



UNIVERSITÀ  
DEGLI STUDI  
DI PADOVA



**DIPARTIMENTO DI INGEGNERIA DELL'INFORMAZIONE**

**CORSO DI LAUREA IN INGEGNERIA DELL'INFORMAZIONE**

**“MODELLI EPIDEMIOLOGICI APPLICATI ALLA DIFFUSIONE DI  
CONTENUTI VIRALI IN RETE”**

**Relatore: Leonardo Badia**

**Tesi di laurea di: Filippo Simonetti**

**ANNO ACCADEMICO 2020 – 2021**

**Data di laurea 23/09/2021**



## **ABSTRACT**

In questa tesi vengono presentati e analizzati alcuni modelli epidemiologici applicati alla diffusione di contenuti virali, nello specifico meme, all'interno della rete Internet. Per prima cosa viene introdotto un sistema compartimentale non lineare basato sul modello SIR. Vengono quindi riportati alcuni risultati matematici notevoli relativi al modello, supportati da simulazioni MATLAB Simulink. Il modello viene in seguito utilizzato per descrivere l'andamento temporale dei dati di ricerca su Google di alcune parole chiave legate a diversi meme, con particolare attenzione a meme di interesse economico e politico; i dati utilizzati a tale scopo sono stati ottenuti tramite Google Trends. Sono infine presentati alcuni modelli compartimentali più complessi, derivati dal modello SIR a partire da ipotesi aggiuntive congruenti con i tipici meccanismi di diffusione di meme all'interno della rete Internet. Viene contestualmente riportata un'analisi di tali modelli, validata tramite alcuni esempi ottenuti mediante simulazioni MATLAB Simulink.



# INDICE

<b>Abstract</b>	1
<b>1. Capitolo 1: Introduzione</b>	5
<b>2. Capitolo 2: Il modello SIR</b>	7
2.1. Introduzione al modello SIR	7
2.2. Numero riproduttivo di base	9
2.3. Risultati del modello SIR	10
<b>3. Capitolo 3: Applicazioni del SIR alla diffusione di meme</b>	13
3.1. Applicazione del modello SIR a dati reali	13
3.2. Una prima modifica al modello SIR: effetto Fallout	14
3.3. Viral marketing e discussioni politiche	16
3.4. Altre applicazioni di modelli epidemiologici	19
<b>4. Capitolo 4: Modifiche al modello SIR</b>	21
4.1. Suscettibilità della classe dei Rimossi	21
4.2. Demografia e rimozione alternativa degli Infetti	24
4.3. Perdita di memoria dei Rimossi	27
4.4. Persuasione alla non condivisione	30
<b>5. Capitolo 5: Conclusione</b>	37
<b>Bibliografia</b>	39



# CAPITOLO 1: Introduzione

Il termine “meme” fu inizialmente coniato nel 1976 dal biologo Richard Dawkins. Nel suo libro “The Selfish Gene” [1] Dawkins definì il meme come un’unità culturale in grado di autoreplicarsi, in possesso di tre caratteristiche tali da determinarne il successo: longevità, fecondità e fedeltà di copia. Nello specifico “meme” deriva dal greco *μίμημα*, *mimema*, ovvero “imitazione”, termine poi ridotto appunto a “meme” per analogia con “gene”. Dawkins infatti sosteneva che esattamente come i geni si propagano da corpo a corpo tramite la riproduzione, così i meme si propagano da mente a mente, tramite un processo comunemente chiamato imitazione. Questa definizione, sebbene sia quella più nota e tipicamente più citata, è stata rivisitata nel corso degli anni per mano di vari autori, tra i quali lo stesso Dawkins. Nel 1986 in “The Blind Watchmaker” [2] e in seguito nel 1993 in “Viruses of the Mind” [3], Dawkins ritrattò la precedente definizione, adottando una prospettiva di carattere virale ed epidemiologico. Ad avvalorare questa tesi, basti pensare che un meme si diffonde di fatto per contatto tra individui ad esso interessati ed individui suscettibili all’influenza dei primi, esattamente come un virus si propaga ai sani a partire dai soggetti infetti.

Nel 2013 C. M. C. Díaz ha raccolto nella sua ricerca [4] i principali contributi all’evoluzione di tale definizione, allo scopo di fornire una caratterizzazione il più possibile coerente e aggiornata di Internet Meme (IM), confrontando le trattazioni teoriche offerte da numerosi studiosi nel corso degli anni e le definizioni di uso comune adottate su Internet dagli utenti stessi. Secondo Díaz viene quindi definito Internet Meme “un’unità di informazione (idea, concetto o convinzione), la quale si replica passando per via Internet, in forma di hyper-link, video, immagine o frase”. Esattamente come fatto da Dawkins nelle sue ultime opere, Díaz abbraccia esplicitamente nel suo lavoro la trattazione di carattere epidemiologico degli Internet Meme.

Considerata quindi la sopracitata analogia tra meme e virus, l’obiettivo di questa tesi è quello di raccogliere una selezione di modelli epidemiologici applicati alla diffusione di contenuti virali, nello specifico meme, all’interno della rete Internet. Per cominciare verrà presentato un primo modello epidemiologico, noto come modello SIR (sane-infected-recovery) [5]. A partire da tale modello sarà possibile ricavare, tramite l’introduzione di nuove ipotesi, modelli sempre più complessi, allo scopo di fornire una panoramica delle meccaniche tipicamente adottate nello studio della diffusione di meme [21, 22, 23, 24]. A partire dalle stesse fonti, verrà quindi riportata un’analisi dei modelli presentati, utile a poterne comprendere il comportamento al variare dei parametri utilizzati. I risultati ottenuti verranno poi validati tramite simulazioni eseguite con MATLAB Simulink.

I modelli qui presentati potrebbero essere sfruttati per predire l’andamento della diffusione di meme in Internet, con valide applicazioni in campi quali il marketing [11, 12, 13] e la politica [13].

Il resto di questa tesi è quindi organizzato come segue. Nel Capitolo 2 verrà presentato il modello SIR, ovvero un semplice modello compartimentale non lineare. Verranno in seguito riportati alcuni risultati di particolare interesse riguardanti il modello, accompagnati da alcune simulazioni MATLAB Simulink. Nel Capitolo 3 i risultati ottenuti dal modello SIR verranno confrontati con alcuni dati reali riguardanti il numero di ricerche su Google di specifiche parole chiave legate ad alcuni meme. Saranno poi approfondite le applicazioni a marketing e politica dei risultati ottenuti. Nel Capitolo 4 verranno infine presentati alcuni modelli più complessi, derivati dal modello SIR. Verrà successivamente riportata un'analisi del comportamento dei modelli proposti al variare dei parametri utilizzati. I risultati ottenuti saranno validati da ulteriori simulazioni ottenute tramite MATLAB Simulink.



# CAPITOLO 2: Il modello SIR

## 2.1 Introduzione al modello SIR

Allo scopo di studiare la diffusione di meme in Internet verranno adottati ed analizzati in questo elaborato il modello SIR, inizialmente introdotto da Kermack e McKendrick nel 1927 [5], e alcuni modelli da esso derivati. Il modello SIR si propone di descrivere l'evoluzione di un'epidemia tramite un sistema di equazioni differenziali di primo ordine. Tali equazioni sono associate alla variazione del numero di individui presenti nelle seguenti tre classi:

- **Suscettibili:** sono gli individui che possono essere infettati dall'ente patogeno. Conservando l'analogia meme-virus, la classe dei Suscettibili contiene gli individui che potrebbero essere influenzati dal meme, ovvero coloro che ancora non vi sono interessati, ma potrebbero provare interesse nel caso in cui venissero esposti. Nel presente lavoro questa classe verrà indicata come  $S$ ;
- **Infetti:** si tratta degli individui che contraggono l'ente patogeno, vale a dire, nell'ambito di studio di questo elaborato, coloro che sono entrati in contatto con il meme e hanno deciso di partecipare alla sua trasmissione, condividendolo a loro volta. Questa classe verrà qui indicata con  $I$ ;
- **Rimossi:** questa classe contiene i soggetti che non possono partecipare alla diffusione del patogeno, ovvero quelli che hanno perso interesse nel meme dopo un'esposizione e non cercano di condividerlo o diffonderlo. Tale classe verrà indicata con  $R$ .

L'unione di queste tre classi rappresenta di fatto l'intera popolazione ospite. Per semplicità di notazione, considerati  $X, Y, Z$  rispettivamente il numero assoluto di individui Suscettibili, Infetti e Rimossi e indicando con  $N$  il numero totale di individui nella popolazione, verranno qui utilizzati i termini  $S, I, R$  come segue:

$$S = X/N, \quad I = Y/N, \quad R = Z/N$$

Risulta quindi:

$$S + I + R = 1 \tag{2.1}$$

Il modello SIR si basa sull'ipotesi che la trasmissione del patogeno, o del meme in questo caso, avvenga tramite contatto tra un individuo suscettibile e uno infetto. A tal proposito, viene assunto che la probabilità di contatto tra due individui qualsiasi della popolazione sia uguale per tutti, ovvero che l'evoluzione dell'epidemia sia identica all'interno dell'intera popolazione. Questa seconda ipotesi in particolare può risultare irrealistica se si considera la normale rete di contatti presente all'interno di

una società o di un social network, tuttavia permette di fornire una descrizione valida e funzionale dell'evolversi dell'epidemia. Sarebbe possibile rilassare tale ipotesi adottando modelli sostanzialmente diversi dal modello SIR. A tale scopo si potrebbe pensare di adattare alla diffusione dei meme il *cellular automata model* proposto da M.A. Fuentes e M.N. Kuperman per lo studio di epidemie con dipendenza spaziale [6]. Ulteriori ipotesi preliminari qui adottate prevedono che le costanti del modello rimangano uguali nel tempo, così come la quantità  $N$  di individui all'interno della popolazione ospite; quest'ultima assunzione verrà rilassata in seguito.

Viste quindi le premesse, viene di seguito presentato il sistema di equazioni differenziali che costituisce il modello SIR [7]:

$$\frac{dS}{dt} = -\beta S I \quad (2.2)$$

$$\frac{dI}{dt} = \beta S I - \gamma I \quad (2.3)$$

$$\frac{dR}{dt} = \gamma I \quad (2.4)$$

Considerando, come anticipato, una popolazione di taglia costante  $N$ , è possibile introdurre un ulteriore vincolo:

$$\frac{dS}{dt} + \frac{dI}{dt} + \frac{dR}{dt} = 0 \quad (2.5)$$

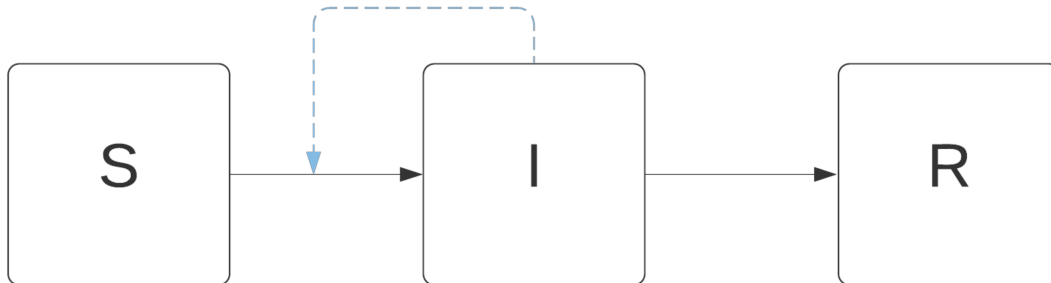
Noto questo vincolo, il modello può essere riassunto da due sole delle tre equazioni del sistema.

Il modello viene poi completato dalle condizioni iniziali poste sulle classi: inizialmente la classe dei Suscettibili conterrà quasi tutta la popolazione ospite, alla classe degli Infetti apparterrà la piccola porzione di popolazione che darà inizio all'epidemia, ovvero gli autori del meme o i primi a dividerlo, mentre la classe dei Rimossi sarà vuota. Le condizioni iniziali del modello sono quindi le seguenti:

$$S(0) \approx 1, \quad I(0) = 1 - S(0) \approx 0, \quad R(0) = 0 \quad (2.6)$$

Il termine  $\beta I$  presente nelle equazioni (2.2) e (2.3) indica il tasso *per capita* con il quale gli individui Suscettibili contraggono l'infezione. Nuovi Infetti sono quindi prodotti con tasso  $\beta S I$ . Verosimilmente  $\beta$  dipenderà sia dalla probabilità di contatto tra individui delle classi di Suscettibili e Infetti, sia dalla probabilità che da un tale contatto derivi una nuova infezione; sarà quindi determinato dalle caratteristiche dell'ente patogeno e da quelle della popolazione ospite. Il parametro  $\gamma$ , noto come *removal rate*, indica invece il tasso di recupero dall'infezione; il suo reciproco  $1/\gamma$  rappresenta quindi il periodo medio di permanenza nella classe degli Infetti.

Il modello appena descritto si presta ad essere graficamente rappresentato tramite un diagramma di flusso. Tale diagramma utilizza delle frecce nere per indicare i flussi tra le classi. La freccia azzurra tratteggiata che collega la classe degli Infetti al flusso tra Suscettibili e Infetti indica la dipendenza che il valore della frazione di Infetti ha sul tasso di infezione:



Sebbene il modello non possa essere risolto esplicitamente, dal momento che non è possibile ottenere una formula analitica che esprima l'andamento di S o di I nel tempo, è comunque possibile risolverlo per via numerica. Possono inoltre essere ricavate dalle equazioni del modello alcune interessanti relazioni che permettono di indagare la diffusione del meme.

## 2.2 Numero riproduttivo di base

Osservando le equazioni del modello è possibile per prima cosa comprendere se la diffusione del meme avrà successo o se viceversa fallirà ancor prima di iniziare. A questo proposito si consideri l'equazione (2.3) nella seguente forma:

$$\frac{dI}{dt} = I (\beta S - \gamma) \quad (2.7)$$

Risulta evidente, ricordando che la frazione degli Infetti sarà senz'altro non negativa, che se inizialmente  $S(0) < \gamma/\beta$  allora la diffusione del meme non sarà possibile, dal momento che  $dI/dt < 0$ . Questo risultato, ricavato per la prima volta dagli stessi Kermack e McKendrick (1927) è noto come *effetto soglia*: la popolazione ospite deve infatti superare una certa soglia critica perché il meme possa diffondersi. Si consideri ora la quantità  $\beta/\gamma$  emersa nella precedente analisi: tale valore è noto come *numero riproduttivo di base*, comunemente indicato come  $R_0$ , il quale rappresenta di fatto il numero medio di casi secondari dovuti ad caso primario in una popolazione interamente suscettibile. Ricordando le condizioni iniziali poste sul modello, la relazione dell'effetto soglia può essere riscritta come segue: considerando una popolazione completamente suscettibile, il meme sarà in grado di diffondersi solo se:

$$R_0 > 1 \quad (2.8)$$

Considerando la definizione intuitiva precedentemente data per il numero riproduttivo di base, questo significa, come intuibile, che il meme potrà diffondersi solo se in media ogni persona interessata ad esso riuscirà a dividerlo suscitando l'interesse di più di un nuovo soggetto all'interno della popolazione.

Ovviamente tale considerazione sulla variazione del numero di Infetti può essere estesa ad ogni istante dell'evolversi del fenomeno epidemiologico: considerando un istante  $t > 0$  è semplice osservare che il numero di Infetti crescerà ( $\frac{dI}{dt} > 0$ ) solo se  $R_0 S(t) > 1$ .

### 2.3 Risultati del modello SIR

Sebbene, come anticipato, il modello SIR non permetta di ricavare una soluzione analitica che descriva l'andamento delle classi nel tempo, risulta comunque possibile risolvere il modello per via numerica. A tale scopo è stato qui utilizzato MATLAB Simulink, adottando un risolutore a passo variabile Dormand-Prince ODE45. Il risultato di tale simulazione, rappresentativo del tipico andamento temporale di un fenomeno epidemico, è riportato in Figura 1.

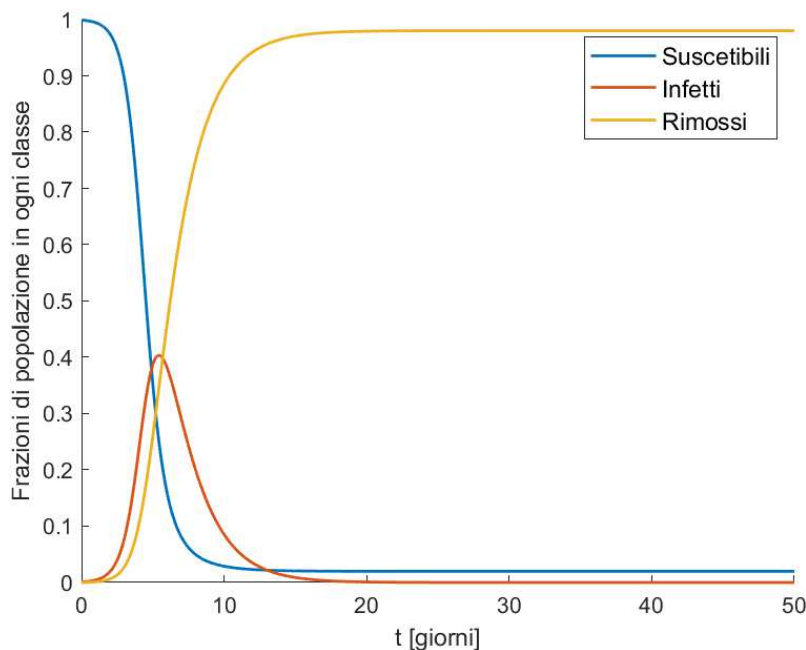


Figura 1: andamento temporale delle classi del modello SIR con  $R_0 = 4$

Risulta inoltre interessante osservare l'andamento del numero di Infetti in funzione del numero di Suscetibili durante lo svolgersi dell'epidemia. Tale relazione è ricavabile [7] a partire dalle equazioni del modello, dividendo l'equazione (2.3) per la (2.2)

$$\frac{dI}{dS} = \frac{\gamma}{\beta S} - 1$$

e integrando, con  $S(0) = 1$ , rispetto ad  $S$ :

$$I = \frac{1}{R_0} \ln S - S + 1 \quad (2.9)$$

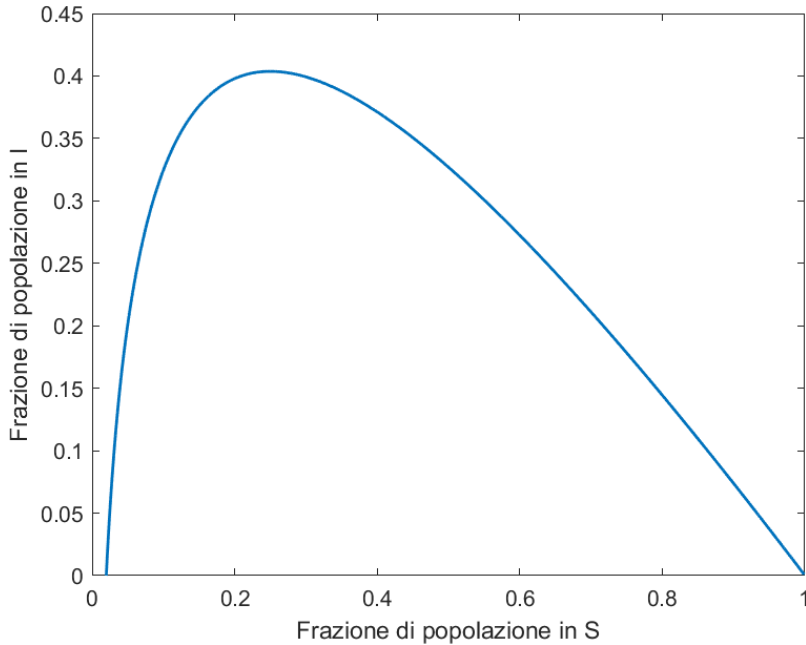


Figura 2: frazione di Infetti in funzione della frazione di Suscettibili con  $R_0 = 4$

Osservando la Figura 2, nella quale è riportata graficamente la relazione tra Infetti e Suscettibili espressa in (2.9), è possibile apprezzare come il numero di Suscettibili, inizialmente costituito dall'intera popolazione ospite ( $S(0) = 1$ ), diminuisca durante l'intera durata dell'epidemia, la quale termina con l'azzerarsi del numero di Infetti. Potrebbe invece stupire notare che la frazione di Suscettibili non si azzerava con il finire dell'epidemia. Tale risultato, valido in generale, è dimostrabile rigorosamente esprimendo  $S$  in funzione di  $R$ . Dividendo l'equazione (2.2) per la (2.4)

$$\frac{dS}{dR} = -R_0 S$$

e integrando rispetto ad  $R$  si ottiene infatti:

$$S = S(0) e^{-R R_0} \quad (2.10)$$

utilizzando come di consueto le condizioni iniziali espresse in precedenza. Dal momento che  $e^{-R R_0} > 0$ , risulta evidente che il numero di Suscettibili sarà sempre strettamente positivo a sua volta. Si può quindi concludere che inevitabilmente ci sarà sempre qualche suscettibile nella popolazione ospite che non si interesserà mai al meme. La catena di condivisioni di conseguenza si interromperà sempre per mancanza di Infetti, non per mancanza di Suscettibili. In Figura 3 tuttavia è possibile osservare che all'aumentare di  $R_0$  la porzione di popolazione che rimane nella classe dei

Suscettibili al termine dell'epidemia diminuisce drasticamente; indicativamente per  $R_0 > 5$  l'intera popolazione finisce per entrare a contatto con il meme [7].

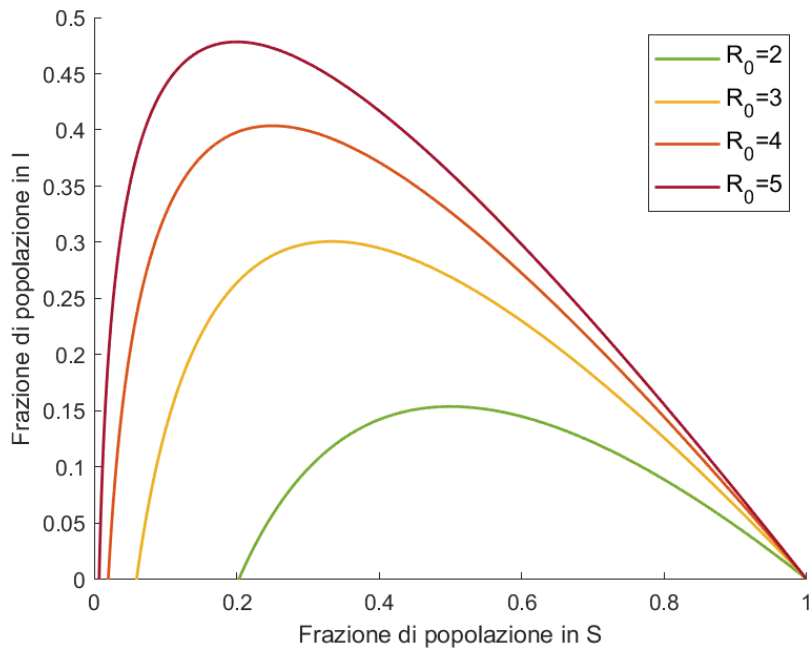


Figura 3: frazione di Infetti in funzione della frazione di Suscettibili dati diversi valori di  $R_0$

Dalla relazione (2.10) è inoltre possibile ricavare la porzione di persone  $R(\infty)$  all'interno della popolazione che si sono interessate al meme durante la sua diffusione:

$$S(\infty) = 1 - R(\infty) = S(0) e^{-R(\infty)R_0} \implies 1 - R(\infty) - S(0) e^{-R(\infty)R_0} = 0 \quad (2.11)$$

Tale equazione tuttavia è trascendente per cui non è possibile ricavare una soluzione esatta.

## CAPITOLO 3: Applicazioni del SIR alla diffusione di meme

### 3.1 Applicazione del modello SIR a dati reali

Il modello SIR finora esposto, per quanto semplice, vanta già alcune applicazioni dirette allo studio della diffusione di meme. Per cominciare, potrebbe risultare interessante verificare se il modello SIR sia effettivamente in grado di approssimare in modo soddisfacente reali fenomeni virali avvenuti all'interno della rete Internet. Di particolare rilievo in questo senso è il lavoro di S.S.S. Peri [8] et al., nel quale il modello SIR viene sfruttato per analizzare la diffusione di specifici *hashtag* su Twitter. Interessante risulta anche lo studio presentato da Perissi et al. [9] nel quale è mostrata un'incoraggiante coincidenza tra i dati di ricerca su Google di vari meme virali e l'andamento tipico di fenomeni epidemiologici descritti dal modello SIR. I dati raccolti da Perissi et al. sono stati ricavati tramite Google Trends [29], un utile strumento in grado di fornire uno storico delle ricerche di specifiche stringhe all'interno di Google a partire dal 2004. Vista la difficoltà riscontrabile nell'ottenere i dati di condivisione di meme all'interno dei vari social network, la possibilità di consultare i dati di ricerca di Google rappresenta senz'altro un'ottima alternativa, se si tiene in considerazione che gli utenti che tendono a cercare un certo meme in rete sono necessariamente interessati – e quindi Infetti – rispetto al meme stesso; si assume quindi che la frequenza delle ricerche di una specifica stringa correlata al meme in analisi sia proporzionale al numero di Infetti presenti nella popolazione ospite.

Sulla falsariga del lavoro svolto da Perissi et al. [9], vengono qui riportati alcuni dati di ricerca raccolti tramite Google Trends, relativi a due fenomeni virali che hanno interessato la rete Internet negli anni passati. Il primo tra questi riguarda la diffusione di “Gangnam Style”, famosa hit del rapper sudcoreano Psy, pubblicata a Luglio 2012; i dati relativi a questo evento sono stati raccolti in relazione alla ricerca della stringa “Gangnam Style” su YouTube, da Luglio 2012 fino ad Agosto 2021, con cadenza mensile.

Il secondo fenomeno osservato concerne la circolazione in Internet di un estratto di un'intervista a Matteo Renzi, noto politico italiano, pubblicata dall'emittente americana BBC News nel 2016; la frase più celebre estrapolata da tale intervista, nonché la stringa qui utilizzata per la raccolta dei dati, è “first reaction shock”, riferita da Renzi in merito ai risultati del referendum del Regno Unito sulla permanenza nell'Unione Europea. Curiosamente questo breve video è divenuto virale solo più tardi, a partire da Gennaio 2021; i dati a riguardo sono stati raccolti quindi da tale data fino ad Agosto 2021, con frequenza settimanale, considerando le sole ricerche eseguite da utenti Italiani. Nell'analizzare i dati raccolti è bene tenere a mente che Google Trends non fornisce l'effettivo numero di ricerche legate ad una determinata stringa, bensì un *search volume index* (SVI), ossia un valore normalizzato

dei dati di ricerca, così presentato per agevolare il confronto tra le caratteristiche di dati relativi a stringhe diverse.

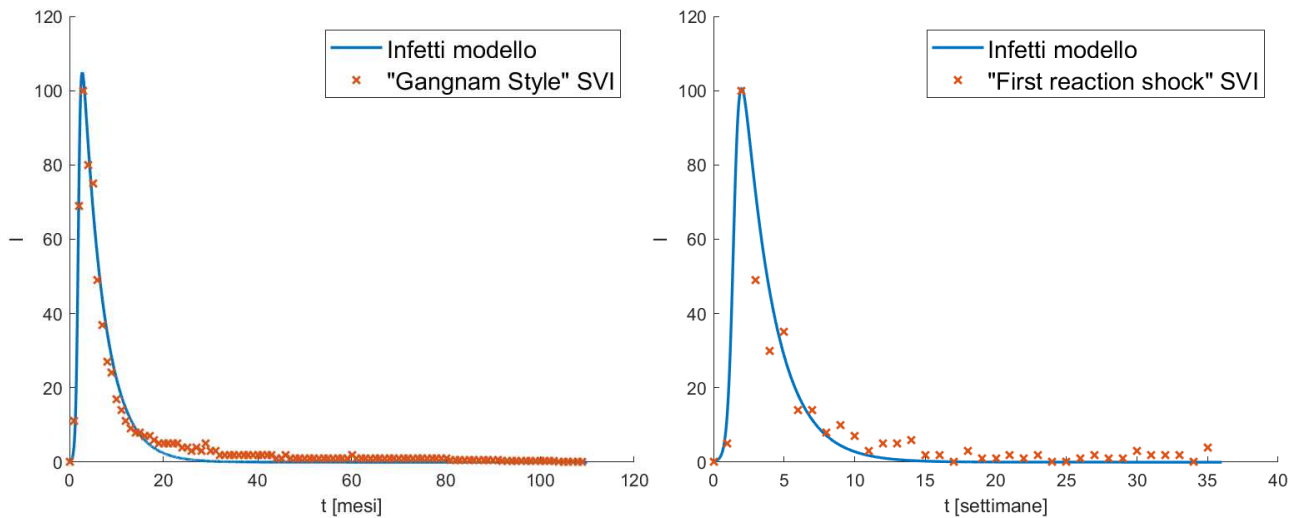


Figura 4: a sinistra il SVI relativo a “Gangnam Style”, confrontato con una simulazione Simulink basata sul modello SIR con  $\beta = 3.8494$ ,  $\gamma = 0.2226$ ,  $N = 135$ ; a destra il SVI di “First reaction shock” in Italia, confrontato con una simulazione analoga con parametri  $\beta = 5.1358$ ,  $\gamma = 0.4637$ ,  $N = 145$

Va sottolineato che i dati raccolti tramite Google Trends, i quali riportano le sole ricerche di parole chiave all’interno del motore di ricerca, così come la diffusione degli *hashtag* di Twitter, rappresentano solo una piccola parte, per quanto indicativa, del fenomeno di diffusione dei meme in Internet, i quali vengono condivisi sia privatamente, all’interno delle chat di portali come Whatsapp, Discord e altri, sia pubblicamente, in bacheche come quelle di Facebook, TikTok, Instagram e simili; i dati relativi alla diffusione di meme all’interno di tali piattaforme sono però, come anticipato, inevitabilmente difficili da reperire.

### 3.2 Una prima modifica al modello SIR: Effetto Fallout

Come osservato nel lavoro precedentemente citato [9] può accadere che l’andamento dei dati relativi alla diffusione di specifici meme raccolti tramite Google Trends non si presti ad essere descritto tramite il modello SIR adottato fino a questo momento. Tuttavia, in alcuni di questi casi i dati in questione possono apparire molto simili agli esempi sopra citati; nello specifico si registra fin da subito un rapido picco di interesse, privo però del punto di flesso tipicamente presente nelle curve degli Infetti descritte dal classico modello SIR. Per tentare di descrivere fenomeni di questo tipo Perissi et al. presentano quindi una prima interessante modifica al modello SIR. In particolare si ipotizza che il comportamento appena descritto sia dovuto alla diffusione di meme inizialmente lanciati da mass media (TV, web channel, ...). Alla luce di questa ipotesi, nel nuovo modello viene conservata la struttura compartimentale tipica del SIR e vengono mantenute inalterate nella loro

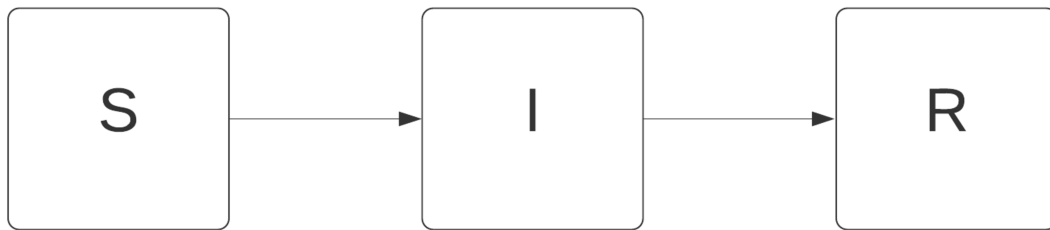


funzione e interpretazione le tre classi di Suscettibili, Infetti e Rimossi descritte come in precedenza; viene però eliminata l'influenza dettata dal numero di Infetti nel modello sul tasso di infezione degli individui Suscettibili. Il fenomeno di recupero resta invece invariato rispetto al modello SIR. Il modello viene quindi adattato come segue:

$$\frac{dS}{dt} = -\beta S \quad (3.1)$$

$$\frac{dI}{dt} = \beta S - \gamma I \quad (3.2)$$

Si ottiene di conseguenza quello che in [9] viene chiamato *effetto fallout*, in analogia con la diffusione di una nube radioattiva: la condivisione infatti non parte da un singolo individuo infetto, bensì coinvolge in massa i Suscettibili tramite un ente esterno, senza necessità di contatto tra Suscettibili e Infetti. Il seguente diagramma di flusso descrive graficamente la struttura del nuovo modello:



In Figura 5 viene riportato il risultato di una simulazione dell'effetto fallout, dove può essere osservata l'assenza del punto di flesso normalmente riscontrabile utilizzando il modello SIR nella fase di crescita del numero di Infetti.

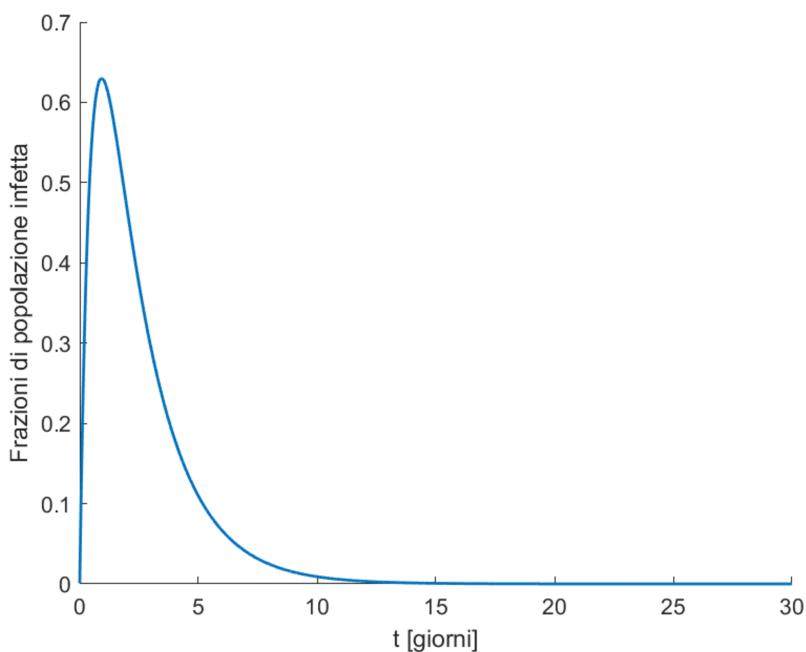


Figura 5: simulazione Simulink nella quale è osservabile l'effetto fallout, basata sul modello descritto dalle equazioni (3.1) e (3.2)

Un valido esempio di questo comportamento, come già osservato da Perissi et al., può essere riscontrato nell'andamento del *search volume index* del termine “Bataclan”, il quale si innalzò enormemente nel Novembre del 2015 in seguito alla diffusione delle notizie relative all’attentato di matrice islamica avvenuto nel teatro omonimo di Parigi. In Figura 6 vengono mostrati i dati forniti da Google Trends relativi al termine “Bataclan” dal primo Novembre 2015 a Natale 2016, raccolti con cadenza settimanale, confrontati con una simulazione MATLAB Simulink basata sul modello descritto in questo paragrafo.

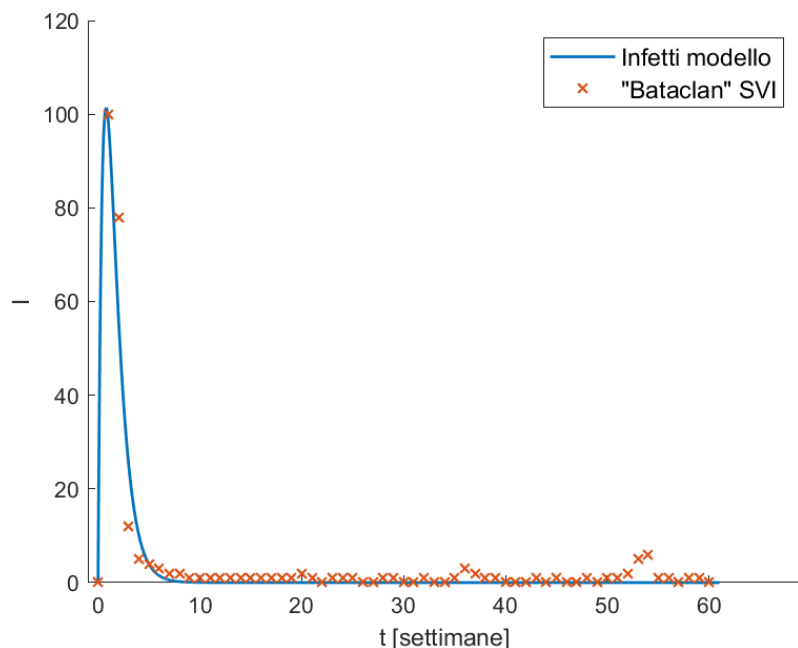


Figura 6: dati relativi al SVI di “Bataclan”, comparati con una simulazione del modello ad effetto fallout con  $\beta = 0.9999$ ,  $\gamma = 1.561$ ,  $N = 350$

### 3.3 Viral marketing e discussioni politiche

Vista la rapidità tramite la quale un’informazione, una notizia, o un meme in generale, possono diffondersi all’interno della rete Internet, risulta evidente come la possibilità di modellare fenomeni simili sia di notevole interesse pratico. In particolare numerosi studi [10, 11, 12, 13] hanno osservato come sia possibile sfruttare il modello SIR – nonché modelli da esso derivati – per analizzare e talvolta prevedere la diffusione di specifici temi di interesse politico, potenzialmente capaci di influenzare l’esito di votazioni ed elezioni, e l’evoluzione di campagne di *viral marketing*, con evidenti conseguenze di natura economica.

Il fenomeno del *viral marketing* nello specifico, anche noto come *network-enhanced word of mouth*, fu inizialmente analizzato da Steve Jurvetson [10], il quale coniò il termine in seguito all’analisi di una campagna promozionale di Hotmail del 1996. In quegli anni il noto servizio di posta elettronica

inseriva infatti un link promozionale in ogni messaggio inviato da un account Hotmail, rendendo in tal modo gli utenti suoi involontari promotori. In generale, è opportuno parlare di *viral marketing* riferendosi ad ogni campagna pubblicitaria nella quale sono gli utenti stessi a promuovere il prodotto o il servizio in questione. Con gli anni strategie simili si sono diffuse enormemente, dal momento che presentano numerosi punti di forza, tra i quali la necessità di un investimento spesso inferiore rispetto alle tradizionali campagne pubblicitarie, una maggior velocità di diffusione e un'efficace autoselezione del target [10, 11]. Ad oggi numerose aziende utilizzano questo strumento, tra le quali, per citarne alcune di particolare rilievo, Microsoft, BMW e Samsung [11].

Visto l'evidente interesse suscitato dal tema, Rodrigues et al. [11] hanno proposto, come anticipato, di studiare il fenomeno del *viral marketing* tramite il modello SIR, basandosi su premesse analoghe a quelle adottate in questa tesi per l'analisi della diffusione di una più ampia classe di meme. A supporto di questo approccio vengono di seguito presentati, in Figura 7, due esempi di meme legati a prodotti o servizi, i quali si sono diffusi seguendo il tipico andamento degli Infetti del modello SIR. Il primo tra questi riguarda il successo dei “fidget spinner”, giocattoli simili a trottole divenuti popolari nella primavera del 2017 anche grazie alla diffusione in Internet di video tutorial ad essi dedicati. I dati a riguardo sono stati ancora una volta raccolti tramite Google Trends, con cadenza settimanale a partire da Febbraio 2017 fino a Dicembre 2018. La seconda raccolta di dati qui mostrata è relativa alla diffusione del termine “Android”, ovvero il nome del noto sistema operativo per dispositivi mobili sviluppato da Google, ad oggi uno tra i sistemi operativi per smartphone più diffusi al mondo [30]. I dati sono raccolti a partire da Novembre 2007 – Android fu presentato il 5 Novembre 2007 – fino ad oggi, con cadenza mensile. È interessante osservare che la diffusione del termine “Android” è tutt'altro che conclusa, sebbene il fenomeno epidemico ad esso associata si trovi evidentemente nella fase di diminuzione del numero di Infetti.

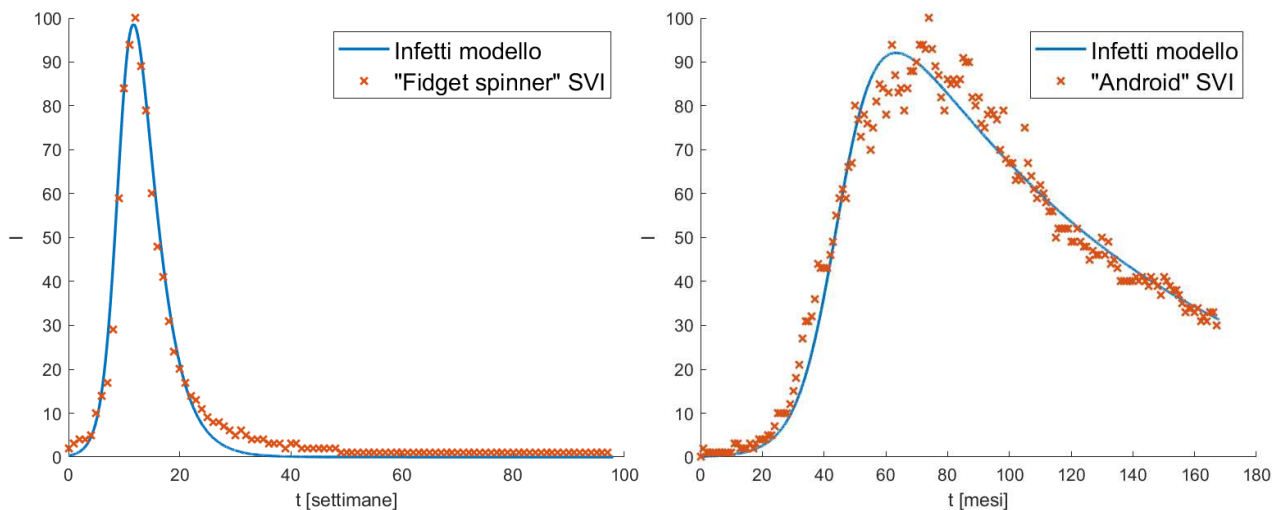


Figura 7: a sinistra i dati relativi al SVI di "Fidget spinner", confrontati con una simulazione del modello SIR con parametri  $\beta = 0.9851$ ,  $\gamma = 0.3302$ ,  $N = 330$ ; a destra i dati raccolti da Google Trends per le ricerche di "Android", confrontati con l'andamento degli infetti del modello SIR posti  $\beta = 0.1642$ ,  $\gamma = 0.0112$ ,  $N = 123$

È bene ricordare che non conoscendo le politiche di marketing delle varie aziende coinvolte nei fenomeni di cui sopra è impossibile sapere se siano state effettivamente adottate campagne di *viral marketing* durante la diffusione dei prodotti interessati, tuttavia il successo di questi ultimi ha inevitabilmente seguito meccanismi analoghi a quelli sfruttati da strategie di marketing simili.

Jiarong Li [12], diversamente da quanto fatto da Rodrigues et al., ha adottato per lo studio dei fenomeni di *viral marketing* un modello derivato dal SIR finora utilizzato. Tale modello nello specifico divide in due classi la popolazione infetta, distinguendo i normali infetti da coloro attivamente interessati alla diffusione del prodotto, per esempio in vista di un qualche premio offerto dall'azienda, e in modo analogo separa in due diverse classi gli individui Rimossi che hanno acquistato il prodotto e quelli che invece sono semplicemente divenuti immuni all'informazione d'interesse. Partendo dall'analisi di tale modello, Li propone poi alcune strategie per migliorare la resa della campagna di marketing in questione: tra queste spiccano la possibilità di incentivare l'interazione tra gli utenti e i propri contatti, e la necessità di mantenere attivo l'interesse della popolazione tramite promozioni proposte frequentemente.

Secondo principi analoghi a quelli considerati finora nell'applicare il modello SIR all'evoluzione di campagne di marketing, Woo et al. hanno esteso la trattazione a meme riguardanti argomenti di natura politica [13]. Questo tipo di approccio permette di comprendere e prevedere l'interesse della popolazione per determinati temi di rilevanza politica e sociale; sarebbe ad esempio possibile intuire quali argomenti possano appassionare gli elettori in vista di campagne elettorali imminenti, o presagire quale candidato possa risultare favorito alle prossime elezioni. Va però osservato che l'attenzione della popolazione per temi politicamente sensibili è spesso influenzata da notizie o fatti di attualità. Per questo motivo è sensato assumere che l'utilizzo di parametri tempo invarianti per il

modello possa rivelarsi inefficace per simili applicazioni. In questi casi può risultare preferibile ridurre l'intervallo di tempo nel quale viene analizzato il fenomeno virale in questione, in modo tale da poter assumere ragionevolmente, come fatto finora, che i parametri del modello siano costanti nel tempo.

Ancora una volta, i dati forniti da Google Trends permettono di mostrare un calzante esempio di quanto detto: in Figura 8 è possibile osservare l'andamento del *search volume index* del termine "Afghanistan" registrato negli Stati Uniti nella settimana tra il 13 Agosto e il 20 Agosto del 2021. Il picco di ricerche relative a tale termine è presumibilmente connesso alla notizia della presa di Kabul da parte dei Talebani, risalente a Domenica 15 Agosto, evento fortemente legato alle politiche estere degli Stati Uniti, ritirati dal territorio afgano a partire da Maggio 2021 [31].

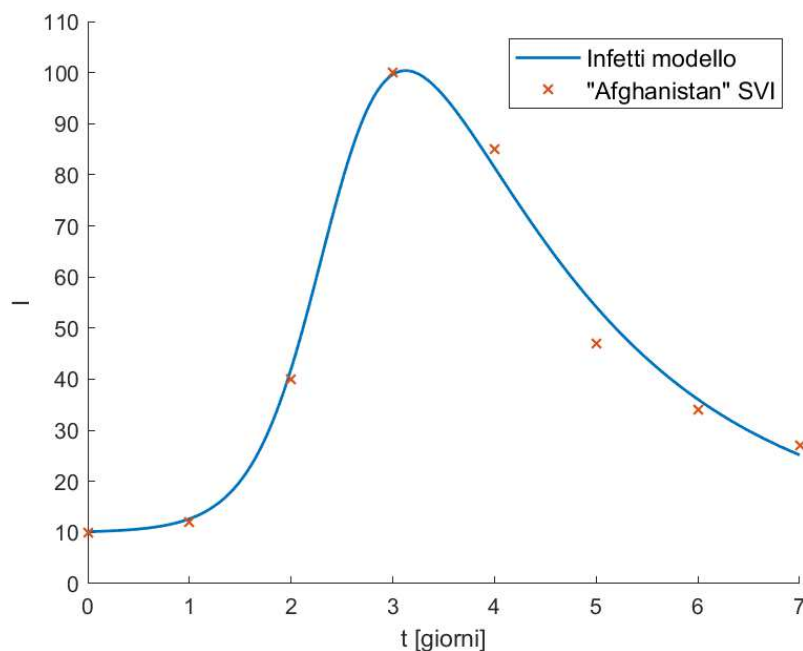


Figura 8: dati raccolti da Google Trends relativi al SVI di "Afghanistan", comparati con una simulazione basata sul modello SIR con parametri  $\beta = 3.3402$ ,  $\gamma = 0.5576$ ,  $N = 169$

### 3.4 Altre applicazioni di modelli epidemiologici

Vari modelli epidemiologici, analoghi a quelli presentati in questa tesi per lo studio della diffusione di meme, sono stati utilizzati con successo nella letteratura scientifica anche in contesti diversi da quello di interesse di questo elaborato. Tra le applicazioni più importanti spiccano senz'altro quelle alla diffusione di virus informatici. Noeiaghdam ad esempio ha sfruttato il modello SIR per lo studio dei computer virus [14]. Similmente, Piqueira et al. [15] hanno adottato per l'analisi dei computer virus un modello SAIR, vale a dire un modello simile al SIR, al quale è stata aggiunta una classe A contenente i nodi della rete equipaggiati con un programma anti-virus efficace. Gao et al. [16] hanno

invece utilizzato il SIR nel definire un modello di rete su due livelli, il primo legato alla rete di natura geografica costituita dalle stazioni radio base, il secondo corrispondente alla rete logica formata dai vari contatti telefonici. Tale modello permette di descrivere efficacemente la diffusione di mobile virus basati su contatti *bluetooth* e *SMS*.

I modelli epidemiologici possono poi essere applicati efficacemente anche in altri ambiti legati alla cyber security. Mishra e Keshri nella loro ricerca [17] hanno adottato un modello SEIRS-V (Suscettibili-Esposti-Infetti-Rimossi-Vaccinati) per analizzare la diffusione di worm informatici all'interno di *wireless sensor network* (WSN). Cisotto e Badia hanno invece applicato le relazioni del SIR all'interno di un modello *cellular automata*, allo scopo di descrivere attacchi informatici a danno di *smart grids* [18].

Un ulteriore esempio di applicazione di modelli epidemiologici è fornito dal lavoro di Yang et al. [19], i quali hanno proposto di modellare tramite diverse strutture compartimentali derivate dal modello SIR la diffusione di file in reti di condivisione *peer-to-peer* (P2P). Zhang e Xiao infine hanno sfruttato un modello *cellular automata* allo scopo di comprendere il fenomeno di polarizzazione della popolazione rispetto a due opinioni contrastanti all'interno di gruppi di discussione in Internet [20].

Gli esempi sopracitati mostrano quindi l'estrema versatilità dei modelli epidemiologici e dei metodi adottati in questa tesi, i quali si rivelano efficaci non solo nella descrizione di fenomeni epidemici veri e propri, ma anche in numerosi ambiti indubbiamente distanti da quello medico-sanitario, tra i quali la sicurezza informatica e lo studio della diffusione di elementi virali in rete.

# CAPITOLO 4: Modifiche al modello SIR

## 4.1 Suscettibilità della classe dei Rimossi

In questo capitolo verranno presentati e analizzati alcuni modelli epidemiologici derivati dal modello SIR, ideati per lo studio della diffusione virale di meme in Internet. Il primo tra questi, proposto da Lin Wang e Brendan C. Wood [21], introduce una nuova ipotesi secondo la quale i soggetti Rimossi sarebbero ancora in condizione tale da poter provare interesse per il meme. Tali individui potrebbero quindi diventare Infetti, in seguito ad un contatto con un altro soggetto Infetto, sebbene con una probabilità di trasmissione minore rispetto a quella riscontrabile per individui appartenenti alla classe dei Suscettibili. A differenza di quanto fatto durante la trattazione del modello SIR, verranno da qui indicate con  $S$ ,  $I$  e  $R$  le quantità rispettivamente di individui Suscettibili, Infetti e Rimossi. Dato  $N$  il numero di individui nella popolazione ospite risulta quindi:

$$S + I + R = N \tag{4.1}$$

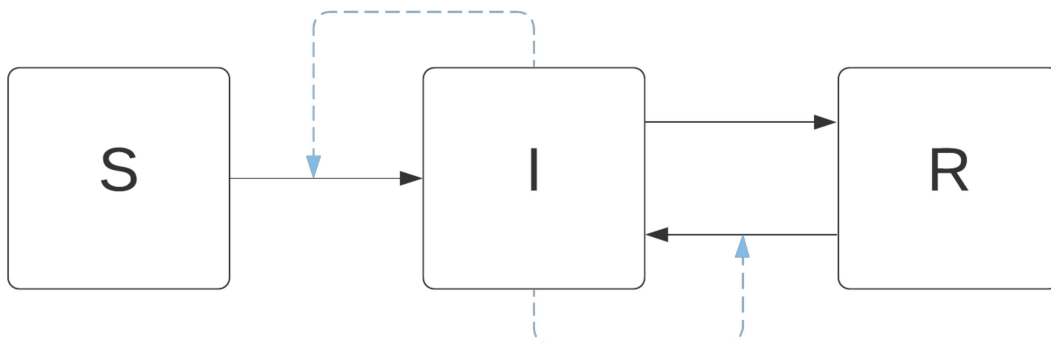
Le equazioni che descrivono il comportamento del nuovo modello sono quindi definite come segue:

$$\frac{dS}{dt} = -\alpha S I \tag{4.2}$$

$$\frac{dI}{dt} = \alpha S I + \beta R I - \gamma I \tag{4.3}$$

$$\frac{dR}{dt} = -\beta R I + \gamma I \tag{4.4}$$

Il termine  $\alpha I$ , come detto per il modello SIR, indica il tasso *per capita* con il quale gli individui Suscettibili si interessano al meme. Analogamente, il termine  $\beta I$ , ragionevolmente con  $\beta < \alpha$ , rappresenta il tasso *per capita* con il quale i Rimossi riacquisiscono interesse per il meme, tornando quindi ad essere Infetti. Il modello è rappresentabile graficamente tramite il seguente diagramma di flusso:



Dal momento che il modello non prevede variazioni demografiche nella popolazione totale, si può ricavare un ulteriore vincolo:

$$\frac{dS}{dt} + \frac{dI}{dt} + \frac{dR}{dt} = 0 \quad (4.5)$$

Tale condizione permette di ridurre il sistema che descrive il modello, considerando il comportamento delle sole classi di Suscettibili e Infetti. Tramite sostituzione in formula (4.3) della relazione (4.1) risulta:

$$\frac{dS}{dt} = -\alpha S I \quad (4.6)$$

$$\frac{dI}{dt} = (\alpha - \beta) S I + (\beta N - \gamma - \beta I) I \quad (4.7)$$

Segue una breve analisi del modello. Data una qualsiasi scelta di condizioni iniziali

$$S(0) > 0, I(0) = N - S(0) > 0, R(0) = 0$$

è possibile innanzitutto notare, partendo dalla relazione (2.3), che:

$$\left. \frac{dI}{dt} \right|_{t=0} = I(0)(\alpha S(0) - \gamma) \geq 0 \iff S(0) \geq \frac{\gamma}{\alpha} \quad (4.8)$$

Ne consegue che se  $S(0) < \frac{\gamma}{\alpha}$  il meme si estingue ancor prima di iniziare a diffondersi.

Si prenda ora in considerazione la condizione di equilibrio per il sistema:

$$\frac{dS}{dt} = -\alpha S I = 0$$

$$\frac{dI}{dt} = (\alpha - \beta) S I + (\beta N - \gamma - \beta I) I = 0$$

Un primo evidente punto di equilibrio è rappresentato dalla condizione di estinzione del meme  $(S^*, 0)$  con  $S^* \leq S(0)$  dipendente dalle condizioni iniziali del sistema. Tuttavia se  $N \frac{\beta}{\gamma} \geq 1$  si ottiene un nuovo equilibrio  $(0, I^*)$  con  $I^* = N - \frac{\gamma}{\beta}$ . La stabilità di tale punto di equilibrio può essere studiata tramite il teorema di linearizzazione [26]. Il sistema viene quindi linearizzato in  $(0, I^*)$  ottenendo la seguente matrice jacobiana:

$$J|_{(0, I^*)} = \begin{bmatrix} -\alpha(N - \frac{\gamma}{\beta}) & 0 \\ (\alpha - \beta)(N - \frac{\gamma}{\beta}) & -\beta(N - \frac{\gamma}{\beta}) \end{bmatrix}$$

Grazie alla struttura triangolare inferiore della matrice risulta evidente che entrambi i suoi autovalori sono negativi, di conseguenza il punto di equilibrio  $(0, I^*)$  è asintoticamente stabile.



I risultati sin qui ottenuti, analoghi a quelli riportati da Wang e Wood [21], vengono ora riassunti e validati dai risultati di alcune simulazioni ottenute tramite MATLAB Simulink. Data una qualunque scelta di condizioni iniziali compatibili con il modello, se  $S(0) > \frac{\gamma}{\alpha}$  allora il numero di Infetti inizialmente aumenta e la sopravvivenza o l'estinzione del meme sono determinate dal valore di  $R_0 = \frac{\beta N}{\gamma}$ . Se  $R_0 \leq 1$  il meme tende all'estinzione, ovvero risulta

$$(S, I) \rightarrow (S^*, 0) \quad \text{per} \quad t \rightarrow \infty$$

con  $S^* \leq S(0)$  dipendente dalle condizioni iniziali. Un esempio di tale comportamento è mostrato in Figura 9.

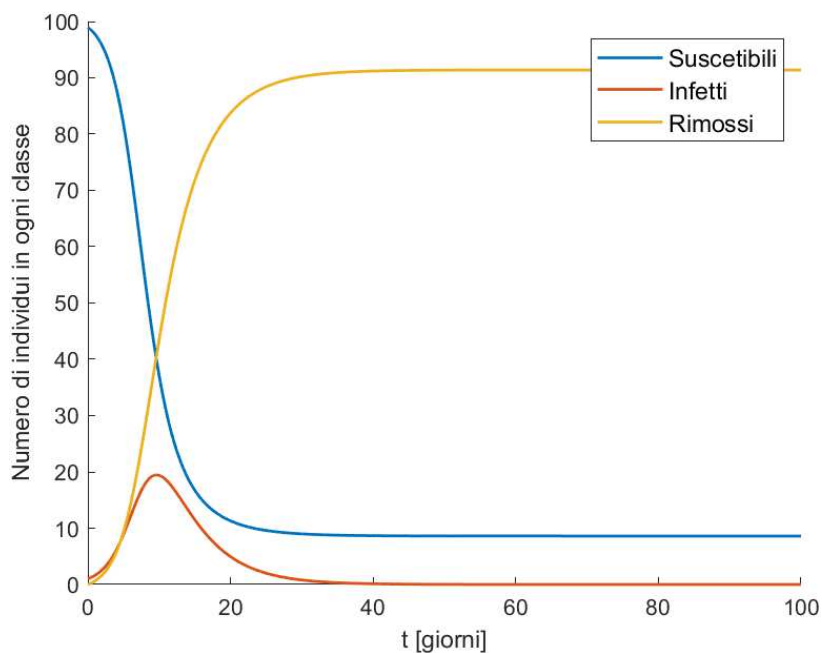


Figura 9: simulazione del modello con  $\alpha = 1, \beta = 0.25, \gamma = 0.5, N = 100, S(0) = 99$

Viceversa, se  $R_0 > 1$  il tasso con il quale gli Infetti perdono interesse per il meme finisce per compensare il tasso di infezione sui Rimossi, per cui si giunge ad una nuova condizione di equilibrio:

$$(S, I) \rightarrow (0, N - \frac{\gamma}{\beta}) \quad \text{per} \quad t \rightarrow \infty$$

Tale condizione è mostrata in Figura 10.

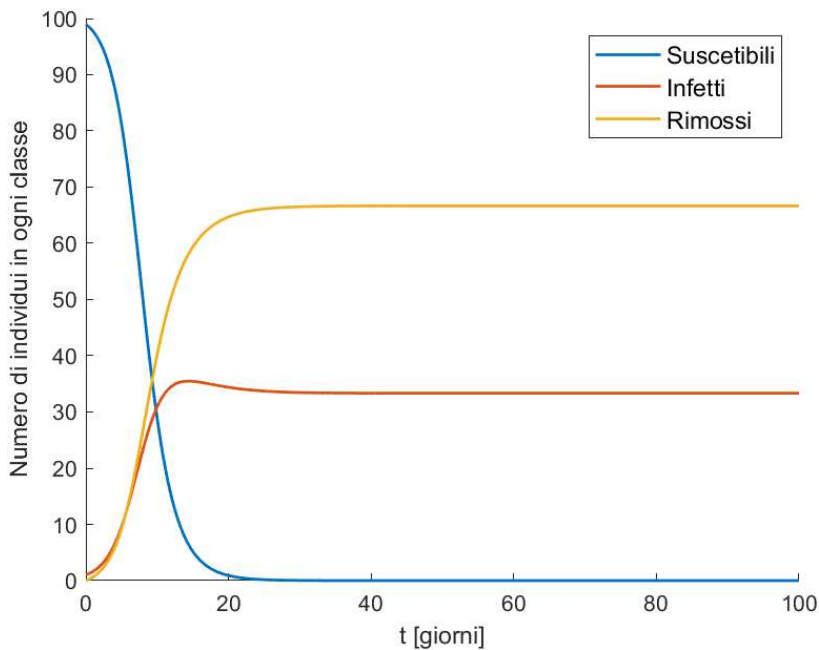


Figura 10: simulazione del modello con  $\alpha = 1$ ,  $\beta = 0.75$ ,  $\gamma = 0.5$ ,  $N = 100$ ,  $S(0) = 99$

Esattamente come fatto da Perissi et al., Wang e Wood hanno mostrato nel loro lavoro un'incoraggiante coincidenza tra i dati di ricerca su Google di alcuni meme e l'andamento del modello [21].

## 4.2 Demografia e rimozione alternativa degli Infetti

Nonostante lo studio della diffusione di Internet meme sia di interesse relativamente recente, un valido contributo può essere ricercato anche in fonti di natura meno attuale. Risulta infatti evidente come i meccanismi che governano la diffusione di contenuti virali in rete possano presentare molti tratti comuni con il comportamento di altri sistemi di carattere sociologico. Nello specifico, Al-Amoudi et al. [22, 23, 24] hanno adottato per lo studio della diffusione di meme alcuni modelli precedentemente introdotti da Kawachi [25], inizialmente nati per l'analisi della propagazione di dicerie. D'altro canto, una diceria altro non è che uno specifico caso di meme, inteso nella più classica definizione data da Dawkins [1]. Dal momento che le ipotesi dei suddetti modelli ben si prestano anche all'analisi della diffusione degli Internet meme, tali modelli verranno presi in considerazione e proposti in questo elaborato. In particolare il primo tra questi verrà introdotto allo scopo di presentare i due successivi, derivati da modifiche ed integrazioni alternative operate sul primo modello.

Per cominciare, viene ora rilassata una delle ipotesi preliminari del modello SIR, introducendo una meccanica demografica all'interno del sistema. La popolazione totale di conseguenza varia sotto l'effetto di fenomeni migratori. Il meccanismo di rimozione dalla classe degli Infetti viene inoltre modificato: si assume ora che gli individui interessati alla condivisione del meme possano perdere interesse per contatto con altri Infetti o con individui Rimossi. Infine viene ipotizzato che l'incontro

tra individui Suscettibili ed individui Infetti possa generare sia nuovi Infetti che nuovi Rimossi. Il nuovo modello è quindi descritto dal seguente sistema di equazioni:

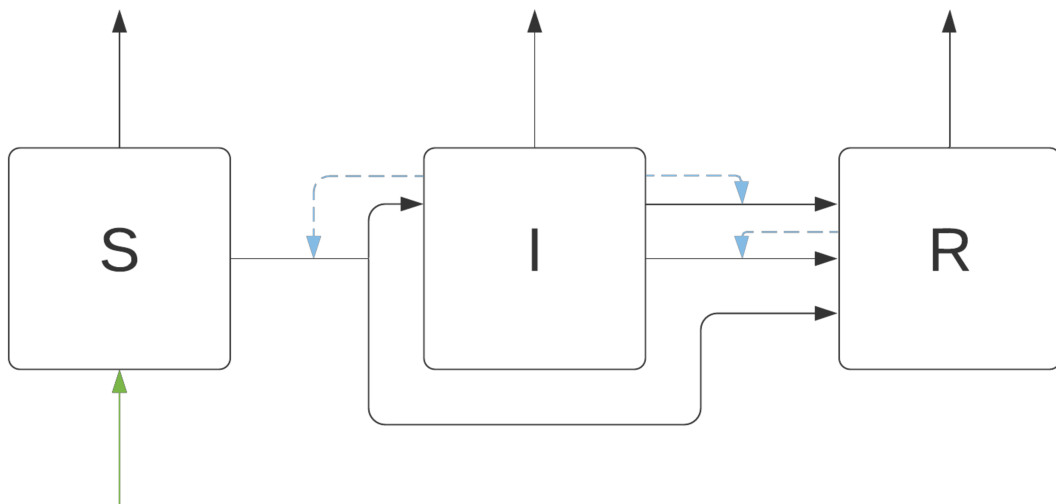
$$\frac{dS}{dt} = B - \alpha S I - \mu S \quad (4.9)$$

$$\frac{dI}{dt} = \alpha \theta S I - \beta I^2 - \gamma I R - \mu I \quad (4.10)$$

$$\frac{dR}{dt} = \alpha (1 - \theta) S I + \beta I^2 + \gamma I R - \mu R \quad (4.11)$$

dove  $B$  è il ritmo di ingresso nel sistema e  $\mu$  quello di uscita, mentre  $\theta \in (0, 1]$  è la frazione di Suscettibili influenzati dagli Infetti ad interessarsi al meme,  $\beta I$  è il tasso di rimozione degli Infetti dovuto al contatto con altri Infetti,  $\gamma R$  è il tasso di rimozione degli Infetti dovuto al contatto con individui Rimossi.

Il seguente diagramma di flusso descrive quindi la struttura del modello:



Risulta evidente, osservando il diagramma di flusso, come questo modello sia significativamente più complesso rispetto a quelli analizzati finora, sebbene ne condivida la filosofia di fondo e buona parte delle meccaniche.

Date le dinamiche demografiche sopra descritte, è possibile ricavare un'ulteriore relazione notevole per il modello:

$$\frac{dN}{dt} = B - \mu N \quad (4.12)$$

con  $N$  numero di individui all'interno del sistema.

Risolvendo l'equazione (4.12) risulta quindi:

$$N = N(0) e^{-\mu t} + \frac{B}{\mu} (1 - e^{-\mu t}) \quad (4.13)$$

Di conseguenza:

$$N \rightarrow \frac{B}{\mu} \quad \text{per} \quad t \rightarrow \infty \quad (4.14)$$

La regione di lavoro del modello è quindi la seguente:

$$\{(S, I, R) \mid S + I + R \leq \frac{B}{\mu}, S > 0, I \geq 0, R \geq 0\} \quad (4.15)$$

Viene ora applicato il metodo della *next generation matrix* [27] per calcolare il numero riproduttivo di base  $R_0$  [22]. Tale operazione verrà eseguita solamente per questo modello tra quelli proposti da Kawachi, dal momento che con passaggi del tutto analoghi sarebbe possibile mostrare che il numero riproduttivo di base così ottenuto non varia fra le tre formulazioni proposte. Si consideri quindi la condizione di equilibrio per il modello:

$$\frac{dS}{dt} = B - \alpha S I - \mu S = 0$$

$$\frac{dI}{dt} = \alpha \theta S I - \beta I^2 - \gamma I R - \mu I = 0$$

$$\frac{dR}{dt} = \alpha (1 - \theta) S I + \beta I^2 + \gamma I R - \mu R = 0$$

Dati  $I = 0$  e  $R = 0$ , il punto di equilibrio *meme free* per il modello è  $(\frac{B}{\mu}, 0, 0)$ . Si consideri ora il modello riscritto come  $(I, R, S)' = \mathcal{F}(I, R, S) - \mathcal{V}(I, R, S)$  dove

$$\mathcal{F}(I, R, S) = \begin{bmatrix} \alpha \theta S I \\ \alpha (1 - \theta) S I \\ 0 \end{bmatrix}, \mathcal{V}(I, R, S) = \begin{bmatrix} \beta I^2 + \gamma I R + \mu I \\ -\beta I^2 - \gamma I R + \mu R \\ -B + \alpha S I + \mu S \end{bmatrix}$$

Vengono quindi calcolate le matrici jacobiane di  $\mathcal{F}(I, R, S)$  e  $\mathcal{V}(I, R, S)$  valutate nel punto di equilibrio  $(\frac{B}{\mu}, 0, 0)$ :

$$J_{\mathcal{F}}|_{(\frac{B}{\mu}, 0, 0)} = \begin{bmatrix} F & 0 \\ 0 & 0 \end{bmatrix} \quad \text{con} \quad F = \begin{bmatrix} \alpha \theta \frac{B}{\mu} & 0 \\ \alpha (1 - \theta) \frac{B}{\mu} & 0 \end{bmatrix}$$

$$J_{\mathcal{V}}|_{(\frac{B}{\mu}, 0, 0)} = \begin{bmatrix} V & 0 \\ J_1 & J_2 \end{bmatrix} \quad \text{con} \quad V = \begin{bmatrix} \mu & 0 \\ 0 & \mu \end{bmatrix}, J_1 = \begin{bmatrix} \alpha \frac{B}{\mu} & 0 \end{bmatrix}, J_2 = [\mu]$$

La next generation matrix è quindi la seguente:

$$F V^{-1} = \begin{bmatrix} \frac{\alpha \theta B}{\mu^2} & 0 \\ \frac{\alpha (1 - \theta) B}{\mu^2} & 0 \end{bmatrix}$$

Il numero riproduttivo di base del sistema può allora essere ricavato come il *raggio spettrale* di tale matrice, ovvero il massimo dei valori assoluti dei suoi autovalori:

$$R_0 = \rho(F V^{-1}) = \frac{\alpha \theta B}{\mu^2} \quad (4.16)$$

Come anticipato, a completare il modello appena descritto verranno introdotte nei prossimi paragrafi due ulteriori ipotesi.

### 4.3 Perdita di memoria dei Rimossi

Con il passare del tempo è ragionevole pensare che chi è venuto a contatto con il meme possa dimenticarsene, così come può capitare che una leggera mutazione del meme – o della sua rappresentazione - faccia sì che questo venga recepito da alcuni individui come un meme completamente nuovo. Per questa ragione un'interessante modifica al modello precedente, introdotta dallo stesso Kawachi [25] e in seguito analizzata da Al-Amoudi et al. [23], consiste nell'introduzione di un nuovo flusso che dalla classe dei Rimossi raccoglie gli individui che perdono memoria del meme e li reinserisce nella classe dei Suscettibili. Il nuovo modello così generato è descritto dalle seguenti equazioni differenziali:

$$\frac{dS}{dt} = B - \alpha S I - \mu S + \eta R \quad (4.17)$$

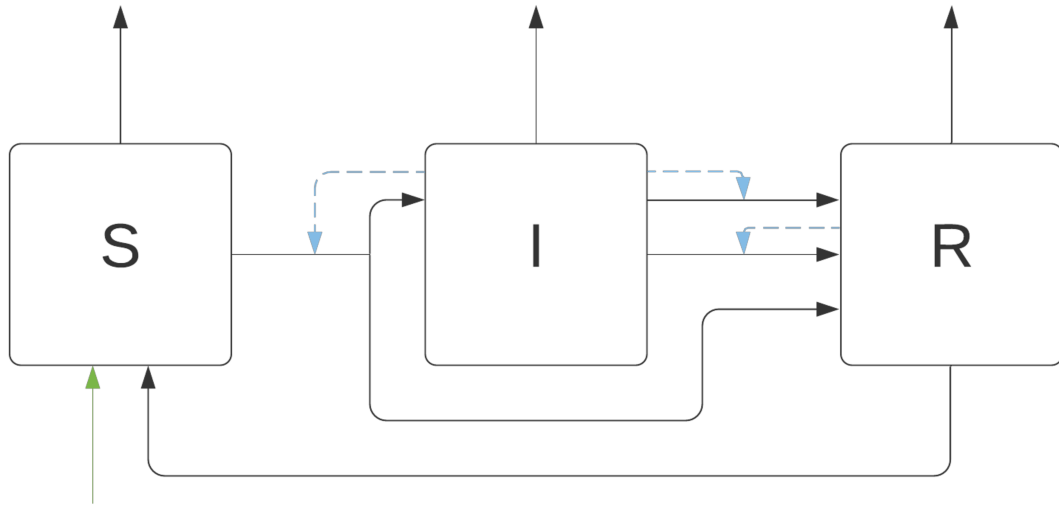
$$\frac{dI}{dt} = \alpha \theta S I - \beta I^2 - \gamma I R - \mu I \quad (4.18)$$

$$\frac{dR}{dt} = \alpha (1 - \theta) S I + \beta I^2 + \gamma I R - \mu R - \eta R \quad (4.19)$$

Il nuovo parametro  $\eta$  rappresenta quindi la frequenza *per capita* con la quale i Rimossi rientrano nella classe dei Suscettibili, perdendo memoria del meme.

Analogamente al modello precedente, una meccanica demografica tiene conto dei fenomeni migratori che influenzano la popolazione del sistema. La regione di lavoro del sistema risulta quindi nuovamente quella introdotta nella formula (4.15).

Il diagramma di flusso del nuovo modello, ottenuto integrando il diagramma del modello precedente con il nuovo flusso che dalla classe dei Rimossi trasferisce gli individui che perdono memoria del meme nella classe dei Suscettibili, si presenta quindi come segue:



Come anticipato, il numero riproduttivo di base risulta essere analogo a quello ricavato per il modello precedente e presentato in formula (2.16), qui riportato per completezza:

$$R_0 = \frac{\alpha \theta B}{\mu^2}$$

Si consideri ora la condizione di equilibrio per il modello:

$$\frac{dS}{dt} = B - \alpha S I - \mu S + \eta R = 0$$

$$\frac{dI}{dt} = \alpha \theta S I - \beta I^2 - \gamma I R - \mu I = 0$$

$$\frac{dR}{dt} = \alpha (1 - \theta) S I + \beta I^2 + \gamma I R - \mu R - \eta R = 0$$

Risolvendo il sistema è possibile ricavare due punti di equilibrio [23]. Il primo, più evidente e sempre coerente con la zona di lavoro del modello per qualsiasi scelta di parametri, identifica lo scenario di equilibrio *meme free*  $(\frac{B}{\mu}, 0, 0)$ . Il secondo punto di equilibrio  $(S^*, I^*, R^*)$  con

$$S^* = \frac{B + \eta R^*}{\alpha I^* + \mu} = \frac{1}{\alpha I^*} [(\mu + \eta) R^* + \mu I^*]$$

$$R^* = \frac{-\mu I^*}{(\alpha I^* + \mu) + \eta} \left( \frac{\alpha I^*}{\mu} + 1 - \frac{\alpha B}{\mu^2} \right)$$

rappresenta lo scenario di persistenza del meme. Tale punto di equilibrio tuttavia risulta incoerente con le ipotesi di lavoro del modello nel caso in cui  $\frac{\alpha B}{\mu^2} \leq 1$  con  $I^* > 0$ , dal momento che si otterrebbe un numero negativo di Rimossi  $R^* < 0$ .

Viene ora investigata la stabilità dei punti di equilibrio precedentemente individuati [23]. Partendo dall'equilibrio  $(\frac{B}{\mu}, 0, 0)$ , si consideri la matrice jacobiana valutata in tale punto:

$$J|_{\left(\frac{B}{\mu}, 0, 0\right)} = \begin{bmatrix} -\mu & -\frac{\alpha B}{\mu} & \eta \\ 0 & \frac{\alpha \theta B}{\mu} - \mu & 0 \\ 0 & \frac{\alpha (1-\theta) B}{\mu} - \mu & -\mu - \eta \end{bmatrix}$$

Risulta semplice ricavare gli autovalori:

$$-\mu < 0 \quad , \quad -\mu - \eta < 0 \quad , \quad \frac{\alpha \theta B}{\mu} - \mu = \mu (R_0 - 1)$$

Mentre i primi due autovalori trovati sono sempre negativi, l'ultimo autovalore risulta negativo se e solo se  $R_0 < 1$ . Si può quindi concludere per teorema di linearizzazione [26] che il punto di equilibrio *meme free*  $\left(\frac{B}{\mu}, 0, 0\right)$  risulta asintoticamente stabile se  $R_0 < 1$  mentre risulta instabile se  $R_0 > 1$ .

Considerando invece il punto di equilibrio positivo, vale a dire quello di persistenza del meme  $(S^*, I^*, R^*)$ , la matrice jacobiana calcolata in tale punto risulta:

$$J|_{(S^*, I^*, R^*)} = \begin{bmatrix} -\frac{(B+\eta R^*)}{S^*} & -\alpha S^* & \eta \\ \alpha \theta I^* & -\beta I^* & -\gamma I^* \\ \alpha (1-\theta) I^* & \beta I^* + \frac{(\mu+\eta) R^*}{I^*} & \gamma I^* - \mu - \eta \end{bmatrix}$$

Il polinomio caratteristico per tale matrice è  $\lambda^3 + a_1 \lambda^2 + a_2 \lambda + a_3 = 0$  dove

- $a_1 = \mu + \eta + \frac{B+\eta R^*}{S^*} + (\beta - \gamma) I^* > 0$  se  $\beta > \gamma$
- $a_2 = \frac{(B+\eta R^*) I^*}{S^*} (\beta - \gamma) + (\mu + \eta) \beta I^* + \gamma (\mu + \eta) R^* + \frac{\mu (B+\eta R^*)}{S^*} + \alpha^2 \theta S^* I^* + \eta \left( \frac{(B+\eta R^*)}{S^*} - \alpha I^* \right) + \eta \alpha \theta I^* > 0$  se  $\beta > \gamma$
- $a_3 = \eta \mu \beta I^* + \alpha S^* I^* (\eta \alpha \theta - \alpha \gamma I^* + \mu \alpha \theta) + \frac{(B+\eta R^*) \beta \mu I^*}{S^*} + \gamma \alpha I^* (\mu + \eta) R^* + (\gamma \mu - \eta \alpha \theta) (\mu + \eta) R^* > 0$  se  $\beta > \gamma \wedge \gamma \mu > \eta \alpha \theta$

Sfruttando il criterio di Routh-Hurwitz [28] è possibile verificare sotto quali condizioni gli autovalori della matrice  $J|_{(S^*, I^*, R^*)}$  hanno tutti parte reale negativa. Omettendo i passaggi intermedi, si può concludere che il punto di equilibrio endemico  $(S^*, I^*, R^*)$  è asintoticamente stabile se  $\beta > \gamma$  e  $\gamma \mu > \eta \alpha \theta$ .

I risultati ottenuti dall'analisi del modello vengono ora riassunti e validati da alcune simulazioni ottenute con MATLAB Simulink. In generale sono stati ricavati due punti di equilibrio: un primo punto di equilibrio *meme free*  $\left(\frac{B}{\mu}, 0, 0\right)$  e un punto  $(S^*, I^*, R^*)$  di equilibrio endemico. Il punto di equilibrio  $\left(\frac{B}{\mu}, 0, 0\right)$  è asintoticamente stabile se e solo se  $R_0 < 1$ . Se poi vale  $\frac{\alpha B}{\mu^2} \leq 1$  il punto di

equilibrio endemico non esiste e il meme finisce per estinguersi, come mostrato in Figura 11.

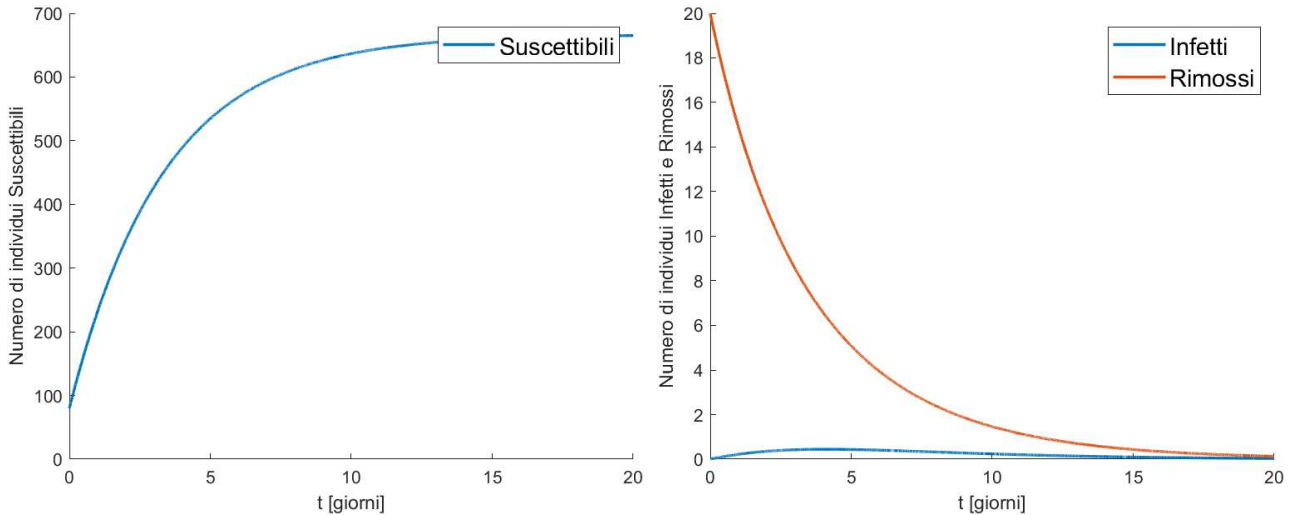


Figura 11: simulazione del modello con  $\alpha = 0.0125$ ,  $\beta = 0.05$ ,  $\gamma = 0.1$ ,  $\eta = 0.02$ ,  $\theta = 0.7$ ,  $\mu = 0.3$ ,  $B = 2$ ,  $N = 100$ ,  $S(0) = 80$

Viceversa per  $R_0 > 0$  il punto di equilibrio non endemico risulta instabile. Se poi valgono  $\beta > \gamma$  e  $\gamma \mu > \eta \alpha \theta$  allora il punto di equilibrio endemico  $(S^*, I^*, R^*)$  è asintoticamente stabile, per cui il meme persiste, come accade nell'esempio mostrato nella simulazione riportata in Figura 12.

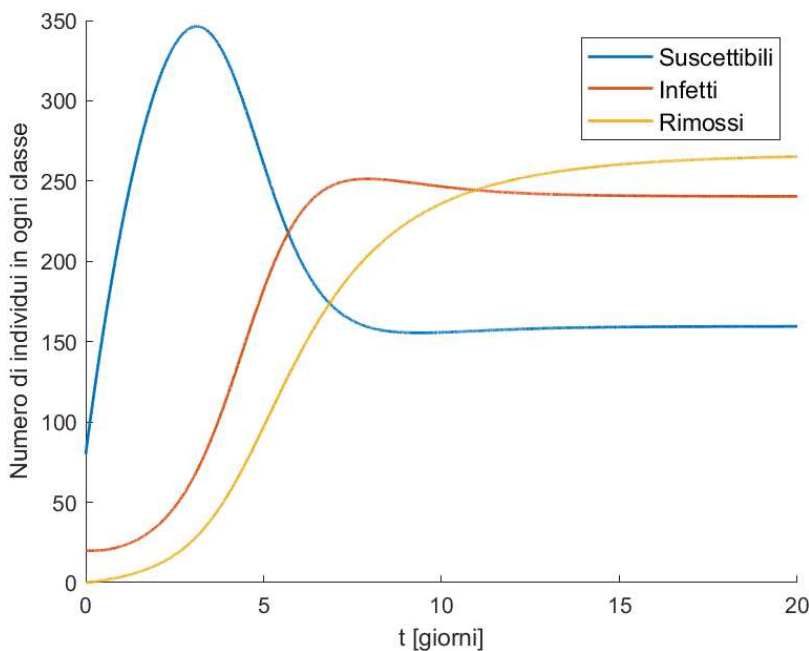


Figura 12: simulazione del modello con  $\alpha = 0.4$ ,  $\beta = 0.05$ ,  $\gamma = 0.01$ ,  $\eta = 0.005$ ,  $\theta = 0.7$ ,  $\mu = 0.3$ ,  $B = 2$ ,  $N = 100$ ,  $S(0) = 80$

#### 4.4 Persuasione alla non condivisione

Viene ora introdotta ed esplorata una nuova ipotesi applicata al modello di Kawachi [25] precedentemente esposto nel paragrafo 4.2, ancora una volta ideata dallo stesso Kawachi e in seguito analizzata da Al-Amoudi et al. [24]: è ragionevole pensare che - esattamente come gli Infetti sono in grado di persuadere i Suscettibili alla condivisione del meme - anche i Rimossi, qualora fossero



detrattori del concetto da esso veicolato, possano convincere i Suscettibili a non condividere il meme, portandoli quindi ad entrare a loro volta nella classe dei Rimossi. Le equazioni che governano il modello, invariate in ogni loro altra meccanica, sono quindi le seguenti:

$$\frac{dS}{dt} = B - \alpha S I - \mu S - \delta S R \quad (4.20)$$

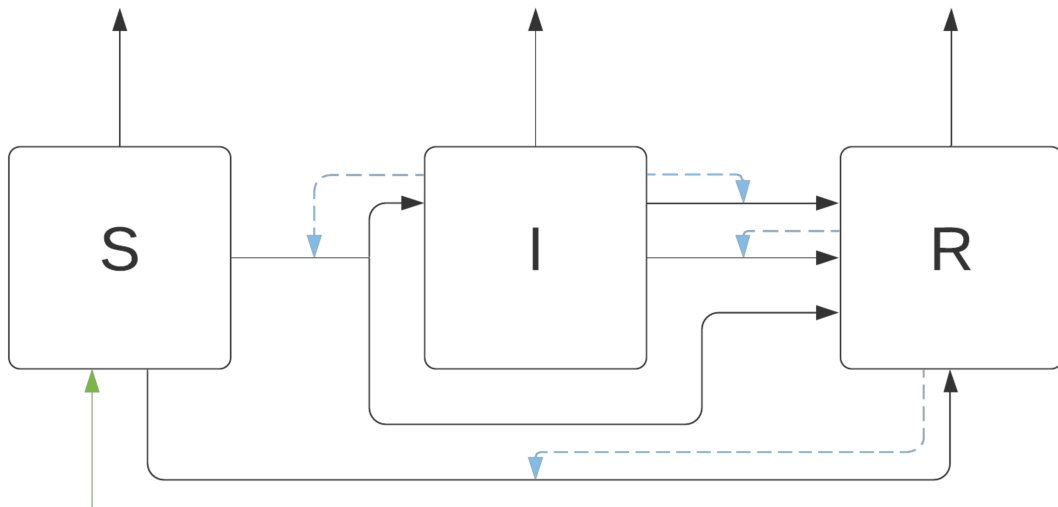
$$\frac{dI}{dt} = \alpha \theta S I - \beta I^2 - \gamma I R - \mu I \quad (4.21)$$

$$\frac{dR}{dt} = \alpha (1 - \theta) S I + \beta I^2 + \gamma I R - \mu R + \delta S R \quad (4.22)$$

dove il termine  $\delta R$  identifica quindi il tasso *per capita* con il quale gli individui Suscettibili vengono persuasi dagli individui Rimossi ad entrare a far parte di questa seconda classe.

Ancora una volta una meccanica demografica descrive i fenomeni migratori ai quali il sistema è sottoposto. La regione di lavoro è nuovamente espressa dalla formula (4.15).

Come di consueto, viene di seguito mostrato il diagramma di flusso che descrive graficamente la struttura del nuovo modello:



Come anticipato, il numero riproduttivo di base è analogo a quello dei due modelli precedenti, ricavato in formula (4.16):

$$R_0 = \frac{\alpha \theta B}{\mu^2}$$

Segue, come sempre, un'analisi dei punti di equilibrio del sistema [24], utile a predirne il comportamento al variare dei parametri. La condizione di equilibrio del sistema è la seguente:

$$\frac{dS}{dt} = B - \alpha S I - \mu S - \delta S R = 0$$

$$\frac{dI}{dt} = \alpha \theta S I - \beta I^2 - \gamma I R - \mu I = 0$$

$$\frac{dR}{dt} = \alpha (1 - \theta) S I + \beta I^2 + \gamma I R - \mu R + \delta S R = 0$$

Risolvendo il sistema si ottengono tre possibili soluzioni: un punto di equilibrio *meme free*  $\left(\frac{B}{\mu}, 0, 0\right)$ , un punto di equilibrio di terminazione del meme  $\left(\frac{\mu}{\delta}, 0, \frac{B}{\mu} - \frac{\mu}{\delta}\right)$  e un punto di equilibrio endemico  $(S^*, I^*, R^*)$  con

$$S^* = \frac{B}{\alpha I^* + \delta R^* + \mu}$$

$$I^* = \frac{\alpha \theta S^* - \delta R^* - \mu}{\beta}$$

$$R^* = \frac{-\alpha (1 - \theta) S^* I^* - \beta I^{*2}}{\delta S^* + \gamma I^* - \mu}$$

Si consideri ora la matrice jacobiana del sistema, valutata nel punto  $\left(\frac{B}{\mu}, 0, 0\right)$ :

$$J|_{\left(\frac{B}{\mu}, 0, 0\right)} = \begin{bmatrix} -\mu & -\frac{\alpha B}{\mu} & -\frac{\delta B}{\mu} \\ 0 & \frac{\alpha \theta B}{\mu} - \mu & 0 \\ 0 & \frac{\alpha (1 - \theta) B}{\mu} & \frac{\delta B}{\mu} - \mu \end{bmatrix}$$

Gli autovalori di tale matrice sono:

$$-\mu < 0$$

$$\mu \left(\frac{\delta B}{\mu^2} - 1\right) < 0 \quad \text{se} \quad \delta < \frac{\mu^2}{B}$$

$$\mu \left(\frac{\alpha \theta B}{\mu^2} - 1\right) = \mu (R_0 - 1) < 0 \quad \text{se} \quad R_0 < 1$$

Ne consegue, per teorema di linearizzazione [26], che il punto di equilibrio  $\left(\frac{B}{\mu}, 0, 0\right)$  è asintoticamente stabile se  $R_0 < 1 \wedge \delta < \frac{\mu^2}{B}$ , mentre è instabile se  $R_0 > 1 \vee \delta > \frac{\mu^2}{B}$ .

Analogamente, si consideri la matrice jacobiana valutata in  $(\frac{\mu}{\delta}, 0, \frac{B}{\mu} - \frac{\mu}{\delta})$ :

$$J|_{(\frac{\mu}{\delta}, 0, \frac{B}{\mu} - \frac{\mu}{\delta})} = \begin{bmatrix} -\frac{\delta B}{\mu} & -\frac{\alpha \mu}{\delta} & -\mu \\ 0 & \frac{\alpha \theta \mu}{\delta} - \gamma \left(\frac{B}{\mu} - \frac{\mu}{\delta}\right) - \mu & 0 \\ \frac{\delta B}{\mu} - \mu & \frac{\alpha (1-\theta)\mu}{\delta} - \gamma \left(\frac{B}{\mu} - \frac{\mu}{\delta}\right) & 0 \end{bmatrix}$$

Dato

$$R_1 = \frac{\delta (B \gamma + \mu^2)}{\mu^2 (\gamma + \alpha \theta)} \quad (4.23)$$

è possibile osservare che gli autovalori di tale matrice sono:

$$-\mu < 0$$

$$-\frac{1}{\mu} (B \delta - \mu^2) < 0$$

$$\frac{1}{\mu \delta} [\mu^2 (\gamma + \alpha \theta) - \delta (B \gamma + \mu^2)] < 0 \quad \text{se} \quad R_1 > 1$$

Di conseguenza il punto di equilibrio  $(\frac{\mu}{\delta}, 0, \frac{B}{\mu} - \frac{\mu}{\delta})$  di terminazione del meme risulta asintoticamente stabile per  $R_1 > 1$ . Nel caso in cui invece dovesse valere  $R_1 < 1$  tale punto di equilibrio risulterebbe instabile.

Si consideri infine la matrice jacobiana del sistema valutata nel punto di equilibrio endemico  $(S^*, I^*, R^*)$ , dopo alcune semplificazioni:

$$J|_{(S^*, I^*, R^*)} = \begin{bmatrix} -\frac{B}{S^*} & -\alpha S^* & -\delta S^* \\ \alpha \theta I^* & -\beta I^* & -\gamma I^* \\ \alpha (1-\theta) I^* + \delta R^* & \beta I^* + \frac{\mu R^*}{I^*} - \frac{\delta S^* R^*}{I^*} & \delta S^* + \gamma I^* - \mu \end{bmatrix}$$

Il polinomio caratteristico di tale matrice risulta  $\lambda^3 + a_1 \lambda^2 + a_2 \lambda + a_3 = 0$  dove

$$a_1 = \beta I^* + \mu + \frac{B}{S^*} - \delta S^*$$

$$a_2 = \alpha \delta (1-\theta) S^* I^* + \delta^2 S^* R^* + \gamma \mu R^* - \gamma \delta S^* R^* - \delta B - \delta \beta S^* I^* - \frac{\gamma B I^*}{S^*} + \frac{B \mu}{S^*} + \mu \beta I^* + \frac{\beta B I^*}{S^*} + \alpha^2 \theta S^* I^*$$

$$a_3 = \alpha \theta \mu S^* R^* - \delta^2 \alpha \theta S^{*2} R^* + \alpha \beta \delta S^* I^{*2} + \beta \delta^2 S^* I^* R^* + \frac{\gamma B \mu R^*}{S^*} - B \gamma \delta R^* - \alpha^2 \gamma S^* I^* - \alpha \gamma \delta S^* I^* R^* - \delta B \beta I^* - \alpha^2 \theta \delta S^{*2} I^* + \frac{\beta B \mu I^*}{S^*} + \alpha^2 \theta \mu S^* I^*$$

Secondo il criterio di Routh-Hurwitz [28] se  $a_1 > 0 \wedge a_3 > 0 \wedge a_1 a_2 > a_3$  allora la matrice  $J|_{(S^*, I^*, R^*)}$  ha solo autovalori a parte reale negativa, pertanto in tal caso il punto di equilibrio endemico  $(S^*, I^*, R^*)$  risulta asintoticamente stabile.

Vengono ora riassunti i risultati riportati sopra. Il punto di equilibrio *meme free*  $(\frac{B}{\mu}, 0, 0)$  risulta asintoticamente stabile se  $R_0 < 1 \wedge \delta < \frac{\mu^2}{B}$ . In caso di convergenza a tale punto di equilibrio la diffusione del meme fallisce, come è possibile osservare nell'esempio in Figura 13.

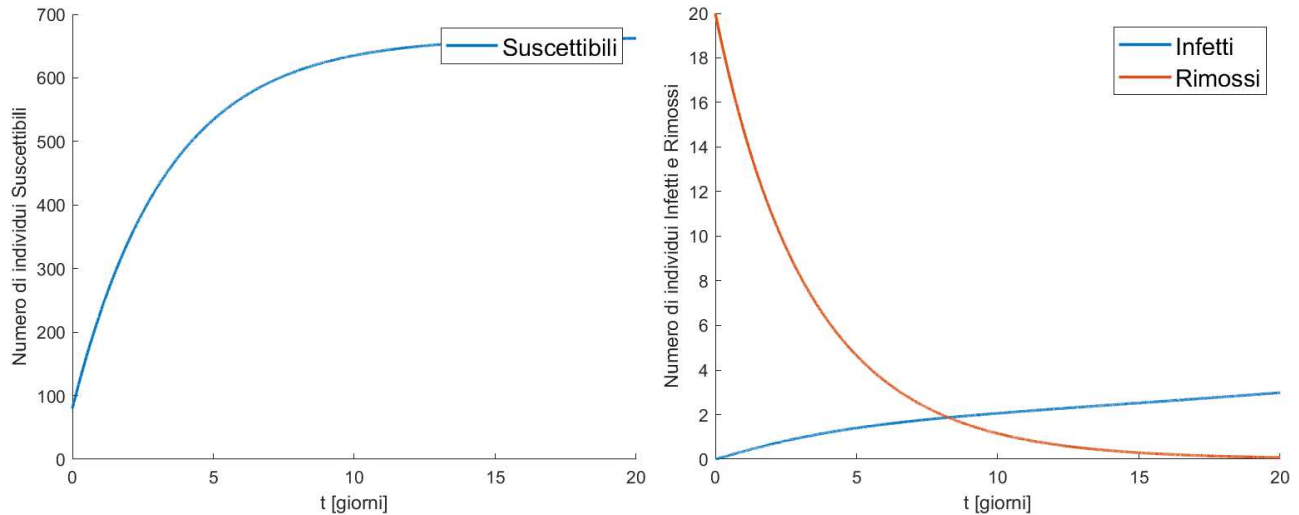


Figura 13: simulazione del modello con  $\alpha = 0.0125, \beta = 0.05, \gamma = 0.015, \delta = 0.05, \theta = 0.3, B = 2, \mu = 0.3$

Il punto di equilibrio di terminazione del meme  $(\frac{\mu}{\delta}, 0, \frac{B}{\mu} - \frac{\mu}{\delta})$  risulta invece asintoticamente stabile se  $R_1$  definito in formula (4.23) è maggiore di 1. In questo caso il meme si diffonde per un primo periodo, tuttavia il numero di Infetti finisce per azzerarsi, lasciando una parte di popolazione nella classe dei Rimossi. Sebbene in termini epidemiologici un numero così alto di Rimossi identifichi con un patogeno che, seppur estinto, ha avuto modo di entrare in contatto con una grossa porzione della popolazione, è bene ricordare nell'interpretare i risultati ottenuti che in questo modello la classe dei Rimossi contiene coloro che hanno perso interesse nel meme o ne sono diventati addirittura detrattori. Il sistema converge ad uno scenario di questo tipo nell'esempio mostrato in Figura 14.

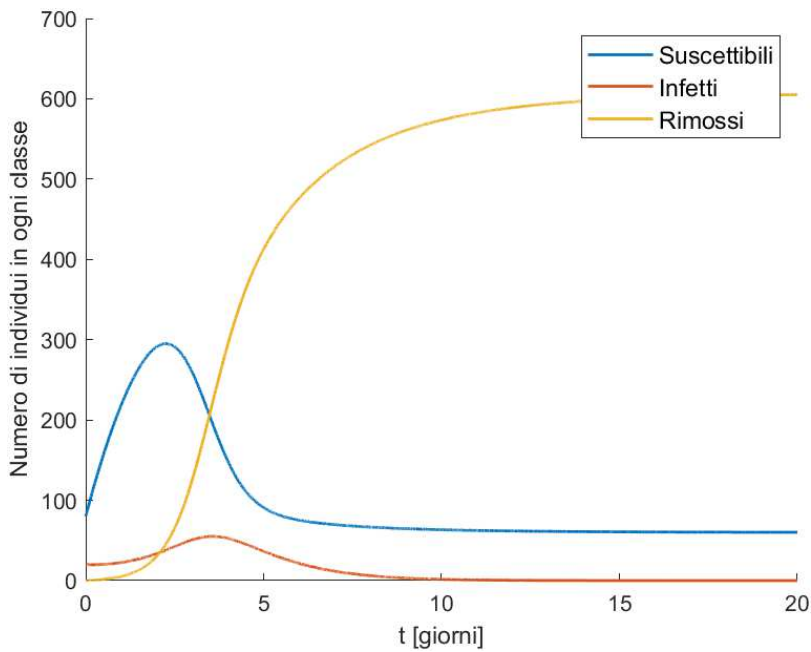


Figura 14: simulazione del modello con  $\alpha = 0.4, \beta = 0.05, \gamma = 0.015, \delta = 0.5, \theta = 0.7, B = 2, \mu = 0.3$

Infine il punto di equilibrio endemico  $(S^*, I^*, R^*)$  risulta asintoticamente stabile nel caso in cui  $a_1 > 0 \wedge a_3 > 0 \wedge a_1 a_2 > a_3$ , secondo il criterio di Routh-Hurwitz [28]. In tal caso il numero di infetti raggiunge un valore costante nel tempo, per cui il meme persiste all'interno della popolazione.

Un esempio di questo tipo è riportato in Figura 15.

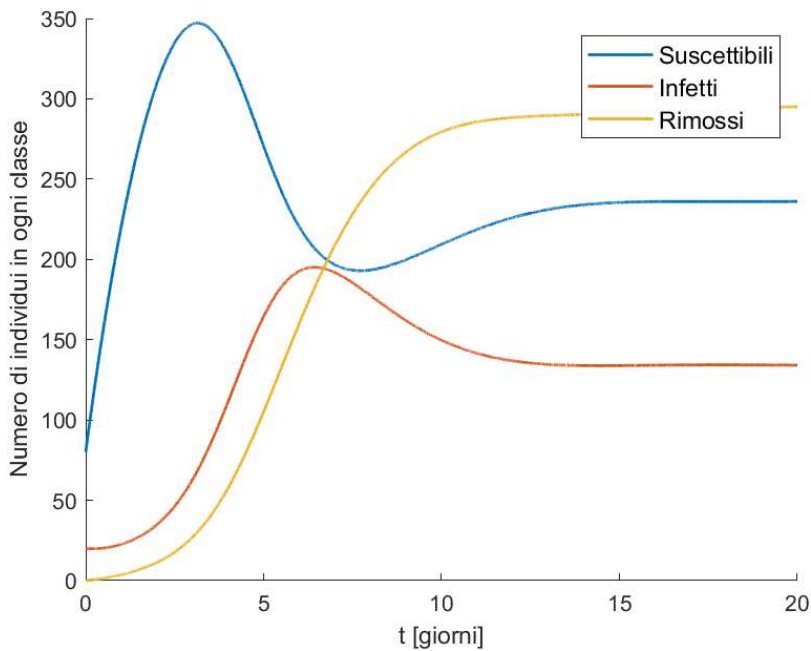


Figura 15: simulazione del modello con  $\alpha = 0.4, \beta = 0.05, \gamma = 0.015, \delta = 0.0035, \theta = 0.7, B = 2, \mu = 0.3$



## CAPITOLO 5: Conclusione

Alla luce dell'analogia riscontrabile tra i meccanismi che governano i fenomeni epidemici di natura virale e quelli che invece guidano la diffusione di un meme, in questa tesi sono stati adottati alcuni modelli epidemiologici allo scopo di descrivere e comprendere i fenomeni di diffusione di meme all'interno della rete Internet.

È stato inizialmente introdotto il modello SIR, ovvero un modello compartimentale non lineare, il quale descrive il fenomeno di diffusione del meme a partire dalla divisione della popolazione in tre classi, vale a dire Suscettibili, Infetti e Rimossi. Contestualmente sono stati presentati alcuni risultati notevoli riguardanti il modello SIR, i quali permettono di comprendere le condizioni necessarie affinché il meme possa diffondersi nella popolazione, nonché di intuire la distribuzione della popolazione nelle varie classi al termine del fenomeno di diffusione del meme. I risultati riportati sono stati validati da alcune simulazioni del modello ottenute mediante l'utilizzo di MATLAB Simulink.

Successivamente tale modello è stato utilizzato per descrivere l'andamento del *search volume index* fornito da Google Trends [29] in relazione alla ricerca di alcune parole chiave legate a specifici meme. È stata osservata un'incoraggiante coincidenza tra i sopracitati dati forniti da Google Trends e l'andamento del numero di Infetti nel modello SIR, osservato in simulazioni Simulink del modello stesso. I parametri utilizzati in tali simulazioni sono stati ricavati tramite un semplice algoritmo basato sul metodo Monte Carlo. Sono state poi approfondite alcune applicazioni di particolare interesse dei temi trattati: è stato infatti proposto di utilizzare modello SIR per descrivere l'andamento di campagne di *viral marketing* e la diffusione di meme di rilievo politico.

Sono stati infine presentati alcuni modelli più complessi, derivati dal modello SIR, allo scopo di descrivere più dettagliatamente i meccanismi di diffusione di meme all'interno della rete Internet. Il primo tra i modelli proposti introduce la possibilità che i soggetti Rimossi possano tornare ad infettarsi, sebbene con probabilità minore dei Suscettibili. Un secondo modello implementa un meccanismo demografico nella popolazione ospite e dettaglia maggiormente il sistema di rimozione degli Infetti. Da tale modello derivano i due restanti: il primo di questi descrive un meccanismo di perdita di memoria degli individui Rimossi, mentre il secondo introduce un'azione di persuasione alla rimozione diretta da parte dei Rimossi nei confronti dei soggetti Suscettibili. Contestualmente sono state riportate le analisi di equilibrio e stabilità locale dei modelli proposti, validate da alcuni esempi ottenuti ancora una volta tramite simulazioni MATLAB Simulink.





## BIBLIOGRAFIA

- [1] Dawkins R., *"The Selfish Gene"*, Oxford University Press, 1976
- [2] Dawkins R., *"The Blind Watchmaker"*, W.W. Norton and Company, 1986
- [3] Dawkins R., *"Viruses of the Mind"*, 1993
- [4] Díaz C.M.C., *"Defining and characterizing the concept of Internet Meme"*, Revista CES Psicología, 2013
- [5] Kermack W.O., McKendrick A.G., *"A contribution to the mathematical theory of epidemics"*, Royal Society Publishing, 1927
- [6] Fuentes M.A., Kuperman M.N., *"Cellular automata and epidemiological models with spatial dependence"*, Elsevier Science, 1999
- [7] Keeling M.J., Rohani P., *"Modeling Infectious Diseases in Humans and Animals"*, Princeton University, 2008
- [8] Peri S.S.S., Chen B., Dougall A.L., Simens G., *"Towards Understanding the Lifespan and Spread of Ideas: Epidemiological Modeling of Participation on Twitter"*, LAK'20 , 2020
- [9] Perissi I., Falsini S., Bardi U., *"Mechanisms of meme propagation in the mediasphere: a system dynamics model"*, Emerald Publishing, 2019
- [10] Jurvetson S., *"What exactly is viral marketing?"*, Red Herring, 2000
- [11] Rodrigues H.S., Fonseca M., *"Viral marketing as an epidemiological model"*, Cornell University, 2005
- [12] Li J., *"Simulation Research on Marketing Effect of Enterprise in Social Network Based on SIR Model"*, Atlantis Press, 2019
- [13] Woo J., Chen H., *"Epidemic model for information diffusion in web forums: experiments in marketing exchange and political dialog"*, Springer Plus, 2016
- [14] Noeighdam S., *"Numerical approximation of modified non-linear SIR model of computer viruses"*, Cornell University, 2019
- [15] Piqueira J.R.C., Araujo V.O., *"A modified epidemiological model for computer viruses"*, Elsevier Science, 2009
- [16] Gao C., Liu J., *"Modeling and Restraining Mobile Virus Propagation"*, IEEE, 2013

- [17] Mishra B.K., Keshri N., “*Mathematical model on the transmission of worms in wireless sensor network*”, Elsevier Science, 2012
- [18] Cisotto G., Badia L., “*Cyber Security of Smart Grids Modeled Through Epidemic Models in Cellular Automata*”, Proc. WoWMoM, 2016
- [19] Yang W., Chang G., Ying W., Zhu Z., “*Research on Dynamic Model of File Diffusion in P2P File-Sharing Networks*”, IEEE, 2008
- [20] Guidolin F., Badia L., Zorzi M., “*A distributed clustering algorithm for coordinated multipoint in LTE networks*”, IEEE Wireless Communications Letters, 2014.
- [21] Wang L., Wood B.C., “*An epidemiological approach to model the viral propagation of memes*”, Elsevier Science, 2011
- [22] Al-Amoudi R., Al-Sheikh S., Al-Tuwairqi S., “*Qualitative Behavior of Solutions to a Mathematical Model of Meme Transmission*”, Science Publishing Corporation, 2014
- [23] Prospero L., Costa R., Badia L., “*Resource sharing in the Internet of Things and selfish behaviors of the agents*”, IEEE Trans. on Circuits and Systems, 2021.
- [24] Badia L., “*On the impact of correlated arrivals and errors on ARQ delay terms*”, IEEE Transactions on Communications 2009
- [25] Kawachi K., “*Deterministic models for rumor transmission*”, Elsevier Science, 2007
- [26] Badia L., “*A Markov analysis of selective repeat ARQ with variable round trip time*”, IEEE Communications Letters, 2013.
- [27] Brauer F., Van der Driessche P., Wu J., “*Mathematical Epidemiology*”, Springer Verlag, 1991
- [28] Routh E.J., “*A Treatise on the Stability of a Given State of Motion: Particularly Steady Motion*”, Macmillan, 1877
- [29] Badia L., Clementel D., Erseghe T., Suitner C., “*Structural and Semantic Impact of Online Collective Action*”, NetSci, 2020.
- [30] Operating System Market Share, <https://netmarketshare.com>, consultato il 17/9/2021
- [31] Madi M., Khalid A., Nizami, S.A., “*Chaos and confusion: The frenzied final hours of the Afghan government*”, BBC News, 8/9/2021, <https://www.bbc.com/news/world-asia-58477131>
- [32] Gindullina E., Badia L., “*An optimization framework for energy topologies in smart cities*,” IEEE WCNC, 2018.