.1093/acprof:oso/9780198705093.001.0001.

[64] K. Binder and A. P. Young, "Spin glasses: Experimental facts, theoretical concepts, and open questions", in: *Rev. Mod. Phys.* **58** (1986), pp. 801–976, DOI: 10.1103/RevModPhys.58.801.

[65] M. Newman, "Models of the Small World", in: *Journal of Statistical Physics* **101** (2000), pp. 819–841, DOI: 10.1023/A:1026485807148.

[66] K. Brinda and S. Vishveshwara, "A network representation of protein structures: implications for protein stability", in: *Biophysical Journal* **89** (2005), pp. 4159–4170, DOI: 10.1529/biophysj.105.064485.

[67] S. L. Pimm, J. H. Lawton, and J. E. Cohen, "Food web patterns and their consequences", in: *Nature* **350** (1991), pp. 669–674, DOI: 10.1038/350669a0.

[68] J. Moreno, *Who shall survive? Foundations of Sociometry, Group Psychotherapy and Sociodrama*, Nervous and mental disease monograph series, Beacon House, 1953, DOI: 10.2307/2785701.

[69] N. H. Gartner and C. Stamatiadis, "Traffic Networks, Optimization and Control of Urban", in: *Encyclopedia of Complexity and Systems Science*, New York, NY: Springer New York, 2009, pp. 9470–9500, DOI: 10.1007/978-0-387-30440-3_563.

[70] A. Broido and K. C. Claffy, "Internet topology: connectivity of IP graphs", in: *Scalability and traffic control in IP networks*, vol. 4526, International Society for Optics and Photonics, SPIE, 2001, pp. 172–187, DOI: 10.1117/12.434393.

[71] G. Cimini et al., "The statistical physics of real-world networks", in: *Nat Rev Phys* **1** (2019), pp. 58–71, DOI: 10.1038/s42254-018-0002-6.

[72] M. Newman, *Networks*, Oxford University Press, 2018, DOI: 10.1093/oso/9780198805090.001.0001.

[73] S. Milgram, "The small world problem", in: *Psychology Today* **1** (1967), pp. 60–67, URL: http://snap.stanford.edu/class/cs224w-readings/milgram67smallworld.pdf.

[74] S. Wasserman and K. Faust, *Social network analysis: methods and applications*, Structural Analysis in the Social Sciences, Cambridge University Press, 1994, DOI: 10.1017/CBO9780511815478.

[75] J. Scott, *Social Network Analysis: A Handbook*, Social Network Analysis: A Handbook, SAGE Publications, 2000, Google Books: meSF2CquIAgC.

[76] P. W. Holland and S. Leinhardt, "Transitivity in structural models of small groups", in: *Comparative Group Studies* **2** (1971), pp. 107–124, DOI: 10.1177/104649647100200201.

[77] T. Silva and L. Zhao, *Machine Learning in Complex Networks*, Springer International Publishing, 2016, DOI: 10.1007/978-3-319-17290-3.

[78] D. G. Rand, S. Arbesman, and N. A. Christakis, "Dynamic social networks promote cooperation in experiments with humans", in: *Proceedings of the National Academy of Sciences* **108** (2011), pp. 19193–19198, DOI: 10.1073/pnas.1108243108.

[79] S. Liu et al., "Heterogeneous network approach to predict individuals' mental health", in: *ACM Trans. Knowl. Discov. Data* **15** (2021), DOI: 10.1145/3429446.

[80] J. You et al., "Graph Structure of Neural Networks", in: *Proceedings of the 37th International Conference on Machine Learning*, vol. 119, Proceedings of Machine Learning Research, PMLR, 2020, pp. 10881–10891, URL: https://proceedings.mlr.press/v119/you20b.html.

[81] C. Matias et al., "Network motifs: mean and variance for the count", in: *REVSTAT-Statistical Journal* **4** (2006), pp. 31–51, URL: https://hal.inrae.fr/hal-02655236/document.

[82] F. Picard et al., "Assessing the Exceptionality of Network Motifs", in: *Journal of Computational Biology* **15** (2008), pp. 1–20, DOI: 10.1089/cmb.2007.0137.

[83] C. Gardiner, *Handbook of Stochastic Methods for Physics, Chemistry, and the Natural Sciences*, Proceedings in Life Sciences, Springer-Verlag, 1985, URL: https://link.springer.com/book/9783540707127.

[84] J. T. Cox, "Coalescing Random Walks and Voter Model Consensus Times on the Torus in $\mathbb{Z}^d$", in: *The Annals of Probability* **17** (1989), pp. 1333–1366, DOI: 10.1214/aop/1176991158.

[85] T. C. Schelling, "Hockey Helmets, Concealed Weapons, and Daylight Saving: A Study of Binary Choices with Externalities", in: *The Journal of Conflict Resolution* **17** (1973), pp. 381–428, DOI: 10.1177/002200277301700302.

[86] I. Z. Kiss, J. C. Miller, and P. L. Simon, "Mean-field approximations for heterogeneous networks", in: *Mathematics of Epidemics on Networks: From Exact to Approximate Models*, Cham: Springer International Publishing, 2017, pp. 165–205, DOI: 10.1007/978-3-319-50806-1_5.

[87] V. Marceau et al., "Adaptive networks: Coevolution of disease and topology", in: *Phys. Rev. E* **82**, 036116 (2010), DOI: 10.1103/PhysRevE.82.036116.

[88] J. Lorenz, S. Battiston, and F. Schweitzer, "Systemic risk in a unifying framework for cascading processes on networks", in: *The European Physical Journal B* **71**, 441 (2009), DOI: 10.1140/epjb/e2009-00347-4.

[89] R. M. Anderson and R. M. May, "Population biology of infectious diseases: Part I", in: *Nature* **280** (1979), pp. 361–367, DOI: 10.1038/280361a0.

[90] R. M. Anderson and R. M. May, *Infectious diseases of humans: dynamics and control*, Oxford University Press, 1992, DOI: 10.1017/S0950268800059896.

[91] P. Gai and S. Kapadia, "Contagion in financial networks", in: *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* **466** (2010), pp. 2401–2423, DOI: 10.2139/ssrn.1577043.

[92] L. Weng et al., "Competition among memes in a world with limited attention", in: *Scientific Reports* **2**, 335 (2012), DOI: 10.1038/srep00335.

[93] D. G. Kendall, "Stochastic processes occurring in the theory of queues and their analysis by the method of the imbedded Markov chain", in: *The Annals of Mathematical Statistics* (1953), pp. 338–354, DOI: 10.1214/aoms/1177728975.

[94] E. Volz and L. A. Meyers, "Susceptible–infected–recovered epidemics in dynamic contact networks", in: *Proceedings of the Royal Society B: Biological Sciences* **274** (2007), pp. 2925–2934, DOI: 10.1098/rspb.2007.1159.

[95] S. Blower and M.-H. Go, "The importance of including dynamic social networks when modeling epidemics of airborne infections: does increasing complexity increase accuracy?", in: *BMC medicine* **9**, 88 (2011), DOI: 10.1186/1741-7015-9-88.

[96] I. Vodenska and A. P. Becker, "Interdependence, vulnerability and contagion in financial and economic networks", in: *New Perspectives and Challenges in Econophysics and Sociophysics*, Springer, 2019, pp. 101–116, DOI: 10.1007/978-3-030-11364-3_8.

[97] P. J. Mucha et al., "Community Structure in Time-Dependent, Multiscale, and Multiplex Networks", in: *Science* **328** (2010), pp. 876–878, DOI: 10.1126/science.1184819.

[98] S. Boccaletti et al., "The structure and dynamics of multilayer networks", in: *Physics Reports* **544** (2014), pp. 1–122, DOI: 10.1016/j.physrep.2014.07.001.

[99] E. Cozzo et al., *Multiplex Networks*, 1st ed., SpringerBriefs in Complexity, Springer International Publishing, 2018, DOI: 10.1007/978-3-319-92255-3.

[100] C. D. Brummitt and T. Kobayashi, "Cascades in multiplex financial networks with debts of different seniority", in: *Physical Review E* **91**, 062813 (2015), DOI: 10.1103/PhysRevE.91.062813.

[101] T. Kobayashi, "A model of financial contagion with variable asset returns may be replaced with a simple threshold model of cascades", in: *Economics Letters* **124** (2014), pp. 113–116, DOI: 10.1016/j.econlet.2014.05.003.

[102] D. Acemoglu, A. Ozdaglar, and A. Tahbaz-Salehi, "Systemic risk and stability in financial networks", in: *American Economic Review* **105** (2015), pp. 564–608, DOI: 10.1257/aer.20130456.

[103] N. Kiyotaki and J. Moore, "Credit chains", 1998, URL: https://www.princeton.edu/~kiyotaki/papers/creditchains.pdf.

[104] F. Boissay, *Credit chains and the propagation of financial distress*, ECB working paper, 2006, DOI: 10.2139/ssrn.872543.

[105] T. Jacobson and E. Von Schedvin, "Trade credit and the propagation of corporate failure: an empirical analysis", in: *Econometrica* **83** (2015), pp. 1315–1371, DOI: 10.3982/ECTA12148.

[106] D. Acemoglu et al., "The network origins of aggregate fluctuations", in: *Econometrica* **80** (2012), pp. 1977–2016, DOI: 10.3982/ECTA9623.

[107] M. G. Hertzel et al., "Inter-firm linkages and the wealth effects of financial distress along the supply chain", in: *Journal of Financial Economics* **87** (2008), pp. 374–387, DOI: 10.1016/j.jfineco.2007.01.005.

[108] E. Oberfield, "A theory of input–output architecture", in: *Econometrica* **86** (2018), pp. 559–589, DOI: 10.3982/ECTA10731.

[109] J. Gao, *The Effects of Firm Network on Banks' Portfolio Consideration*, tech. rep., Mimeo. Kelly School of Business, Indiana University, 2015, DOI: 10.2139/ssrn.2829988.

[110] S. Battiston et al., "Credit chains and bankruptcy propagation in production networks", in: *Journal of Economic Dynamics and Control* **31** (2007), pp. 2061–2084, DOI: 10.1016/j.jedc.2007.01.004.

[111] S. Košćak et al., *Hub Think Tank Broj 3: Arhitektura Mreže Dugovanja Hrvatskih Tvrtki*, (Text is in Croatian), 2015, URL: https://www.hub.hr/hr/hub-think-tank-broj-3-arhitektura-mreze-dugovanja-hrvatskih-tvrtki.

[112] M. Torricelli, M. Karsai, and L. Gauvin, "weg2vec: Event embedding for temporal networks", in: *Scientific Reports* **10**, 7164 (2020), DOI: 10.1038/s41598-020-63221-2.

[113] S. Lehmann, "Fundamental Structures in Temporal Communication Networks", in: *Temporal Network Theory*, Springer, 2019, pp. 25–48, DOI: 10.1007/978-3-030-23495-9_2.

[114] A. Mellor, *Analysing collective behaviour in temporal networks using event graphs and temporal motifs*, 2018, arXiv: 1801.10527.

[115] S. S. Shen-Orr et al., "Network motifs in the transcriptional regulation network of Escherichia coli", in: *Nature Genetics* **31** (2002), pp. 64–68, DOI: 10.1038/ng881.

[116] U. Alon, "Network motifs: theory and experimental approaches", in: *Nature Reviews Genetics* **8** (2007), pp. 450–461, DOI: 10.1038/nrg2102.

[117] K. Ristl, S. J. Plitzko, and B. Drossel, "Complex response of a food-web module to symmetric and asymmetric migration between several patches", in: *Journal of Theoretical Biology* **354** (2014), pp. 54–59, DOI: 10.1016/j.jtbi.2014.03.009.

[118] T. Ohnishi, H. Takayasu, and M. Takayasu, "Network motifs in an inter-firm network", in: *Journal of Economic Interaction and Coordination* **5** (2010), pp. 171–180, DOI: 10.1007/s11403-010-0066-6.

[119] F. W. Takes et al., "Multiplex network motifs as building blocks of corporate networks", in: *Applied Network Science* **3**, 39 (2018), DOI: 10.1007/s41109-018-0094-z.

[120] V. Zlatić et al., "Wikipedias: Collaborative web-based encyclopedias as complex networks", in: *Physical Review E* **74**, 016115 (2006), DOI: 10.1103/PhysRevE.74.016115.

[121] A. Paranjape, A. R. Benson, and J. Leskovec, "Motifs in temporal networks", in: *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining*, 2017, pp. 601–610, DOI: 10.1145/3018661.3018731.

[122] P. V. Paulau, C. Feenders, and B. Blasius, "Motif analysis in directed ordered networks and applications to food webs", in: *Scientific Reports* **5**, 11926 (2015), DOI: 10.1038/srep11926.

[123] A. C. Schwarze and M. A. Porter, "Motifs for Processes on Networks", in: *SIAM Journal on Applied Dynamical Systems* **20** (2021), pp. 2516–2557, DOI: 10.1137/20M1361602.

[124] T. LaRock et al., "HYPA: Efficient Detection of Path Anomalies in Time Series Data on Networks", in: *Proceedings of the 2020 SIAM International Conference on Data Mining (SDM)*, pp. 460–468, DOI: 10.1137/1.9781611976236.52.

[125] I. Barjašić et al., "Causal motifs and existence of endogenous cascades in directed networks with application to company defaults", in: *Scientific Reports* **11**, 24028 (2021), DOI: 10.1038/s41598-021-02976-8.

[126] I. Barjašić et al., *Supplementary information*, URL: https://github.com/ibarjasi94/default_process.

[127] F. G. Elfadaly, P. H. Garthwaite, and J. R. Crawford, "On point estimation of the abnormality of a Mahalanobis index", in: *Computational Statistics & Data Analysis* **99** (2016), pp. 115–130, DOI: 10.1016/j.csda.2016.01.014.

[128] R. C. Merton, "On the Pricing of Corporate Debt: The Risk Structure of Interest Rates", in: *The Journal of Finance* **29** (1974), pp. 449–470, DOI: 10.2307/2978814.

[129] F. Rehm and M. Rudolf, "KMV Credit Risk Modeling", in: *Risk Management: Challenge and Opportunity*, Berlin, Heidelberg: Springer Berlin Heidelberg, 2000, pp. 141–154, DOI: 10.1007/978-3-662-04008-9_8.

[130] D. Duffie and K. Singleton, "Modeling Term Structures of Defaultable Bonds", in: *Review of Financial Studies* **12** (1999), pp. 687–720, DOI: 10.1093/rfs/12.4.687.

[131] C. H. Furfine, "Interbank Exposures: Quantifying the Risk of Contagion", in: *Journal of Money, Credit and Banking* **35** (2003), pp. 111–128, DOI: 10.2139/ssrn.169089.

[132] L. C. G. Rogers and L. A. M. Veraart, "Failure and Rescue in an Interbank Network", in: *Management Science* **59** (2013), pp. 882–898, DOI: 10.2139/ssrn.1932911.

[133] T. Suzuki, "Valuing Corporate Debt: The Effect of Cross-Holdings of Stock and Debt", in: *Journal of the Operations Research Society of Japan* **2** (2002), pp. 123–144, DOI: 10.15807/jorsj.45.123.

[134] T. Banerjee, A. Bernstein, and Z. Feinstein, *Dynamic Clearing and Contagion in Financial Networks*, 2018, arXiv: 1801.02091.

[135] A. Sklar, "Fonctions de Répartition À N Dimensions Et Leurs Marges", in: *Publications de l'Institut Statistique de l'Université de Paris* **8** (1959), pp. 229–231, Google Books: nreSmAEACAAJ.

[136] P. Krupskii and H. Joe, "Factor copula models for multivariate data", in: *Journal of Multivariate Analysis* **120** (2013), pp. 85–101, DOI: 10.1016/j.jmva.2013.05.001.

[137] J. Ward et al., "Micro-scale foundation with error quantification for the approximation of dynamics on networks", in: *Communications Physics* **5**, 71 (2022), DOI: 10.1038/s42005-022-00834-1.

[138] M. E. Newman, A.-L. E. Barabási, and D. J. Watts, *The structure and dynamics of networks*. Princeton University Press, 2006, URL: https://www.jstor.org/stable/j.ctt7ssgv.

[139] R. Pastor-Satorras and A. Vespignani, "Epidemic spreading in scale-free networks", in: *Physical Review Letters* **86** (2001), pp. 3200–3203, DOI: 10.1103/PhysRevLett.86.3200.

[140] R. R. Nadakuditi and M. E. Newman, "Graph spectra and the detectability of community structure in networks", in: *Physical Review Letters* **108**, 188701 (2012), DOI: 10.1103/PhysRevLett.108.188701.

[141] P. W. Holland, K. B. Laskey, and S. Leinhardt, "Stochastic blockmodels: First steps", in: *Social Networks* **5** (1983), pp. 109–137, DOI: 10.1016/0378-8733(83)90021-7.

# Curriculum vitae

Irena Barjašić was born on the 13th of December 1994 in Zagreb, Croatia. After completing her secondary education at Gimnazija Jurja Barakovića in 2012, she enrolled on the Physics programme at the Department of Physics, Faculty of Science, University of Zagreb. In 2017 she defended her Master's thesis, titled "Symmetry inheritance", under the supervision of Ivica Smolić, PhD. In 2018 she enrolled on the Biophysics doctoral programme at the Department of Physics, working as a teaching assistant at the Scientific Center of Excellence for Quantum and Complex Systems, and Representations of Lie Algebras (QuantiXLie). She defended her doctoral thesis title "Network effects in systemic risk propagation" ("Mrežni utjecaj u propagaciji sistemskog rizika") under the supervision of Vinko Zlatić, PhD in 2020.

During her doctoral studies, in 2019, she visited the group of Prof. Dirk Helbing at ETH Zurich. Under the supervision of Nino Antulov-Fantulin, PhD, she worked on improving Bitcoin volatility prediction on a minute scale by including external signals into the GARCH model, which resulted in a publication. In 2020 she was awarded the Swiss Government Excellence Scholarship, to do a one-year-long research visit to the University of Zurich. There, in the group of Prof. Stefano Battiston and under his supervision, she researched the contribution of the credit exposure network to the Credit Valulation Adjustment.

## Publications

1. Barjašić, I., Štefančić, H., Pribičević, V., Zlatić, V. (2021) Causal motifs and existence of endogenous cascades in directed networks with application to company

defaults. Scientific reports, 11, 24028. doi:10.1038/s41598-021-02976-8

2. Barjašić I. and Antulov-Fantulin N. (2021) Time-Varying Volatility in Bitcoin Market and Information Flow at Minute-Level Frequency. Front. Phys. 9, 644102. doi:10.3389/fphy.2021.644102

3. Zlatić V, Barjašić I., Kadović A., Štefančić H., Gabrielli A. (2020) Bi-stability of SUDR+K model of epidemics and test kits applied to COVID-19. Nonlinear Dynamics 101, 1635–1642. doi:10.1007/s11071-020-05888-w

4. Barjašić, I., Smolić, I. (2018) On symmetry inheritance of nonminimally coupled scalar fields. Classical and quantum gravity 35, 075002. doi:10.1088/1361-6382/aaabfc

5. Barjašić, I., Gulin, L., Smolić, I. (2017) Nonlinear electromagnetic fields and symmetries. Physical review. D 95, 124037. doi:10.1103/PhysRevD.95.124037