

# Studie av olika strategier för att förhindra spridningen av ett datorvirus inom ett e-postnätverk

Carl-Michael Lindblom Berglund

Kandidatuppsats i matematisk statistik Bachelor Thesis in Mathematical Statistics

Kandidatuppsats 2016:6 Matematisk statistik Juni 2016

www.math.su.se

Matematisk statistik Matematiska institutionen Stockholms universitet 106 91 Stockholm

# Matematiska institutionen



Matematisk statistik Stockholms universitet Kandidatuppsats **2016:6** http://www.math.su.se/matstat

## Studie av olika strategier för att förhindra spridningen av ett datorvirus inom ett e-postnätverk

Carl-Michael Lindblom Berglund\*

## Juni 2016

#### Sammanfattning

I denna uppsats undersöks hur man med hjälp av immunisering kan förhindra spridningen av en smitta inom ett riktat stokastiskt nätverk konstruerat enligt konfigurationsmodellen. Tanken är att nätverket ska beskriva ett e-postnätverk där noderna motsvarar användare av en eposttjänst och där varie förekomst av en användares e-postadress i en annan användares adressbok motsvaras av en riktad kant. Smittan motsvarar ett datorvirus som skickas med e-post från en infekterad användare till samtliga adresser i dennes adressbok, det vill säga till de noder, vilka i uppsatsen benämns som grannar, som kan nås från den infekterade noden via en riktad kant. Immuniseringen av en nod motsvarar installation av ett antivirusprogram på en användares dator, och innebär att denna användare ej längre är mottaglig för datorviruset. Tre olika immuniseringsstrategier studeras i uppsatsen: där de noder som görs immuna väljs (1) proportionellt mot nodernas utgrader, (2) proportionellt mot nodernas ingrader respektive (3) oberoende av nodernas utgrader och ingrader. Utgraden hos en nod v är antalet grannar till v och ingraden hos en nod v är antalet noder som har v som granne. I huvudsak undersöks vilken påverkan de olika immuniseringsstrategierna har på smittspridningen, och hur denna påverkan beror på nodernas gradfördelningar samt när under smittprocessen immuniseringen utförs. I uppsatsen görs uträkningar för att bestämma villkor som avgör när sannolikheten för en stor epidemi är större än 0 för var och en av immuniseringsstrategierna. Med hjälp av simulering undersöks även hur epidemins slutstorlek, det vill säga antalet noder som någon gång blir smittade, för de olika strategierna beror på valet av parametrar i modellen.

<sup>\*</sup>Postadress: Matematisk statistik, Stockholms universitet, 106 91, Sverige. E-post: cabe5146@student.su.se. Handledare: Maria Deijfen, Pieter Trapman.

## Innehåll

1	Introduktion				
	1.1	Frågeställning	3		
2 Modellen		dellen	3		
	2.1	Nätverket	3		
		2.1.1 Konfigurationsmodellen	4		
		2.1.2 Fördelningar för in- och utgrader	6		
	2.2	Smittmodellen	7		
	2.3	Immunisering	8		
3	Uträkningar				
	3.1	Ingen immunisering	10		
	3.2	Slumpmässig immunisering proportionell mot nodernas utgrader	11		
	3.3	Slumpmässig immunisering proportionell mot nodernas ingrader	12		
	3.4	Slumpmässig immunisering oberoende av nodernas ut- och ingrader	13		
	3.5	Jämförelse av de olika immuniseringsstrategierna $\ .\ .\ .\ .$ .	14		
4	$\mathbf{Sim}$	ulering	16		
	4.1	Resultat	17		
		4.1.1 Tungsvansade fördelningar med ändligt andramoment	17		
		4.1.2 Tungsvansade fördelningar med o ändligt andramoment	20		
		4.1.3 Icke-tungsvansade fördelningar	21		
		4.1.4 Olika värden på parametern $s$	23		
<b>5</b>	Dis	kussion	25		
$\mathbf{A}_{\mathbf{j}}$	Appendix 28				

## 1 Introduktion

I denna uppsats undersöks olika strategier för att förhindra spridningen av en smitta inom ett riktat stokastiskt nätverk. Tanken är att den nätverksmodell som används ska beskriva ett e-postnätverk där varje användare av en viss eposttjänst motsvaras av en nod i nätverket och varje förekomst av en användares e-postadress i en annan användares adressbok motsvaras av en riktad kant. Nätverksmodellen som används är en riktad variant av konfigurationsmodellen, vilket innebär att varje nods ingrad respektive utgrad väljs slumpmässigt från två fördelningar F och G som har samma väntevärde, och att de riktade kanterna skapas genom att de utgående och ingående halvkanterna paras ihop slumpmässigt så länge det finns ingående och utgående halvkanter att para ihop.[3] Modellen kommer att vara en *erased configuration model*, vilket innebär att eventuella loopar, det vill säga riktade kanter från en not till sig själv, och multipla kanter som uppstår vid ihopparningen tas bort från nätverket.[3]

Smittan motsvarar ett datorvirus som sprids med hjälp av e-postmeddelanden från en infekterad användares konto till samtliga e-postadresser i användarens adressbok, det vill säga till de noder, vilka i uppsatsen kommer att benämnas som grannar, som kan nås från motsvarande infekterade nod via en riktad kant. En användare som tar emot ett sådant meddelande får sin dator infekterad om användaren öppnar en länk i meddelandet. I en mer verklighetstrogen modell hade det varit rimligt att anta att smittspridningen sker i kontinuerlig tid, men för att modellen inte ska bli alltför komplicerad antas smittspridningen ske i diskret tid. I modellen för smittspridningen blir en slumpmässigt utvald nod smittad i tidssteg 1, och varje mottaglig granne till en nod som har blivit smittad i tidssteg  $t \ge 1$  kan i sin tur bli smittad i det efterföljande tidssteget t + 1, vilket sker med med en sannolikhet p, som är en given konstant. En infekterad nod antas alltså vara smittsam i ett tidssteg efter det att den smittas, och därefter återhämtad och immun.

Immunisering av en nod motsvarar installation av ett antivirusprogram på en användares dator. I modellen utförs immuniseringen på en andel s av noderna i det tidssteg T då antalet smittade noder för första gången är större eller lika med en given konstant h > 0, vilket motsvarar tidpunkten då epidemin upptäcks och identifieras. För en smittsam nod v antas sannolikheten  $\tilde{p}$  att en tidigare mottaglig granne till v som har gjorts immun genom immunisering blir smittad av v vara lika med 0. En infekterad nod som görs immun genom immunisering återhämtar sig genast från infektionen och förblir immun. Om immuniseringen utförs på en redan återhämtad nod, så påverkas denna nod inte alls.

I uppsatsen studeras följande tre immuniseringsstrategier:

- *Strategi 1:* de noder som görs immuna väljs proportionellt mot nodernas utgrader,
- *Strategi 2:* de noder som görs immuna väljs proportionellt mot nodernas ingrader,
- *Strategi 3:* de noder som görs immuna väljs slumpmässigt oberoende av nodernas ingrader och utgrader.

I studien undersöks med hjälp av uträkningar samt simulering, vilken görs i programmet R,[10] hur resultatet av immuniseringen skiljer sig mellan de tre strategierna för olika val av gradfördelningar F och G, samt parametern h.

## 1.1 Frågeställning

- Vilken påverkan har de olika immuniseringsstrategierna på smittspridningen?
- Hur beror resultatet av immuniseringen på in- och utgradernas fördelningar *F* respektive *G*?
- Hur beror resultatet av immuniseringen på valet av parametern *h*, epidemins storlek då immuniseringen görs?

## 2 Modellen

#### 2.1 Nätverket

Ett nätverk beskrivs lämpligast med hjälp av en graf (V, E), vilken består av en mängd noder V och en mängd kanter E, där varje kant mellan två noder  $v_1, v_2 \in V$  talar om att det finns någon form av relation mellan  $v_1$  och  $v_2$ .[1] Om nätverket är riktat är alla element i E ordnade par av noder  $(v_1, v_2)$ , där  $v_1, v_2 \in V$ , och om nätverket är oriktat är alla element i E oordnade par av noder  $\{v_1, v_2\}$ , där  $v_1, v_2 \in V$ .[8] Graden hos en nod v i ett oriktat nätverk är antalet kanter i E där noden v ingår, och betecknas deg(v).[8] I ett riktat nätverk är utgraden od(v) och ingraden id(v) hos en nod v antalet kanter i Edär v förekommer som det första respektive det andra elementet i det ordnade paret.[3] En kant där samma nod förekommer som första och andra element i paret kallas för loop.[1] I denna uppsats kommer begreppet granne (till en nod  $v \in V$ ) att användas som benämning på en nod som kan nås från v via en riktad kant inom en riktad graf, det vill säga en nod  $v_1$  sådan att  $(v, v_1) \in V$ . En graf kallas för slumpgraf om det finns någon form av slump i hur grafen är konstruerad.[1] Slumpgrafer är vanliga modeller för att beskriva verkliga nätverk, som exempelvis sociala nätverk, elnätverk eller internet.[1] Den nätverkmodell som studeras i denna uppsats är tänkt att beskriva ett e-postnätverk där användarna av en viss e-posttjänst motsvaras av noder i nätverket och där förekomsten av användare  $v_2$ :s e-postadress i användare  $v_1$ :s adressbok motsvaras av en riktad kant  $(v_1, v_2)$  mellan noderna  $v_1$  och  $v_2$ , för alla  $v_1, v_2 \in V$ .

#### 2.1.1 Konfigurationsmodellen

Konfigurationsmodellen är en nätverksmodell där nodernas grader slumpas från en viss diskret fördelning och där kanterna skapas genom slumpmässig ihopparning av halvkanter som ges av nodernas grader. [1, 3] Den modell som studeras i den här uppsatsen är en riktad variant av konfigurationsmodellen, där nodernas utgrader respektive ingrader slumpas från de diskreta fördelningarna F och G som har samma väntevärde, och där de riktade kanterna skapas genom att de utgående och ingående halvkanterna paras ihop slumpmässigt så länge det finns ingående och utgående halvkanter att para ihop.[3] Att använda en konfigurationsmodell för att beskriva ett verkligt nätverk är lämpligt då gradfördelningarna skiljer sig markant från Poissonfördelningen, som är den asymptotiska gradfördelningen hos den enklaste slumpgrafen Erdős-Rényi-grafen, på så sätt att sannolikheten för extrema utfall är relativt stor.[1] Sådana fördelningar kallas tungsvansade (heavy-tailed på engelska),[4] och är vanliga i modeller för e-postnätverk.[6] Eftersom det kan anses osannolikt att någon användare av en e-posttjänst väljer att lägga in samma e-postadress flera gånger i sin adressbok och ointressant för smittspridningen om användaren har sin egen epostadress i sin addressbok, är det lämpligast att beskriva e-postnätverket med en enkel graf, det vill säga en graf som saknar loopar och multipla kanter.[1] För att de nätverk som studeras i denna uppsats ska uppfylla detta används en erased configuration model, vilket innebär att loopar och multipla kanter har tagits bort efter att ihopparningen har gjorts.[3] Om vi låter  $F_n$  och  $G_n$ beteckna ingradernas respektive utgradernas fördelning i denna reducerade riktade konfigurationsmodellen då antalet noder är n, och om F och G har ändligt väntevärde, så gäller att

$$F_n \to F$$

och

$$G_n \to G$$

då  $n \to \infty$ .[3] I Figur 1 illustreras hur ett nätverk bestående av fem noder skapas enligt konfigurationsmodellen. De nätverk som studeras i uppgiften kommer att ha ett betydligt större antal noder, men kommer att skapas genom samma process.



Figur 1: Illustration av hur ett riktat nätverk bestående av fem noder (numrerade 1–5) skapas enligt konfigurationsmodellen. I delfigur (a) har var och en av noderna tilldelats en utgrad och en ingrad, vilka representeras av antalet utgående respektive ingående halvkanter (där pilarna anger halvkanternas riktning). I delfigur (b) har varje utgående halvkant parats ihop med en slumpmässigt vald ingående halvkant. I delfigur (c) har de ingående halvkanterna som blev över vid ihopparningen (totalt 5 stycken) tagits bort och i delfigur (d) har samtliga multipla kanter (inga) och loopar (kanten (4,4)) som uppstod vid ihopparningen tagits bort.

#### 2.1.2 Fördelningar för in- och utgrader

I modeller för e-postnätverk är det vanligt att nodernas grader följer en potenslag (*power law* på engelska),[6] vilket innebär att gradfördelningen har en sannolikhetsfunktion  $P_x$  sådan att

$$P_x \simeq C x^{-k}, \ k > 1, \ C > 0,$$

för stora värden på x.[9] På motsvarande sätt följer en kontinuerlig stokastisk variabel en potenslag om dess fördelning har en täthetsfunktion med mosvarande egenskaper som sannolikhetsfunktionen i det diskreta fallet.[9] Fördelningar som uppfyller detta är tungsvansade och har ändligt väntevärde om k > 2 och ändligt andramoment om k > 3.[9] För att kunna undersöka nätverk där ingradernas fördelning F och utgradernas fördelning G är två olika tungsvansade fördelningar med samma väntevärde, kommer nätverksmodeller med gradfördelningar  $F = Po(X_1)$  och  $G = Po(X_2)$ , där  $X_1 \sim Pareto(c_1, \alpha_1)$  och  $X_2 \sim$ Pareto $(c_2, \alpha_2)$ , att studeras. Paretofördelningen Pareto $(c, \alpha)$ , där  $c, \alpha > 0$ , är en kontinuerlig fördelning som följer en potenslag.[9] Täthetsfunktionen, väntevärdet och variansen för Pareto $(c, \alpha)$ -fördelningen ges av[7]

$$f(x) = \begin{cases} \frac{\alpha c^{\alpha}}{x^{\alpha+1}} & x \ge c\\ 0 & x < c \end{cases},$$
$$\mu = \begin{cases} \infty & \alpha \le 1\\ \frac{\alpha c}{\alpha-1} & \alpha > 1 \end{cases}$$

respektive

$$\sigma^{2} = \begin{cases} \infty & 1 < \alpha \leq 2\\ \left(\frac{c}{\alpha - 1}\right)^{2} \frac{\alpha}{\alpha - 2} & \alpha > 2 \end{cases}$$

Eftersom Poissonfördelningen Po(a) har väntevärdet a och variansen a, [7] och eftersom E(X) = E(E(X|Y)) och  $\operatorname{Var}(X) = \operatorname{Var}(E(X|Y)) + E(\operatorname{Var}(X|Y))$  för alla stokastiska variabler X, Y, [7] så gäller att om  $X \sim \operatorname{Po}(Y)$ , där  $Y \sim \operatorname{Pareto}(c, \alpha)$ , så har X väntevärdet

$$E(X) = E(E(X|Y)) = E(Y) = \mu$$

och variansen

$$\operatorname{Var}(X) = \operatorname{Var}(E(X|Y)) + E(\operatorname{Var}(X|Y)) = \operatorname{Var}(Y) + E(Y) = \sigma^2 + \mu$$

Eftersom Paretofördelningen är tungsvansad följer det av detta att även X har en tungsvansad fördelning. För att undersöka nätverk med icke-tungsvansade gradfördelningar kommer även modeller där F eller G är geometriska fördelningar,

och alltså icke-tungsvansade, att användas. En geometrisk fördelning med parameter q har sannolikhetsfunktionen

$$p(x) = q(1-q)^x,$$

väntevärdet  $\frac{1-q}{q}$  och variansen  $\frac{1-q}{q^2}$ .[7]

I en mer verklighetstrogen modell hade det varit rimligt att anta att en nods ingrad beror på dess utgrad, eftersom det kan anses troligt att en användare som skickar många e-postmeddelanden också tar emot många e-postmeddelanden, men för att undvika att modellen blir alltför komplicerad och opraktisk kommer detta att bortses ifrån.

## 2.2 Smittmodellen

I modellen motsvarar smittan ett datorvirus som vid den tidpunkt då en användares dator blir infekterad skickas med ett e-postmeddelande till samtliga e-postadresser i användarens adressbok. När en användare klickar på en länk i ett sådant meddelande blir dennes dator infekterad. I modellen antas smittspridningen ske i diskret tid, där en slumpmässigt vald nod smittas i tidssteg 1, och där varje mottaglig granne till en nod v som har blivit smittad i tidssteg  $t \ge 1$ , kan bli smittad av v i det efterföljande tidssteget t + 1 vilket sker med sannolikhet p, som är en given konstant. I en mer verklighetstrogen modell hade det möjligen varit lämpligt att låta smittspridningen ske i kontinuerlig tid, men eftersom det skulle innebära att modellen blev mycket mer komplicerad, så används i stället en smittmodell i diskret tid. De faser en nod kan genomgå är

- 1. mottaglig
- 2. infekterad (smittsam)
- 3. återhämtad (ej smittsam, immun)

Smittmodellen har formen av en SIR-modell, där S står för susceptible (mottaglig), I står för infected (infekterad) och R står för removed eller recovered (återhämtad).[2] En nod räknas som infekterad i ett tidssteg efter att den blev smittad. Därefter antas noden vara och förbli återhämtad, vilket innebär att noden inte längre kan smitta någon annan nod eller själv bli smittad igen. Det hade även varit möjligt att undersöka modeller där en nod antas vara infekterad och smittsam i  $m \geq 1$  tidssteg, för olika värden på heltalet m, men eftersom det skulle innebära onödigt komplicerade uträkningar och ytterligare en parameter att ta hänsyn till vid simuleringen i Avsnitt 4, som därmed kan bli alltför tidskrävande, begränsar sig denna studie till en smittmodell där m = 1.

#### 2.3 Immunisering

Immunisering av en nod motsvarar installation av ett antivirusprogram på en användares dator. Denna åtgärd utförs på en andel s av noderna i det tidssteg Tdå antalet smittade noder för första gången är större eller lika med en given konstant h > 0, vilket motsvarar tidpunkten då epidemin upptäcks och identifieras. Om en infekterad nod v har en tidigare mottaglig granne som har gjorts immun genom immunisering, så antas sannolikheten  $\tilde{p}$  att denna granne blir smittad av v vara lika med 0. Om en infekterad nod görs immun genom immunisering så återhämtar den sig från infektionen och förblir immun. En nod som redan är återhämtad förblir återhämtad om den görs immun genom immunisering. En nod som görs immun genom immunisering kan alltså gå från att vara mottaglig till att vara immun med en mycket låg smittsannolikhet  $\tilde{p}$ , eller gå från att vara infekterad (smittsam) eller återhämtad (ej smittsam, immun) till att vara återhämtad (ej smittsam, immun). Följande tre immuniseringsstartegier undersöks:

- *Strategi 1:* de noder som görs immuna väljs proportionellt mot nodernas utgrader,
- *Strategi 2:* de noder som görs immuna väljs proportionellt mot nodernas ingrader,
- *Strategi 3:* de noder som görs immuna väljs slumpmässigt oberoende av nodernas ingrader och utgrader.

Dessa strategier är rimliga, eftersom information om användarnas adressböcker i e-posttjänsten bör vara lagrad i en databas och alltså tillgänglig för e-posttjänstens ägare. Att noderna som görs immuna väljs proportionellt mot nodernas utgrader innebär att, för varje nod  $v_1$ ,

 $P(v_1 \text{ blir vald}) \propto \mathrm{od}(v_1),$ 

det vill säga sannolikheten att  $v_1$  blir vald ges av dess utgrad multiplicerat med någon konstant, som är lika med sannolikheten för en nod med utgrad 1 att bli vald. Motsvarande gäller då noder väljs proportionellt mot ingrad.

## 3 Uträkningar

I detta avsnitt kommer uträkningar att göras för att bestämma villkor som avgör när sannolikheten för ett stor epidemi är strikt större än 0, för var och en av immuniseringsstrategierna. Epidemins slutstorlek S, det vill säga antalet antalet noder i nätverket som någon gång blir smittade, för var och en av immuniseringsstrategierna, och hur reultatet av immuniseringen beror på parametern h, kommer att undersökas med hjälp av simuleringar i Avsnitt 4. Att ingen immunisering sker kommer i fortsättningen att benämnas *Immuniseringsstrategi*  $\theta$ . Eftersom antalet noder n är stort i modellen kan antagandet att  $n \to \infty$  göras inledningsvis. Detta innebär att att sannolikheten att en infekterad nod har en annan infekterad nod som granne eller delar granne med en annan infekterad nod konvergerar mot 0, och alltså kan antas vara lika med 0. Därför kan utvecklingen av epidemin ses som en förgreningsprocess i diskret tid, där varje nod som blir infekterad i ett tidssteg i sin tur ger upphov till ett visst antal infekterade noder i efterföljande tidssteg oberoende av andra infekterade noder, så länge smittspridningen befinner sig i ett tidigt stadium. Fördelnngarna F och G kommer att antas ha samma väntevärde  $\mu < \infty$ , vilket innebär att F och G kan antas vara fördelningarna för nodernas ingrader respektive utgrader, eftersom antagandet att  $n \to \infty$  görs.[3] I Tabell 1 listas de i modellen givna parametrarna, och i Tabell 2 listas de stokastiska variabler som undersöks.

Beteckning	Betydelse
n	antalet noder i nätverket
F	fördelningen för ingraderna då $n \to \infty$
G	fördelningen för utgraderna då $n \to \infty$
h	minsta antalet smittade i det tidssteg då immuniseringen görs
p	sannolikheten att en granne till en infekterad nod $\boldsymbol{v}$ smittas av $\boldsymbol{v}$
s	andelen noder som görs immuna
Ι	immuniseringsstrategi

Tabell 1: Givna parametrar i modellen

Beteckning	Betydelse
L	antalet grannar som en infekterad nod $v$ smittar
$S^{(t)}$	antalet noder som smittas i tidssteg $t$
S	epidemins slutstorlek, det vill säga antalet noder som någon gång
	blir smittade
T	tidssteget då antalet smittade noder överstiger $h$ , det vill säga
	tidssteget då immuniseringen görs

Tabell 2: Stokastiska variabler som undersöks

#### 3.1 Ingen immunisering

Om v antas vara en nod som smittas i tidssteg  $t \ge 1$  så är antalet mottagliga grannar till v i tidssteg t + 1 lika med  $od(v) \sim G$  för alla  $t \ge 1$ . Sannolikheten att x noder blir smittade av v i tidssteg t + 1 givet att od(v) = y ges då av uttrycket

$$p^x(1-p)^{y-x}\binom{y}{x},$$

vilket är sannolikhetsfunktionen för  $\operatorname{Bin}(y, p)$ -fördelningen.[7] Från detta följer att

$$L|\mathrm{od}(v) \sim \mathrm{Bin}(\mathrm{od}(v), p)$$

Väntevärdet och variansen för L ges av[7]

$$E(L) = E(E(L|\mathrm{od}(v))) = E(p \cdot \mathrm{od}(v)) = pE(\mathrm{od}(v))$$

respektive

$$\operatorname{Var}(L) = \operatorname{Var}(E(L|\operatorname{od}(v))) + E(\operatorname{Var}(L|\operatorname{od}(v))) =$$
$$\operatorname{Var}(p \cdot \operatorname{od}(v))) + E(p(1-p)\operatorname{od}(v)) = p^{2}\operatorname{Var}(\operatorname{od}(v)) + p(1-p)E(\operatorname{od}(v)).$$

Antalet noder  $S^{(t)}$  som smittas i tidssteg t > 1 fås genom att summera antalet noder som smittas av var och en av de smittsamma noderna. Eftersom m antas vara lika med 1 är de smittsamma noderna de noder som blev smittade i tidssteg t - 1. Smittspridningen beskrivs inledningsvis av en förgreningsprocess där pE(od(v)) är det förväntade antalet smittade som en infekterad nod ger upphov till. Väntevärdet och variansen för totala antalet noder som smittas i varje tidssteg t ges av[7]

$$E(S^{(t)}) = (E(L))^{t-1} = p^{t-1}(E(\mathrm{od}(v)))^{t-1}$$

respektive

$$Var(S^{(t)}) = Var(L)((E(L))^{t-2} + (E(L))^{t-1} + \dots + (E(L))^{2t-4}) =$$
$$(p^{2}Var(od(v)) + p(p-1)E(od(v)))(p^{t-2}(E(od(v)))^{t-2} +$$
$$p^{t-1}(E(od(v)))^{t-1} + \dots + p^{2t-4}(E(od(v)))^{2t-4})$$

Väntevärdet av det totala antalet smittade noder som smittas i tidsstegen  $1, \dots, t$ , ges av[7]

$$E\Big(\sum_{i=0}^{t} S^{(i)}\Big) = \sum_{i=0}^{t} E(S^{(i)}) = \sum_{i=0}^{t} p^{i-1}(E(\mathrm{od}(v)))^{i-1}).$$

Enligt en känd sats[7] är sannolikheten att epidemin dör ut innan den når ett stort utbrott 1 om pE(od(v)) < 1 och mindre än 1 om  $pE(od(v)) \ge 1$ .

## 3.2 Slumpmässig immunisering proportionell mot nodernas utgrader

I detta avsnitt undersöks smittspridningen i tidssteg t > T, det vill säga tidssteg efter att immuniseringen har skett. För tidssteg  $t \leq T$  har immuniseringen ännu inte gjorts vilket innebär att smittspridningen inte skiljer sig jämfört med för immuniseringsstrategi 0. Eftersom en nods utgrad antas vara oberoende av dess ingrad i modellen gäller, för t > T att sannolikheten att en granne till en smittad nod v är mottaglig är (1-s) och att sannolikheten att en granne till en smittad nod v har gjorts immun är s. Som i föregående avsnitt antas epidemin befinna sig i ett så pass tidigt stadium att utvecklingen kan beskrivas av en förgreningsprocess. Eftersom smittsannolikheten  $\tilde{p}$  för en immun granne till en infekterad nod är lika med 0 i modellen, gäller att sannolikheten att en granne  $v_1$  till en infekterad nod  $v_2$  blir smittad är

 $P(v_1 \text{ smittad av } v_2) =$ 

 $\underbrace{P(v_1 \text{ smittad av } v_2|v_1 \text{ har gjorts immun})}_{\hat{p}=0} P(v_1 \text{ har gjorts immun}) + \underbrace{P(v_1 \text{ smittad av } v_2|v_1 \text{ är mottaglig})}_{p(v_1 \text{ ar mottaglig})} P(v_1 \text{ är mottaglig}) = \underbrace{p(1-s)}$ 

Sannolikhetsfunktionen för antalet grannar till en smittsam nod, det vill säga för utgraden od(v) hos en nod v givet att noden är smittsam, ges av

$$\begin{split} P(\mathrm{od}(v) &= x | v \text{ är smittsam}) = P(\mathrm{od}(v) = x | v \text{ har inte gjorts immun}) = \\ & \frac{P(v \text{ har gjorts immun} | \mathrm{od}(v) = x) P(\mathrm{od}(v) = x)}{P(v \text{ har inte gjorts immun})} = \\ & \frac{(1 - P(v \text{ har gjorts immun} | \mathrm{od}(v) = x)) P(\mathrm{od}(v) = x)}{1 - s} \end{split}$$

där P(v har gjorts immun|od(v) = x) är en funktion som beror entydigt av G, fördelningen för utgraderna. Sannolikhetsfunktionen för antalet grannar L som en nod v smittar givet att noden har y grannar är

$$p(1-s))^{x}(1-(p(1-s)))^{y-x} \begin{pmatrix} y\\ x \end{pmatrix}$$

vilket är sannolikhetsfunktionen för  $\mathrm{Bin}(y,p(1-s))\text{-}$ fördelningen. [7] Från detta följer att

$$E(L) = E(L|(\mathrm{od}(v)|v \text{ är smittsam})) = p(1-s)E(\mathrm{od}(v)|v \text{ är smittsam}) = p(1-s)\sum_{i=0}^{\infty} iP(\mathrm{od}(v) = i|v \text{ är smittsam}) =$$

$$\begin{split} p(1-s)\sum_{i=0}^{\infty} i \frac{(1-P(v \text{ har gjorts immun}|\mathrm{od}(v)=i))P(\mathrm{od}(v)=i)}{1-s} = \\ p\sum_{i=0}^{\infty} i(1-P(v \text{ har gjorts immun}|\mathrm{od}(v)=i))P(\mathrm{od}(v)=i) \end{split}$$

Antalet noder  $S^{(t)}$  som smittas i tidssteg t > T kan beskrivas som en förgreningsprocess som inleds i tidssteg T, då ett visst antal noder smittas. Vi såg i föregående avsnitt att väntevärdet för detta antal är

$$p^{T-1}(E(\mathrm{od}(v)))^{T-1}.$$

Väntevärdet för antalet noder som smittas i tidssteg t > T är därför[7]

$$p^{T-1}(E(\mathrm{od}(v)))^{T-1} \left( p \sum_{i=0}^{\infty} i(1 - P(v \text{ har gjorts immun}|\mathrm{od}(v) = i)) P(\mathrm{od}(v) = i)) \right)^{t-T}$$

Enligt en känd sats[7] är sannolikheten 1 att epidemin dör ut innan den når ett stort utbrott om

$$p \sum_{i=0}^{\infty} i(1 - P(v \text{ har gjorts immun}|\mathrm{od}(v) = i))P(\mathrm{od}(v) = i) < 1$$

och mindre än 1 om

$$p\sum_{i=0}^{\infty} i(1 - P(v \text{ har gjorts immun}|\mathrm{od}(v) = i))P(\mathrm{od}(v) = i) \ge 1,$$

givet att epidemin inte har dött ut i någon tidssteg  $t \leq T$ .

## 3.3 Slumpmässig immunisering proportionell mot nodernas ingrader

För Immuniseringsstrategi 2, då immuniseringen görs proportionellt mot nodernas ingrader, är sannolikheten att en granne v till en infekterad nod är immun

$$p^{\star} := P(v \text{ har gjorts immun}) = \sum_{i=0}^{\infty} P(v \text{ har gjorts immun}|\mathrm{id}(v) = i)P(\mathrm{id}(v) = i),$$

för tidssteg t > T. Sannolikheten att en granne v till en infekterad nod är mottaglig är därför  $1 - p^*$ . Det är rimligt att anta att sannolikheten att en nod har smittats före tidssteg T + 1 beror på nodens ingrad, vilket skulle innebära att både det förväntade antalet infekterade noder som återhämtar sig genom immunisering och det förväntade antalet redan återhämtade noder som görs immuna till ingen nytta är större i detta fall än för övriga immuniseringsstrategier. På motsvarande sätt som för Immuniseringsstrategi 1 gäller, för tidssteg t > T, att sannolikheten att en granne  $v_1$  till en infekterad nod  $v_2$  blir smittad är

 $P(v_1 \text{ smittad av } v_2) = \underbrace{P(v_1 \text{ smittad av } v_2 | v_1 \text{ har gjorts immun})}_{\tilde{p} = 0} P(v_1 \text{ har gjorts immun})$ 

 $P(v_1 \text{ smittad av } v_2 | v_1 \text{ är mottaglig}) P(v_1 \text{ är mottaglig}) =$ 

$$p(1-p^{\star})$$

Antalet grannar till en smittsam nod v, i tidssteg t > T, är oberoende av nodernas ingrad och alltså oberoende av immuniseringen, och är därför lika med od(v). Sannolikhetsfunktionen för antalet noder L som blir smittade av vi tidssteg t givet att od(v) = y är därför

$$P(L = x | od(v) = y) = (p(1 - p^{\star}))^{x} (1 - p(1 - p^{\star}))^{y - x} {y \choose x},$$

vilket är sannolikhetsfunktionen för $\mathrm{Bin}(y,p(1-p^{\star}))\text{-}$ fördelningen. [7] Från detta följer att

$$E(L) = E(E(L|od(v))) = E(p(1-p^{\star})od(v)) = p(1-p^{\star})E(od(v)).$$

På motsvarande sätt som i föregående avsnitt ges väntevärdet för antalet noder  $S^{(t)}$  som smittas i tidssteg t > T av

$$E(S^{(t)}) = p^{t-1} (E(\mathrm{od}(v)))^{t-1} (1-p^{\star})^{t-T}$$

så länge smittspridningen befinner sig i ett tillräckligt tidigt stadium. Sannolikheten att epidemin dör ut är 1 om

$$pE(\mathrm{od}(v))(1-p^{\star}) < 1$$

och mindre än 1 om

$$pE(\mathrm{od}(v))(1-p^{\star}) \ge 1,$$

givet att epidemin inte har dött ut i någon tidssteg  $t \leq T$ .

## 3.4 Slumpmässig immunisering oberoende av nodernas utoch ingrader

För tidssteg t > T är sannolikheten att en granne till en smittad nod v är mottaglig (1-s), sannolikheten att en granne till en smittad nod v är immun är s, och sannolikheten att en infekterad nod återhämtar sig genom immunisering är s. Om t > T, så länge smittspridningen befinner sig i ett tillräckligt tidigt stadium, ges sannolikheten att en infekterad nod v har x mottagliga grannar, givet att dess utgrad är y, av

$$(1-s)^x s^{y-x} \begin{pmatrix} y \\ x \end{pmatrix},$$

vilket är sannolikhetsfunktionen för Bin(y, 1 - s)-fördelningen.[7] Från detta följer att det förväntade antalet mottagliga grannar till en smittsam nod v är (1-s)E(od(v)). På samma sätt ges sannolikheten att x mottagliga grannar blir

smittade av en smittsam nod v, givet att antalet mottagliga grannar till v är y, av

$$p^x(1-p)^{y-x}\binom{y}{x},$$

vilket är täthetsfunktionen för Bin(y, p)-fördelningen, [7] vilket medför att

$$E(L) = E(E(L| \text{ antalet mottagliga grannar till } v)) = p(1-s)E(\operatorname{od}(v)),$$

så länge epidemin befinner sig så pass tidigt stadium att utvecklingen av antalet smittade kan approximeras med en förgreningsprocess. På samma sätt som i tidigare avsnitt ges väntevärdet för antalet noder  $S^{(t)}$  som smittas i ett tidssteg t > T av

$$\begin{split} E(S^{(t)}) &= p^{T-1}(E(\mathrm{od}(v)))^{T-1}p^{t-T}(1-s)^{t-T}(E(\mathrm{od}(v)))^{t-T} = \\ p^{t-1}(1-s)^{t-T}(E(\mathrm{od}(v)))^{t-1}, \end{split}$$

och sannolikheten att epidemin dör ut är lika med 1 om

$$pE(\mathrm{od}(v))(1-s) < 1$$

och mindre än 1 om

$$pE(\mathrm{od}(v))(1-s) \ge 1,$$

givet att epidemin inte har dött ut i någon tidssteg  $t \leq T$ .

### 3.5 Jämförelse av de olika immuniseringsstrategierna

I Avsnitt 3.1–3.4 bestäms, för var och en av immuniseringsstrategierna, villkor som avgör när sannolikheten att epidemin dör ut innan den når ett stort utbrott är mindre än 1, eller ekvivalent när sannolikheten för ett stort utbrott är större än 0. De kritiska variablerna, det vill säga väntevärdet för antalet grannar Lsom en infekterad nod smittar, ges av

$$pE(\mathrm{od}(v))$$

för Immuniseringsstrategi 0 (ingen immunisering),

$$p\sum_{i=0}^{\infty} i(1 - P(v \text{ har gjorts immun}|\mathrm{od}(v) = i))P(\mathrm{od}(v) = i),$$

för Immuniseringsstrategi 1 (immunisering proportionell mot nodernas utgrader),

$$pE(\mathrm{od}(v))(1-p^{\star}), \quad p^{\star} = \sum_{i=0}^{\infty} P(v \text{ har gjorts immun}|\mathrm{id}(v)=i)P(\mathrm{id}(v)=i),$$

för Immuniseringsstrategi2 (immunisering proportionell mot nodernas ingrader) samt

$$pE(\mathrm{od}(v))(1-s)$$

för Immuniseringsstrategi 3 (immunisering oberoende av mot nodernas utgrader och ingrader), där  $v \in V$  är någon nod i tidssteg t > T (eller t > 0 för Immuniseringsstrategi 0) Då någon av dessa är större eller lika med 1 är sannolikheten för ett stort utbrott lika med 0 för motsvarande immuniseringsstrategi.[7] Eftersom  $0 \leq P(v \text{ har gjorts immun}|od(v) = i) < 1$ , i tidssteg t > T för Immuniseringsstrategi 1, för alla noder v och alla icke-negativa heltal i, gäller att

$$p\sum_{i=0}^{\infty} i(1 - P(v \text{ har gjorts immun}|od(v) = i))P(od(v) = i) < p\sum_{i=0}^{\infty} iP(od(v) = i) = pE(od(v)).$$

På liknande sätt, eftersom  $0 < p^* < 1$  och 0 < s < 1, i tidssteg t > T för Immuniseringsstrategi 2 respektive Immuniseringsstrategi 3, gäller att

$$pE(\mathrm{od}(v))(1-p^{\star}) < pE(\mathrm{od}(v))$$

 $\operatorname{och}$ 

$$pE(\mathrm{od}(v))(1-s) < pE(\mathrm{od}(v)).$$

Genom att anta att, i tidssteg t > T för Immuniseringsstrategi 2,

$$P(v \text{ har gjorts immun}|\mathrm{id}(v) = i) \approx \frac{sni}{\sum_{j \in V} \mathrm{id}(j)}$$

fås att

$$\frac{p^{\star}}{s} = \frac{\sum_{i=0}^{\infty} P(v \text{ har gjorts immun}|\mathrm{id}(v) = i)P(\mathrm{id}(v) = i)}{s} \approx \frac{\sum_{i=0}^{\infty} sniP(\mathrm{id}(v) = i)}{s\sum_{j \in V} \mathrm{id}(j)} = \frac{E(\mathrm{id}(v))}{\sum_{j \in V} \mathrm{id}(j)} \to \frac{E(\mathrm{id}(v))}{E(\mathrm{id}(v))} = 1,$$

då  $n \to \infty$ , enligt stora talens lag.[7] Detta medför att  $pE(od(v))(1 - p^*) \approx pE(od(v))(1 - s)$ , eftersom antagandet att  $n \to \infty$  görs. På motsvarande sätt, i tidssteg t > T för Immuniseringsstrategi 1, antas att

$$P(v \text{ har gjorts immun}|\mathrm{od}(v) = i) \approx \frac{sni}{\sum_{j \in V} \mathrm{od}(j)}$$

Insättning av detta i motsvarande kritiska variabel ger

$$p\sum_{i=0}^{\infty} i(1 - P(v \text{ har gjorts immun}|\mathrm{od}(v) = i))P(\mathrm{od}(v) = i) \approx$$

$$\begin{split} p \sum_{i=0}^{\infty} i(1 - \frac{sni}{\sum_{j \in V} \operatorname{od}(j)}) P(\operatorname{od}(v) = i) = \\ p \sum_{i=0}^{\infty} iP(\operatorname{od}(v) = i) - \frac{psn}{\sum_{j \in V} \operatorname{od}(j)} \sum_{i=0}^{\infty} i^2 P(\operatorname{od}(v) = i) = \\ pE(\operatorname{od}(v)) - \frac{psE((\operatorname{od}(v))^2)}{\sum_{j \in V} \operatorname{od}(j)} \to pE(\operatorname{od}(v)) - \frac{psE((\operatorname{od}(v))^2)}{E(\operatorname{od}(v))} = \\ pE(\operatorname{od}(v)) \left(1 - \frac{sE((\operatorname{od}(v))^2)}{(E(\operatorname{od}(v))^2)}\right) < pE(\operatorname{od}(v))(1 - s), \end{split}$$

då  $n \to \infty$ , enligt stora talens lag,[7] eftersom  $E((\mathrm{od}(v))^2) - (E(\mathrm{od}(v)))^2 = \operatorname{Var}(\mathrm{od}(v)) > 0$  för samtliga fördelningar som undersöks.[7] Eftersom antagandet att  $n \to \infty$  görs, följer av detta att stora epidemier kan förväntas inträffa oftare för Immuniseringsstrategi 1 än för övriga immuniseringsstrategier, medan de kan förväntas inträffa ungefär lika ofta för Immuniseringasstrategi 2 och 3.

## 4 Simulering

Simuleringen görs i programmet R och har programmerats i R-kod.[10] Programmet låter heltalen  $1, 2, \dots, n$  benämna noderna i en graf, slumpar fram nodernas ingrader och utgrader från givna fördelningar F respektive G, och skapar genom slumpmässig hopparning riktade kanter mellan noderna. Dessa sparas i en matris bestående av två kolonner där varje rad beskriver en riktad kant. Den första kolonnen innehåller det första elementet i de riktade kanterna och den andra kolonnen innehåller det andra elementet. Genom att ta bort alla loopar och multipla kanter från listan över kanter fås en lista som definierar en enkel graf som beskriver nätverket i modellen. Eftersom antalet noder n antas vara stort i modellen, samtidigt som det är tidskrävande att skapa många alltför stora nätverk, har bedömningen gjorts att det är lämpligast att låta  $n = 10\,000$  för samtliga simuleringar. För att programmet ska bli mindre tidskrävande vid skapandet av nätverket används R-paketet dplyr.[5] Simulering av smittspridningen görs enligt modellen, det vill säga genom att låta alla noder vara mottagliga från början, låta en nod smittas i tidssteg 1, och låta varje infekterad nod smitta sina mottagliga grannar med sannolikhet p i varje tidssteg som noden är smittsam. Om Immuniseringsstrategi 0 har valts kommer inga noder att göras immuna. Om Immuniseringsstrategi 1 har valts kommer  $s \cdot n$  noder att göras immuna proportionellt mot nodernas utgrad, vilket innebär att programmet går igenom följande steg:

- Steg 1: en lista  $\mathscr{L}$  över  $s \cdot n$  noder skapas och till varje position i  $\mathscr{L}$  väljs varje nod v med sannolikhet  $\frac{\operatorname{od}(v)}{\sum_{i=1}^{n} \operatorname{od}(i)}$ ,
- Steg 2: eventuella dubletter i  $\mathscr{L}$  ersätts med andra noder genom att varje nod v som ännu inte finns med i  $\mathscr{L}$  väljs med sannolikhet  $\frac{\mathrm{od}(v)}{\sum_{i=1}^{n} \mathrm{od}(i) - \sum_{j \in A} \mathrm{od}(j)}$ , där A är mängden av noder som redan har valts till  $\mathscr{L}$ ,
- Steg 3: steg 2 upprepas tills hela listan  $\mathscr L$  är fri från dubletter. Noderna i  $\mathscr L$  görs sedan immuna.

Motsvarande gäller för Immuniseringsstrategi 2 där noder i stället väljs proportionellt mot nodernas ingrader. Om Immuniseringsstrategi 3 har valts görs  $s \cdot n$  noder immuna, alla med samma sannolikhet s.

Maximala antalet tidssteg i simuleringen betecknas Y i programmet, och för samtliga simuleringar har värdet Y = 50 valts. Övriga parametrar för simuleringen är samma som i Tabell 1, Avsnitt 3.

### 4.1 Resultat

Graferna i detta avsnitt visar hur medelvärdet av slutstorleken S beror på smittsannolikheten p. Färgen på graferna anger vilken immuniseringsstrategi som har använts vid simuleringarna; röd graf betyder att Immuniseringsstrategi 0 har använts, grön graf betyder att Immuniseringsstrategi 1 har använts, blå graf betyder att Immuniseringsstrategi 2 har använts och svart graf betyder att Immuniseringsstrategi 3 har använts. De vänstra delfigurerna visar medelvärdena för simuleringar där antalet smittade överstiger h i någon tidpunkt medan de högra visar medelvärdena för samtliga simuleringar. För varje kombination av parametrar som undersöks har 20 simuleringar gjorts.

#### 4.1.1 Tungsvansade fördelningar med ändligt andramoment

I Figur 2 visar graferna medelvärdet av slutstorlekarna hos simuleringar där  $F = G = \text{Po}(Z), Z \sim \text{Pareto}(0.5, 2.7), n = 10\,000, h = 1000, \text{ och } s = 0.05.$  I detta fall är väntevärdet av ingraderna och utgraderna lika med  $\frac{2.7 \cdot 0.5}{2.7 - 1} \approx 0.79$  och gradfördelningarna har ändligt andramoment, eftersom Paretofördelningens andra parameter är större än 2. Figur 3 innehåller motsvarande grafer för h = 100 och har, till skillnad från fallet h = 1000, lägre medelvärde av slutstorlekarna för Immuniseringsstrategi 2 (blå graf) än för Immuniseringsstrategi 3 (svart graf), för samtliga värden på smittsannolikheten p. Denna tendens syns även i motsvarande figur för h = 10, Figur 14 i Appendix.



Figur 2: Diagram som visar hur medelslutstorleken beror på p, för modeller där  $F = G = \text{Po}(Z), Z \sim \text{Pareto}(0.5, 2.7), n = 10\,000, h = 1000$  och s = 0.05. Röd graf visar resultatet av simuleringar då I = 0, grön graf då I = 1, blå graf då I = 2 och svart graf då I = 3. Högra grafen visar medelvärdet för samtliga simuleringar (20 stycken för varje kombination av parametrar), medan vänstra grafen visar medevärdet av de slutstorlekar som överstiger h. Saknade värden i vänstra diagrammet beror på att inga slutstorlekar överstiger h för den kombinationen av parametrar.



Figur 3: Diagram som visar hur medelslutstorleken beror på p, för modeller där  $F = G = \text{Po}(Z), Z \sim \text{Pareto}(0.5, 2.7), n = 10\,000, h = 100$  och s = 0.05. Röd graf visar resultatet av simuleringar då I = 0, grön graf då I = 1, blå graf då I = 2 och svart graf då I = 3. Högra grafen visar medelvärdet för samtliga simuleringar (20 stycken för varje kombination av parametrar), medan vänstra grafen visar medevärdet av de slutstorlekar som överstiger h. Saknade värden i vänstra diagrammet beror på att inga slutstorlekar överstiger h för den kombinationen av parametrar.

I Figurer 4, 5 och 6 visas motsvarande grafer för simuleringar där i stället  $F = G = \text{Po}(Z), Z \sim \text{Pareto}(c, 2.7)$ , för värdena  $\frac{85}{108}$ , 1 respektive  $\frac{170}{81}$  på parametern c. Motsvarande väntevärden för in- och utgraderna är då  $\frac{5}{4} = 1.25$ ,  $\frac{27}{17} \approx 1.59$  respektive  $\frac{10}{3} \approx 3.33$ . Om  $c = \frac{85}{108}$  och p = 0.8 eller om  $c = \frac{170}{81}$  och p = 0.3 så gäller att pE(od(v)) = 1, för någon nod v i nätverket, därav valet av parametervärdena  $\frac{85}{108}$  och  $\frac{170}{81}$ . En tendens som tydligt framgår av graferna i dessa figurer är att slutstorlekarna för Immuniseringsstrategi 1 tenderar att skil-

ja sig mindre från slutstorlekarna för övriga immuniseringsstrategier för höga värden på parametern c, det vill säga för högre väntevärden för in- och utgraderna. Till skillnad från i Figur 1 verkar grafen för Immuniseringsstrategi 2 (den blå) tendera att ligga något högre än grafen för Immuniseringsstrategi 3 (den svarta), åtminstone för höga värden på smittsannolikheten p. I Appendix finns motsvarande figurer för  $c = \frac{85}{108}$  och  $c = \frac{170}{81}$ , Figurer 15–18, där i stället h = 100 eller h = 10. Liknande tendenser förekommer i dessa grafer som i motsvarande grafer för c = 0.5.



Figur 4: Diagram som visar hur medelslutstorleken beror på p, för modeller där  $F = G = \text{Po}(Z), Z \sim \text{Pareto}(\frac{85}{108}, 2.7), n = 10\,000, h = 1000$  och s = 0.05. Röd graf visar resultatet av simuleringar då I = 0, grön graf då I = 1, blå graf då I = 2 och svart graf då I = 3. Högra grafen visar medelvärdet för samtliga simuleringar (20 stycken för varje kombination av parametrar), medan vänstra grafen visar medevärdet av de slutstorlekar som överstiger h. Saknade värden i vänstra diagrammet beror på att inga slutstorlekar överstiger h för den kombinationen av parametrar.



Figur 5: Diagram som visar hur medelslutstorleken beror på p, för modeller där  $F = G = \text{Po}(Z), Z \sim \text{Pareto}(1, 2.7), n = 10\,000, h = 1000$  och s = 0.05. Röd graf visar resultatet av simuleringar då I = 0, grön graf då I = 1, blå graf då I = 2 och svart graf då I = 3. Högra grafen visar medelvärdet för samtliga simuleringar (20 stycken för varje kombination av parametrar), medan vänstra grafen visar medevärdet av de slutstorlekar som överstiger h. Saknade värden i vänstra diagrammet beror på att inga slutstorlekar överstiger h för den kombinationen av parametrar.



Figur 6: Diagram som visar hur medelslutstorleken beror på p, för modeller där  $F = G = \text{Po}(Z), Z \sim \text{Pareto}(\frac{170}{81}, 2.7), n = 10\,000, h = 1000$  och s = 0.05. Röd graf visar resultatet av simuleringar då I = 0, grön graf då I = 1, blå graf då I = 2 och svart graf då I = 3. Högra grafen visar medelvärdet för samtliga simuleringar (20 stycken för varje kombination av parametrar), medan vänstra grafen visar medevärdet av de slutstorlekar som överstiger h. Saknade värden i vänstra diagrammet beror på att inga slutstorlekar överstiger h för den kombinationen av parametrar.

#### 4.1.2 Tungsvansade fördelningar med oändligt andramoment

I Figur 7 visas medelvärdet av slutstorlekarna hos simuleringar där  $F = G = Po(Z), Z \sim Pareto(0.5, 1.7), n = 10\,000, h = 1000, s = 0.05$ . I detta fall är F och G likafördelade med oändligt andramoment, eftersom Paretofördelningens andra parameter är mindre än 2. Väntevärdet för in- och utgraderna är i det här fallet lika med  $\frac{0.5 \cdot 1.7}{0.7} \approx 1.21$ . Motsvarande grafer i Appendix där i stället  $F = Po(Z_1), Z_1 \sim Pareto(\frac{289}{378}, 2.7)$  (Figur 19) eller  $G = Po(Z_2), Z_2 \sim Pareto(\frac{289}{378}, 2.7)$ 

(Figur 20), där Paretofördelningens parametervärden har valts så att in- respektive utgraderna får ändligt andramoment och samma väntevärde som ovan, skiljer sig inte nämnvärt från Figur 7.



Figur 7: Diagram som visar hur medelslutstorleken beror på p, för modeller där  $F = G = \text{Po}(Z), Z \sim \text{Pareto}(0.5, 1.7), n = 10\,000, h = 1000$  och s = 0.05. Röd graf visar resultatet av simuleringar då I = 0, grön graf då I = 1, blå graf då I = 2 och svart graf då I = 3. Högra grafen visar medelvärdet för samtliga simuleringar (20 stycken för varje kombination av parametrar), medan vänstra grafen visar medevärdet av de slutstorlekar som överstiger h. Saknade värden i vänstra diagrammet beror på att inga slutstorlekar överstiger h för den kombinationen av parametrar.

#### 4.1.3 Icke-tungsvansade fördelningar

I Figur 8 visas medelvärdet av slutstorlekarna hos simuleringar där  $F = Po(Z), Z \sim$ Pareto $(\frac{170}{81}, 2.7), G = Ge(\frac{3}{13}), n = 10\,000, h = 10$  och s = 0.05. G är i detta fall icke-tungsvansad och både F och G har väntevärdet  $\frac{10}{3} \approx 3.33$ . I Appendix förekommer motsvarande figurer för h = 1000 (Figur 24) och h = 100 (Figur 25). Det är endats i fall som detta, då utgradernas fördelning är icke-tungsvansad samtidigt som ingradernas fördelning är tungsvansad, som slutstorleken för Immuniseringsstrategi 2 (blå graf) tenderar att vara mindre än för Immuniseringsstrategi 1 (grön graf).



Figur 8: Diagram som visar hur medelslutstorleken beror på p, för modeller där  $F = Po(Z), Z \sim Pareto(\frac{170}{81}, 2.7), G = Ge(\frac{3}{13}), n = 10\,000, h = 10$  och s = 0.05. Röd graf visar resultatet av simuleringar då I = 0, grön graf då I = 1, blå graf då I = 2 och svart graf då I = 3. Högra grafen visar medelvärdet för samtliga simuleringar (20 stycken för varje kombination av parametrar), medan vänstra grafen visar medevärdet av de slutstorlekar som överstiger h. Saknade värden i vänstra diagrammet beror på att inga slutstorlekar överstiger h för den kombinationen av parametrar.

I Figur 9 visas medelvärdet av slutstorlekarna hos simuleringar där  $F = \text{Ge}(\frac{3}{13}), G = \text{Po}(Z), Z \sim \text{Pareto}(\frac{170}{81}, 2.7), n = 10\,000, h = 10 \text{ och } s = 0.05.$  F är i detta fall icke-tungsvansad och både F och G har väntevärdet är  $\frac{10}{3} \approx 3.33$ , det vill säga samma som ovan. I Figur 26 i Appendix visas motsvarande grafer för h = 100.



Figur 9: Diagram som visar hur medelslutstorleken beror på p, för modeller där  $F = \operatorname{Ge}(\frac{3}{13})$ ,  $G = \operatorname{Po}(Z), Z \sim \operatorname{Pareto}(\frac{170}{81}, 2.7), n = 10\,000, h = 10$  och s = 0.05. Röd graf visar resultatet av simuleringar då I = 0, grön graf då I = 1, blå graf då I = 2 och svart graf då I = 3. Högra grafen visar medelvärdet för samtliga simuleringar (20 stycken för varje kombination av parametrar), medan vänstra grafen visar medevärdet av de slutstorlekar som överstiger h. Saknade värden i vänstra diagrammet beror på att inga slutstorlekar överstiger h för den kombinationen av parametrar.

I Figur 10 visas medelvärdet av slutstorlekarna hos simuleringar där  $F = G = \text{Ge}(\frac{3}{13})$ ,  $n = 10\,000$ , h = 10 och s = 0.05. I 27 i Appendix visas motsvarande grafer för h = 100. I detta fall skiljer sig medelvärdet av slutstorlekarna inte

mycket för Immuniseringsstrategi 1 (grön graf) och 2 (blå graf).



Figur 10: Diagram som visar hur medelslutstorleken beror på p, för modeller där  $F = G = \operatorname{Ge}(\frac{3}{13}), n = 10\,000, h = 10$  och s = 0.05. Röd graf visar resultatet av simuleringar då I = 0, grön graf då I = 1, blå graf då I = 2 och svart graf då I = 3. Högra grafen visar medelvärdet för samtliga simuleringar (20 stycken för varje kombination av parametrar), medan vänstra grafen visar medevärdet av de slutstorlekar som överstiger h. Saknade värden i vänstra diagrammet beror på att inga slutstorlekar överstiger h för den kombinationen av parametrar.

Figurer 21–23 i Appendix motsvarar Figurer 5–7, men har i stället den tungsvansade fördelningen  $Po(Z), Z \sim Pareto(\frac{85}{108}, 2.7)$  och den icke-tungsvansade fördelningen  $Ge(\frac{4}{9})$ . Gradfördelningarnas väntevärde är i detta fall  $\frac{5}{4} = 1.25$ 

#### 4.1.4 Olika värden på parametern s

Fallet då  $F = G = \text{Po}(Z), Z \sim \text{Pareto}(\frac{85}{108}, 2.7)$  och  $n = 10\,000$  har undersökts för olika värden på parametern s, det vill säga på andelen noder som görs immuna i tidssteg T. Figurer 11 och 12 visar hur medelvärdet av slutstorleken beror på smittsannolikheten då s = 0.1 för h = 1000 och h = 10. Figur 13 visar motsvarande grafer för fallet då s = 0.3 för h = 100. Figurer 28–36 i Appendix innehåller motsvarande grafer för olika värden på s och h. Tre tydliga tendenser framgår av figurerna i detta avsnitt och i motsvarande stycke i Appendix:

- 1. stora värden på s ger större skillnader mellan medelvärdet av slutstorlekarna för de olika immuniseringsstrategierna,
- 2. för små värden på h ligger grafen för Immuniseringsstrategi 2 (den blå grafen) över grafen Immuniseringsstrategi 1 (den gröna grafen), och under grafen för Immuniseringsstrategi 3 (den svarta grafen) för så gott som alla värden på smittsannolikheten p,
- 3. för stora värden på h ligger grafen för Immuniseringsstrategi 2 (den blå grafen) under grafen för Immuniseringsstrategi 3 (den svarta grafen) för

små värden på smittsannolikheten p, och över samma graf för stora värden på p.



Figur 11: Diagram som visar hur medelslutstorleken beror på p, för modeller där  $F = G = \text{Po}(Z), Z \sim \text{Pareto}(\frac{85}{108}, 2.7), n = 10\,000, h = 1000$  och s = 0.1. Röd graf visar resultatet av simuleringar då I = 0, grön graf då I = 1, blå graf då I = 2 och svart graf då I = 3. Högra grafen visar medelvärdet för samtliga simuleringar (20 stycken för varje kombination av parametrar), medan vänstra grafen visar medevärdet av de slutstorlekar som överstiger h. Saknade värden i vänstra diagrammet beror på att inga slutstorlekar överstiger h för den kombinationen av parametrar.



Figur 12: Diagram som visar hur medelslutstorleken beror på p, för modeller där  $F = G = \text{Po}(Z), Z \sim \text{Pareto}(\frac{85}{108}, 2.7), n = 10\,000, h = 10\text{ och } s = 0.1$ . Röd graf visar resultatet av simuleringar då I = 0, grön graf då I = 1, blå graf då I = 2 och svart graf då I = 3. Högra grafen visar medelvärdet för samtliga simuleringar (20 stycken för varje kombination av parametrar), medan vänstra grafen visar medevärdet av de slutstorlekar som överstiger h. Saknade värden i vänstra diagrammet beror på att inga slutstorlekar överstiger h för den kombinationen av parametrar.



Figur 13: Diagram som visar hur medelslutstorleken beror på p, för modeller där  $F = G = \text{Po}(Z), Z \sim \text{Pareto}(\frac{85}{108}, 2.7), n = 10\,000, h = 100$  och s = 0.3. Röd graf visar resultatet av simuleringar då I = 0, grön graf då I = 1, blå graf då I = 2 och svart graf då I = 3. Högra grafen visar medelvärdet för samtliga simuleringar (20 stycken för varje kombination av parametrar), medan vänstra grafen visar medevärdet av de slutstorlekar som överstiger h. Saknade värden i vänstra diagrammet beror på att inga slutstorlekar överstiger h för den kombinationen av parametrar.

## 5 Diskussion

I frågeställningen i inledningen lyder den första frågan:

• Vilken påverkan har de olika immuniseringsstrategierna på smittspridningen?

Det framgår tydligt av figurerna i Avsnitt 4 och i Appendix är att Immuniseringsstrategi 1, det vill säga immunisering proportionellt mot nodernas utgrader, är den strategi som resulterar i minst slutstorlek S för så gott som alla modeller med tungsvansade gradfördelningar som undersöks. Medelslutstorleken kan i vissa fall vara mindre än hälften så stor som motsvarande medelvärde för övriga immuniseringsstrategier. Detta är inte helt oväntat, eftersom resultatet i Avsnitt 3 antyder att Immuniseringsstrategi 1 medför en mindre risk för en stor epidemi än övriga immuniseringsstrategier.

Resultatet i Avsnitt 3 antyder även att risken för en stor epidemi inte skiljer sig för Immuniseringsstrategi 2 och 3, immunisering proportionellt mot nodernas ingrader respektive oberoende av in- och utgrader. Att många figurer i Avsnitt 4 och i Appendix i många fall, för låga värden på p och h, visar ett lägre medelvärde av slutstorlekarna för Immuniseringsstrategi 2, eller, för höga värden på p och h, ett högre medelvärde av slutstorlekarna för Immuniseringsstrategi 2, än för Immuniseringsstrategi 3, kan förklaras med att många noder med hög ingrad smittas tidigt under smittprocessen och att immunisering proportionellt mot ingrad därför innebär att många redan smittade noder görs immuna. Svaret på frågan ovan är alltså att Immuniseringsstrategi 1 har störst negativ påverkan på smittspridningen, det vill säga den innebär minst slutstorlek S, då nätverket har tungsvansade gradfördelningar, vilket är det mest realistiska fallet, medan Immuniseringsstrategi 2 och 3 har mindre negativ påverkan. Vilken av dessa två strategier som är att föredra beror på parametrarna p och h. Hur resultatet ser ut för andra gradfördelningar tas upp under nästa fråga från frågreställningen:

• Hur beror resultatet av immuniseringen på in- och utgradernas fördelningar *F* respektive *G*?

Av Figurerna i Avsnitt 4 och i Appendix framgår att immuniseringens påverkan då nätverkets gradfördelningar är tungsvansade (Poissonfördelningar med Paretofördelade parametrar) med ändliga andramoment, för samtliga immuniseringsstrategier (1-3), till stor del beror på gradfördelningarnas väntevärde. Ett större värde på Paretofördelningens första parameter c, som i sin tur medför ett större väntevärde, innebär mindre negativ påverkan. Ingen närmare undersökning av fördelningar med oändligt andramoment har gjorts och det är därför svårt att dra några avgörande slutsatser utifrån de få resultat som finns presenterade, mer än att Immuniseringsstrategi 1 har betydligt större negativ påverkan på smittspridningen än andra immuniseringsstrategier även i detta fall. För modeller där både ingradernas och utgradernas fördelningar är icke-tungsvansade verkar Imunniseringsstrategi 1 och Immuniseringsstrategi 2 vara lika effektiva och bättre än Immuniseringsstrategi 3. För nätverk där bara en av fördelningarna F och G är tungsvansad, vilket möjligen inte är särskilt realistiskt, verkar den immuniseringsstrategi där immuniseringen görs proportionellt mot motsvarande grad ha störst negativ påverkan. Frågeställningens tredje fråga lyder:

• Hur beror resultatet av immuniseringen på valet av parametern *h*, epidemins storlek då immuniseringen görs?

Det kan verka rimligt att på förhand förvänta sig att immunisering proportionellt mot nodernas ingrader har en särskilt stor negativ påverkan på smittspridningen om immuniseringen görs i ett tidigt stadium, det vill säga då h är litet, eftersom många noder med stor ingrad kan förväntas smittas tidigt under smittprocessen. Enligt de grafer som studeras verkar också parametern h vara avgörande för effekten av Immuniseringsstrategi 2, och möjligen även för effeken av Immuniseringsstrategier 1 och 3. Trots att inga uträkningar i Avsnitt 3 tyder på någon större eller mindre risk för ett stort utbrott för Immuniseringsstrategi 2 än för Immuniseringsstrategi 3 i tidssteg t > T, verkar slutstorleken ofta vara mindre för Immuniseringsstrategi 2 för små värden på h och för små värden på smittsannolikheten p, medan stora värden på både h och p däremot verkar innebära en större slutstorlek för Immuniseringsstrategi 2 än för Immuniseringsstrategi 3. En trolig förklaring till detta är att många noder med hög ingrad smittas tidigt under smittprocessen och att tidig immunisering proportionellt mot ingrad därför innebär att många som blir smittade i tidssteg Tockså görs immuna på samma gång och därför inte hinner föra smittan vidare. För stora värden på h och p är det däremot många med stor ingrad som redan har varit smittade sedan en tid tillbaka och därför har hunnit återhämta sig. Ett större antal sådana noder kan förväntas göras immuna till ingen nytta än för någon annan immuniseringsstrategi, vilket innebär att färre mottagliga eller infekterade noder görs immuna, vilket i sin tur leder till större slutstorlek än för övriga immuniseringsstrategier.

## Referenser

- T. Britton, M. Deijfen, A. Martin-Löf. Generating simple random graphs with prescribed degree distribution. 2006. Journal of Statistical Physics 124.6: 1377-1397.
- [2] A. S. Carstea, B. Grammaticos, A. Ramani, J. Satsuma, R. Willox. *Extending the SIR epidemic model*. 2004. Physica A: Statistical Mechanics and its Applications, 336(3), 369-375.
- [3] N. Chen, M. Olvera-Cravioto. Directed Random Graphs with Given Degree Distributions. 2012. Stochastic Systems 3.1: 147-186.
- [4] A. Clauset, M. E. J. Newman C. R. Shalizi. Power-law Distributions in Empirical Data. 2009. SIAM review 51.4 (2009): 661-703.
- R. Francois, H. Wickham. *dplyr: A Grammar of Data Manipulation.* 2015.
  R package version 0.4.1. http://CRAN.R-project.org/package=dplyr.
- [6] W. Gong, D. Towsley, C. C. Zou. *Email virus propagation modeling and analysis*. 2003. Department of Electrical and Computer Engineering, Univ. Massachusetts, Amherst, Technical Report: TR-CSE-03-04.
- [7] A. Gut. An Intermediate Course in Probability. 1995. Springer-Verlag.
- [8] R. van der Hofstad. Random Graphs and Complex Networks. Vol. I. 2016. Available on http://www.win.tue.nl/~rhofstad/NotesRGCN.pdf.
- M. E. J. Newman. Power laws, Pareto distributions and Zipf's law. 2005. Contemporary physics 46.5: 323-351.
- [10] R Core Team. R: A language and environment for statistical computing. 2014. R Foundation for Statistical Computing, Vienna, Austria. URL http: //www.R-project.org/.

## Appendix

Här presenteras de diagram som inte fick plats i Avsnitt 4. Diagrammen visar hur medelvärdet av slutstorleken för simuleringarna beror på smittsannolikheten p. De röda graferna visar medelvärdet av slutstorleken då immuniseringsstrategin I = 0, de gröna då I = 1, de blå då I = 2 och de svarta då I = 3. De vänstra diagrammen i varje figur visar medelvärdet av slutstorleken för de simuleringar då antalet smittade noder överstiger h, medan de högra diagrammen visar medelvärdet av samtliga simuleringar (totalt 20 stycken för varje kombination av parametrar). I varje simulering är antalet noder n = 10000.

## Tungsvansade fördelningar med ändligt andramoment



Figur 14: Diagram som visar hur medelslutstorleken beror på p, för modeller där  $F = G = \text{Po}(Z), Z \sim \text{Pareto}(0.5, 2.7), n = 10\,000, h = 10 \text{ och } s = 0.05$ . Röd graf visar resultatet av simuleringar då I = 0, grön graf då I = 1, blå graf då I = 2 och svart graf då I = 3. Högra grafen visar medelvärdet för samtliga simuleringar (20 stycken för varje kombination av parametrar), medan vänstra grafen visar medevärdet av de slutstorlekar som överstiger h. Saknade värden i vänstra diagrammet beror på att inga slutstorlekar överstiger h för motsvarande kombinationer av parametrar.



Figur 15: Diagram som visar hur medelslutstorleken beror på p, för modeller där  $F = G = \text{Po}(Z), Z \sim \text{Pareto}(\frac{85}{108}, 2.7), n = 10\,000, h = 100 \text{ och } s = 0.05$ . Röd graf visar resultatet av simuleringar då I = 0, grön graf då I = 1, blå graf då I = 2 och svart graf då I = 3. Högra grafen visar medelvärdet för samtliga simuleringar (20 stycken för varje kombination av parametrar), medan vänstra grafen visar medevärdet av de slutstorlekar som överstiger h. Saknade värden i vänstra diagrammet beror på att inga slutstorlekar överstiger h för den kombinationen av parametrar.



Figur 16: Diagram som visar hur medelslutstorleken beror på p, för modeller där  $F = G = \text{Po}(Z), Z \sim \text{Pareto}(\frac{85}{108}, 2.7), n = 10\,000, h = 10$  och s = 0.05. Röd graf visar resultatet av simuleringar då I = 0, grön graf då I = 1, blå graf då I = 2 och svart graf då I = 3. Högra grafen visar medelvärdet för samtliga simuleringar (20 stycken för varje kombination av parametrar), medan vänstra grafen visar medevärdet av de slutstorlekar som överstiger h. Saknade värden i vänstra diagrammet beror på att inga slutstorlekar överstiger h för den kombinationen av parametrar.



Figur 17: Diagram som visar hur medelslutstorleken beror på p, för modeller där  $F = G = \text{Po}(Z), Z \sim \text{Pareto}(\frac{170}{81}, 2.7), n = 10\,000, h = 100$  och s = 0.05. Röd graf visar resultatet av simuleringar då I = 0, grön graf då I = 1, blå graf då I = 2 och svart graf då I = 3. Högra grafen visar medelvärdet för samtliga simuleringar (20 stycken för varje kombination av parametrar), medan vänstra grafen visar medevärdet av de slutstorlekar som överstiger h. Saknade värden i vänstra diagrammet beror på att inga slutstorlekar överstiger h för den kombinationen av parametrar.



Figur 18: Diagram som visar hur medelslutstorleken beror på p, för modeller där  $F = G = \text{Po}(Z), Z \sim \text{Pareto}(\frac{170}{81}, 2.7), n = 10\,000, h = 10 \text{ och } s = 0.05$ . Röd graf visar resultatet av simuleringar då I = 0, grön graf då I = 1, blå graf då I = 2 och svart graf då I = 3. Högra grafen visar medelvärdet för samtliga simuleringar (20 stycken för varje kombination av parametrar), medan vänstra grafen visar medevärdet av de slutstorlekar som överstiger h. Saknade värden i vänstra diagrammet beror på att inga slutstorlekar överstiger h för den kombinationen av parametrar.

### Tungsvansade fördelningar med oändligt andramoment



Figur 19: Diagram som visar hur medelslutstorleken beror på p, för modeller där  $F = \text{Po}(Z_1), Z_1 \sim \text{Pareto}(\frac{289}{378}, 2.7), G = \text{Po}(Z_2), Z_2 \sim \text{Pareto}(0.5, 1.7), n = 10\,000, h = 1000 \text{ och } s = 0.05$ . Röd graf visar resultatet av simuleringar då I = 0, grön graf då I = 1, blå graf då I = 2 och svart graf då I = 3. Högra grafen visar medelvärdet för samtliga simuleringar (20 stycken för varje kombination av parametrar), medan vänstra grafen visar medevärdet av de slutstorlekar som överstiger h. Saknade värden i vänstra diagrammet beror på att inga slutstorlekar överstiger h för den kombinationen av parametrar.



Figur 20: Diagram som visar hur medelslutstorleken beror på p, för modeller där  $F = \text{Po}(Z_1), Z_1 \sim \text{Pareto}(0.5, 1.7), G = \text{Po}(Z_2), Z_2 \sim \text{Pareto}(\frac{289}{378}, 2.7), n = 10\,000, h = 1000$  och s = 0.05. Röd graf visar resultatet av simuleringar då I = 0, grön graf då I = 1, blå graf då I = 2 och svart graf då I = 3. Högra grafen visar medelvärdet för samtliga simuleringar (20 stycken för varje kombination av parametrar), medan vänstra grafen visar medevärdet av de slutstorlekar som överstiger h. Saknade värden i vänstra diagrammet beror på att inga slutstorlekar överstiger h för den kombinationen av parametrar.

#### Icke-tungsvansade fördelningar



Figur 21: Diagram som visar hur medelslutstorleken beror på p, för modeller där  $F = \text{Po}(Z), Z \sim \text{Pareto}(\frac{85}{108}, 2.7), G = \text{Ge}(\frac{4}{9}), n = 10\,000, h = 10$  och s = 0.05. Röd graf visar resultatet av simuleringar då I = 0, grön graf då I = 1, blå graf då I = 2 och svart graf då I = 3. Högra grafen visar medelvärdet för samtliga simuleringar (20 stycken för varje kombination av parametrar), medan vänstra grafen visar medevärdet av de slutstorlekar som överstiger h. Saknade värden i vänstra diagrammet beror på att inga slutstorlekar överstiger h för den kombinationen av parametrar.



Figur 22: Diagram som visar hur medelslutstorleken beror på p, för modeller där  $F = \operatorname{Ge}(\frac{4}{9}), G = \operatorname{Po}(Z), Z \sim \operatorname{Pareto}(\frac{85}{108}, 2.7), n = 10\,000, h = 10$  och s = 0.05. Röd graf visar resultatet av simuleringar då I = 0, grön graf då I = 1, blå graf då I = 2 och svart graf då I = 3. Högra grafen visar medelvärdet för samtliga simuleringar (20 stycken för varje kombination av parametrar), medan vänstra grafen visar medevärdet av de slutstorlekar som överstiger h. Saknade värden i vänstra diagrammet beror på att inga slutstorlekar överstiger h för den kombinationen av parametrar.



Figur 23: Diagram som visar hur medelslutstorleken beror på p, för modeller där  $F = G = \operatorname{Ge}(\frac{4}{9}), n = 10\,000, h = 10$  och s = 0.05. Röd graf visar resultatet av simuleringar då I = 0, grön graf då I = 1, blå graf då I = 2 och svart graf då I = 3. Högra grafen visar medelvärdet för samtliga simuleringar (20 stycken för varje kombination av parametrar), medan vänstra grafen visar medevärdet av de slutstorlekar som överstiger h. Saknade värden i vänstra diagrammet beror på att inga slutstorlekar överstiger h för den kombinationen av parametrar.



Figur 24: Diagram som visar hur medelslutstorleken beror på p, för modeller där  $F = \text{Po}(Z), Z \sim \text{Pareto}(\frac{170}{81}, 2.7), G = \text{Ge}(\frac{3}{13}), n = 10\,000, h = 1000 \text{ och } s = 0.05$ . Röd graf visar resultatet av simuleringar då I = 0, grön graf då I = 1, blå graf då I = 2 och svart graf då I = 3. Högra grafen visar medelvärdet för samtliga simuleringar (20 stycken för varje kombination av parametrar), medan vänstra grafen visar medevärdet av de slutstorlekar som överstiger h. Saknade värden i vänstra diagrammet beror på att inga slutstorlekar överstiger h för den kombinationen av parametrar.



Figur 25: Diagram som visar hur medelslutstorleken beror på p, för modeller där  $F = \text{Po}(Z), Z \sim \text{Pareto}(\frac{170}{81}, 2.7), G = \text{Ge}(\frac{3}{13}), n = 10\,000, h = 100$  och s = 0.05. Röd graf visar resultatet av simuleringar då I = 0, grön graf då I = 1, blå graf då I = 2 och svart graf då I = 3. Högra grafen visar medelvärdet för samtliga simuleringar (20 stycken för varje kombination av parametrar), medan vänstra grafen visar medevärdet av de slutstorlekar som överstiger h. Saknade värden i vänstra diagrammet beror på att inga slutstorlekar överstiger h för den kombinationen av parametrar.



Figur 26: Diagram som visar hur medelslutstorleken beror på p, för modeller där  $F = \operatorname{Ge}(\frac{3}{13}), G = \operatorname{Po}(Z), Z \sim \operatorname{Pareto}(\frac{170}{81}, 2.7), n = 10\,000, h = 100$  och s = 0.05. Röd graf visar resultatet av simuleringar då I = 0, grön graf då I = 1, blå graf då I = 2 och svart graf då I = 3. Högra grafen visar medelvärdet för samtliga simuleringar (20 stycken för varje kombination av parametrar), medan vänstra grafen visar medevärdet av de slutstorlekar som överstiger h. Saknade värden i vänstra diagrammet beror på att inga slutstorlekar överstiger h för den kombinationen av parametrar.



Figur 27: Diagram som visar hur medelslutstorleken beror på p, för modeller där  $F = G = \text{Ge}(\frac{3}{13}), n = 10\,000, h = 100$  och s = 0.05. Röd graf visar resultatet av simuleringar då I = 0, grön graf då I = 1, blå graf då I = 2 och svart graf då I = 3. Högra grafen visar medelvärdet för samtliga simuleringar (20 stycken för varje kombination av parametrar), medan vänstra grafen visar medevärdet av de slutstorlekar som överstiger h. Saknade värden i vänstra diagrammet beror på att inga slutstorlekar överstiger h för den kombinationen av parametrar.

## Olika värden på parametern s



Figur 28: Diagram som visar hur medelslutstorleken beror på p, för modeller där  $F = G = \text{Po}(Z), Z \sim \text{Pareto}(\frac{85}{108}, 2.7), n = 10\,000, h = 100$  och s = 0.1. Röd graf visar resultatet av simuleringar då I = 0, grön graf då I = 1, blå graf då I = 2 och svart graf då I = 3. Högra grafen visar medelvärdet för samtliga simuleringar (20 stycken för varje kombination av parametrar), medan vänstra grafen visar medevärdet av de slutstorlekar som överstiger h. Saknade värden i vänstra diagrammet beror på att inga slutstorlekar överstiger h för den kombinationen av parametrar.



Figur 29: Diagram som visar hur medelslutstorleken beror på p, för modeller där  $F = G = \text{Po}(Z), Z \sim \text{Pareto}(\frac{85}{108}, 2.7), n = 10\,000, h = 1000$  och s = 0.15. Röd graf visar resultatet av simuleringar då I = 0, grön graf då I = 1, blå graf då I = 2 och svart graf då I = 3. Högra grafen visar medelvärdet för samtliga simuleringar (20 stycken för varje kombination av parametrar), medan vänstra grafen visar medevärdet av de slutstorlekar som överstiger h. Saknade värden i vänstra diagrammet beror på att inga slutstorlekar överstiger h för den kombinationen av parametrar.



Figur 30: Diagram som visar hur medelslutstorleken beror på p, för modeller där  $F = G = \text{Po}(Z), Z \sim \text{Pareto}(\frac{85}{108}, 2.7), n = 10\,000, h = 100$  och s = 0.15. Röd graf visar resultatet av simuleringar då I = 0, grön graf då I = 1, blå graf då I = 2 och svart graf då I = 3. Högra grafen visar medelvärdet för samtliga simuleringar (20 stycken för varje kombination av parametrar), medan vänstra grafen visar medevärdet av de slutstorlekar som överstiger h. Saknade värden i vänstra diagrammet beror på att inga slutstorlekar överstiger h för den kombinationen av parametrar.



Figur 31: Diagram som visar hur medelslutstorleken beror på p, för modeller där  $F = G = \text{Po}(Z), Z \sim \text{Pareto}(\frac{85}{108}, 2.7), n = 10\,000, h = 10$  och s = 0.15. Röd graf visar resultatet av simuleringar då I = 0, grön graf då I = 1, blå graf då I = 2 och svart graf då I = 3. Högra grafen visar medelvärdet för samtliga simuleringar (20 stycken för varje kombination av parametrar), medan vänstra grafen visar medevärdet av de slutstorlekar som överstiger h. Saknade värden i vänstra diagrammet beror på att inga slutstorlekar överstiger h för den kombinationen av parametrar.



Figur 32: Diagram som visar hur medelslutstorleken beror på p, för modeller där  $F = G = \text{Po}(Z), Z \sim \text{Pareto}(\frac{85}{108}, 2.7), n = 10\,000, h = 1000$  och s = 0.2. Röd graf visar resultatet av simuleringar då I = 0, grön graf då I = 1, blå graf då I = 2 och svart graf då I = 3. Högra grafen visar medelvärdet för samtliga simuleringar (20 stycken för varje kombination av parametrar), medan vänstra grafen visar medevärdet av de slutstorlekar som överstiger h. Saknade värden i vänstra diagrammet beror på att inga slutstorlekar överstiger h för den kombinationen av parametrar.



Figur 33: Diagram som visar hur medelslutstorleken beror på p, för modeller där  $F = G = \text{Po}(Z), Z \sim \text{Pareto}(\frac{85}{108}, 2.7), n = 10\,000, h = 100$  och s = 0.2. Röd graf visar resultatet av simuleringar då I = 0, grön graf då I = 1, blå graf då I = 2 och svart graf då I = 3. Högra grafen visar medelvärdet för samtliga simuleringar (20 stycken för varje kombination av parametrar), medan vänstra grafen visar medevärdet av de slutstorlekar som överstiger h. Saknade värden i vänstra diagrammet beror på att inga slutstorlekar överstiger h för den kombinationen av parametrar.



Figur 34: Diagram som visar hur medelslutstorleken beror på p, för modeller där  $F = G = \text{Po}(Z), Z \sim \text{Pareto}(\frac{85}{108}, 2.7), n = 10\,000, h = 10\text{ och } s = 0.2$ . Röd graf visar resultatet av simuleringar då I = 0, grön graf då I = 1, blå graf då I = 2 och svart graf då I = 3. Högra grafen visar medelvärdet för samtliga simuleringar (20 stycken för varje kombination av parametrar), medan vänstra grafen visar medevärdet av de slutstorlekar som överstiger h. Saknade värden i vänstra diagrammet beror på att inga slutstorlekar överstiger h för den kombinationen av parametrar.



Figur 35: Diagram som visar hur medelslutstorleken beror på p, för modeller där  $F = G = \text{Po}(Z), Z \sim \text{Pareto}(\frac{85}{108}, 2.7), n = 10\,000, h = 1000$  och s = 0.3. Röd graf visar resultatet av simuleringar då I = 0, grön graf då I = 1, blå graf då I = 2 och svart graf då I = 3. Högra grafen visar medelvärdet för samtliga simuleringar (20 stycken för varje kombination av parametrar), medan vänstra grafen visar medevärdet av de slutstorlekar som överstiger h. Saknade värden i vänstra diagrammet beror på att inga slutstorlekar överstiger h för den kombinationen av parametrar.



Figur 36: Diagram som visar hur medelslutstorleken beror på p, för modeller där  $F = G = \text{Po}(Z), Z \sim \text{Pareto}(\frac{85}{108}, 2.7), n = 10\,000, h = 10\text{ och } s = 0.3$ . Röd graf visar resultatet av simuleringar då I = 0, grön graf då I = 1, blå graf då I = 2 och svart graf då I = 3. Högra grafen visar medelvärdet för samtliga simuleringar (20 stycken för varje kombination av parametrar), medan vänstra grafen visar medevärdet av de slutstorlekar som överstiger h. Saknade värden i vänstra diagrammet beror på att inga slutstorlekar överstiger h för den kombinationen av parametrar.