



Università degli Studi di Padova

DIPARTIMENTO DI INGEGNERIA DELL'INFORMAZIONE

Corso di Laurea in Ingegneria dell'Informazione

**Diffusione in rete di contenuti epidemici
per scenari di mobilità'**

Candidato:

Eleonora Canato

Relatore:

Leonardo Badia

Anno Accademico 2013/2014

24 settembre 2014

Indice

1	Introduzione	1
2	Stato dell'arte	5
2.1	Modello SIR deterministico e stocastico	5
2.2	Lavori collegati	9
3	Mobilità e compartimentazione	13
3.1	Catena di Markov	13
3.2	Il modello teorico	16
3.3	Le implementazioni pratiche	17
4	Risultati	21
4.1	Deterministico vs Stocastico	21
4.2	Effetto della compartimentazione	25
4.3	Correlazione del movimento	29
4.4	Calcolo di R_0	34
5	Conclusioni e sviluppi futuri	37
	Bibliografia	39

Elenco delle figure

2.1	schema dell'evoluzione nel modello SIR	6
3.1	esempio di diagramma di transizione	14
3.2	diagramma di transizione a tre stati delle implementazioni pratiche	18
4.1	stato 1: modello deterministico vs stocastico	23
4.2	stato 2: modello deterministico vs stocastico	23
4.3	stato 3: modello deterministico vs stocastico	24
4.4	dinamica totale: modello deterministico vs stocastico	24
4.5	stato 1: modello mobile vs statico con parametri uguali	26
4.6	stato 2: modello mobile vs statico con parametri uguali	26
4.7	stato 3: modello mobile vs statico con parametri uguali	27
4.8	stato 1: modello mobile vs statico con picchi uguali con $\beta = 0.00021$ e $\gamma = 0.041$ nel modello statico	27
4.9	stato 2: modello mobile vs statico con picchi uguali con $\beta = 0.0015$ e $\gamma = 0.1$ nel modello statico	28
4.10	stato 3: modello mobile vs statico con picchi uguali con $\beta = 0.0008$ e $\gamma = 0.1$ nel modello statico	28
4.11	effetto della correlazione: media del massimo numero di infetti	31
4.12	effetto della correlazione: varianza del massimo numero di infetti	31
4.13	effetto della correlazione: casi in cui la malattia non si diffonde	32
4.14	stato 1: massimo numero di infetti al variare di β	33
4.15	stato 2: massimo numero di infetti al variare di β	33
4.16	stato 3: massimo numero di infetti al variare di β	34

Abstract

Questa tesi presenta un adattamento del modello epidemiologico SIR (Suscettibili-Infetti-Rimossi) per creare un nuovo modello matematico del processo di diffusione della malattia in uno scenario di mobilità.

Ciò contribuisce a rendere il modello più realistico e permette di valutare nuovi effetti legati alla circolazione degli individui.

Il processo di mobilità viene rappresentato con una catena di Markov. Il modello risultante è stato studiato tramite il software Matlab, sia analiticamente che tramite simulazioni, mostrando una buona concordanza di entrambi.

Variando i parametri del nuovo modello e considerando un diverso grado di spostamento degli individui si determina che le caratteristiche della rete di mobilità hanno un significativo impatto sul corso della propagazione del contagio.

Capitolo 1

Introduzione

I modelli matematici sono strumenti utili per il progresso scientifico, tecnologico, industriale, economico e sociale per il loro ruolo di predizione e comprensione [1]. Grazie al linguaggio della matematica è possibile modellare sistemi e fenomeni in modo preciso e sotto diversi punti di vista, così da fornire un'immagine dettagliata della realtà. Formulare un modello per un particolare problema è un compromesso tra tre importanti elementi: la precisione, cioè l'abilità di riprodurre i dati osservati e di predire affidabili dinamiche future; la trasparenza che deriva dalla capacità di capire come le varie componenti del modello interagiscono e influenzano la dinamica; la flessibilità che misura la facilità con la quale il modello può essere adattato a nuove situazioni.

Un modello matematico può essere costruito seguendo un approccio deterministico oppure stocastico [2]. Il maggior vantaggio dei modelli deterministici sta nella loro più semplice analisi in quanto non si tiene in considerazione l'incertezza associata alle variabili. Al contrario, i modelli stocastici tengono in considerazione le variazioni dei parametri e quindi forniscono risultati in termini di probabilità. In genere, i modelli stocastici hanno una struttura più complessa di quelli deterministici ma risultano più affidabili in quanto sono capaci di fornire risultati più aderenti alla realtà.

Modelli di entrambi i tipi sono usati, per esempio, nello studio di reti di comunicazione (fissa, mobile, Internet), in applicazioni nelle scienze sociali, in biologia, nello studio della diffusione di malattie infettive [1]-[9]. Una delle primarie ragioni per studiare la diffusione delle malattie infettive è quella di rendere le misure di controllo più efficienti e di rimuovere tempestivamente l'infezione dalla popolazione, ottimizzando l'uso delle risorse limitate [1]. In questo senso i modelli permettono di predire la dinamica dell'epidemia a livello della popolazioni dalla conoscenza di fattori epidemici a livello individuale, il comportamento a lungo termine dalle prime dinamiche dell'invasione, o l'impatto della vaccinazione nella propagazione della malattia.

Lo scopo della tesi è duplice in quanto i modelli epidemiologici possono anche essere applicati:

- per contrastare la propagazione di virus informatici o di *malware* nelle reti di cellulari mobili [10];
- per la valutazione di strategie di disseminazione dell'informazione in *mobile ad hoc network* [3];
- per lo studio della diffusione di contenuti multimediali in piattaforme sociali virtuali [8];
- per modellare il processo di invio dei messaggi nello scenario di *Delay-Tolerant Network* caratterizzato da *epidemic routing* [4];
- per analizzare come il fenomeno di trasmissione, che include la divulgazione di informazioni, innovazioni, idee, tendenze, ecc., è influenzato dalla struttura della rete sottostante [5].

Quando ci si accinge alla costruzione di un modello che descriva la dinamica di propagazione di un virus, la conoscenza della rete di contatti sociali è un aspetto fondamentale. Per uno studio efficace di una rete, sia essa di comunicazione, virtuale o di interazioni umane reali, è importante considerare il ruolo giocato dalla mobilità nella dinamica del processo studiato, per esempio:

- in [11] si presenta un modello flessibile e pratico che è adatto per implementare e supportare molte caratteristiche della distribuzione e mobilità degli utenti per eseguire simulazioni delle dinamiche della rete, destinate allo studio delle tecniche di gestione delle risorse in reti cellulari;
- in [12] è svolto uno studio del controllo sull'ammissione al servizio nelle reti UMTS: algoritmi noti sono valutati in termini di correttezza e di prestazioni generali quando sono presi in considerazione modelli realistici per la mobilità; l'approccio al controllo di ammissione alla chiamata che è consapevole della diversa mobilità tra gli utenti, che può essere tracciata dalle stazioni base a cui sono connessi, permette di formulare miglioramenti necessari per algoritmi che potrebbero risultare globalmente buoni.

Perciò sono necessarie la revisione e l'estensione dei modelli affermati in letteratura [1] per lo sviluppo di altri modelli nuovi e più sofisticati.

Il contributo di questa tesi è l'ideazione di un nuovo modello, di tipo markoviano, che tiene conto della mobilità nella rete analizzata. Ciò consente di studiare la diffusione di un virus all'interno di una popolazione di individui che si muovono tra stati diversi. Sulla base di tale modello teorico vengono poi sviluppati due programmi in Matlab e grazie alle simulazioni si possono valutare i cambiamenti causati dalla mobilità. I risultati dimostrano che la mobilità

ha un impatto maggiore sugli stati più transitori, favorendo l'aumento e la velocizzazione del contagio a causa della grande affluenza di infetti da altre aree della rete considerata.

I successivi capitoli di questa tesi sono organizzati come segue. Nel capitolo 2 si descrive il modello SIR (Suscettibili-Infetti-Rimossi), seguendo prima l'approccio deterministico e poi quello stocastico. Segue poi un'esposizione di lavori collegati al tema di questa tesi. Nel capitolo 3 si richiama la teoria delle catene di Markov per introdurre nel modello SIR il processo stocastico che descrive la mobilità. Inoltre vengono presentate le implementazioni pratiche che sono state realizzate con Matlab per simulare le dinamiche dell'epidemia in uno scenario di mobilità. Nel capitolo 4 si discutono i risultati raggiunti. Infine il capitolo 5 conclude la trattazione sviluppando anche possibili studi futuri.

Capitolo 2

Stato dell'arte

2.1 Modello SIR deterministico e stocastico

Il classico modello matematico per la diffusione di una malattia infettiva è il modello deterministico di Kermack and McKendrick del 1927 [13], conosciuto come il modello epidemiologico SIR, in cui le lettere S, I, R indicano rispettivamente i termini “Suscellibile”, “Infetto” e “Rimosso”. Il modello considera che la durata della malattia sia molto più breve della durata della vita della popolazione, per cui il numero N degli individui che compongono la popolazione è fisso e nascite e morti, non collegate alla malattia, sono ignorate.

Durante lo sviluppo dell'infezione ogni individuo si trova in uno stato specifico, in accordo con la sua condizione di salute, così la popolazione risulta suddivisa in tre gruppi:

- S, il gruppo dei suscettibili: individui in grado di contrarre la malattia e diventare infetti;
- I, il gruppo degli infetti: individui che sono stati contagiati e sono in grado di trasmettere la malattia ad altri individui;
- R, il gruppo dei rimossi: individui che, dopo aver contratto la malattia, sono morti oppure sono guariti diventando immuni al contagio.

La trasmissione della malattia è rappresentata dalla transizione dallo stato Suscellibile verso lo stato Infetto, mentre la guarigione dalla malattia dalla transizione da Infetto a Rimosso.

Il passaggio $S \rightarrow I$ è determinato da tre distinti fattori: la prevalenza di infetti, la sottostante rete di contatti della popolazione e la probabilità di successo di contagio dato il contatto.

Gli infetti si spostano nella classe dei rimossi, $I \rightarrow R$, quando hanno sconfitto l'infezione e la probabilità di tale movimento dipende da quanto tempo l'individuo ha passato nella classe degli infetti, il così detto “periodo infettivo”.

Lo schema dell'andamento è visualizzato in figura 2.1:

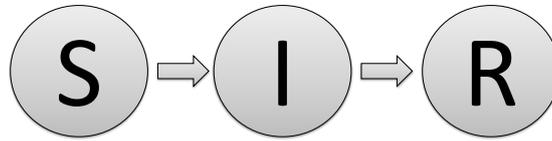


Figura 2.1: schema dell'evoluzione nel modello SIR

Per introdurre le equazioni che governano il modello si definiscono $S(t)$, $I(t)$ e $R(t)$, rispettivamente il numero di suscettibili, infetti e rimossi al tempo t .

Poiché si è ipotizzato che la popolazione totale rimane sempre in quantità costante pari alla somma dei tre compartimenti, risulta

$$S(t) + I(t) + R(t) = N$$

per ogni $t \geq 0$. Si suppone inoltre che [9]:

- non vi sia tempo di incubazione, ossia che ogni infetto sia immediatamente infettivo;
- il contagio avvenga per contatto diretto;
- la probabilità di incontro tra due qualsiasi individui della popolazione sia uguale: ipotesi di *small world* [14], ossia che tutti possano contattare tutti;
- ogni individuo malato abbia una probabilità di guarigione per unità di tempo costante, questo si traduce nel fatto che la probabilità di un individuo di muoversi dallo stato I allo stato R è indipendente da quanto tempo ha trascorso nella classe I (proprietà di assenza di memoria [15]).

Gli individui sono uniformemente mescolati all'interno della popolazione e il numero di nuovi contagiati, per unità di tempo, è proporzionale al numero di contatti tra individui nella classe S ed individui nella classe I . Pertanto si assume un numero di contagi proporzionale al prodotto SI . D'altra parte, nella stessa unità di tempo vi sono degli individui infetti che guariscono; per ipotesi ciò avverrà in proporzione al numero di malati. L'evoluzione dell'epidemia è descritta

quindi dalle seguenti equazioni differenziali, con β e γ parametri positivi [13]

$$\begin{aligned}\frac{dS(t)}{dt} &= -\beta S(t)I(t) \\ \frac{dI(t)}{dt} &= +\beta S(t)I(t) - \gamma I(t) \\ \frac{dR(t)}{dt} &= +\gamma I(t)\end{aligned}$$

Il parametro β è il coefficiente di trasmissione che moltiplicato per il numero di suscettibili e infetti rappresenta il tasso al quale nuovi individui sono contagiati.

Il parametro γ è chiamato tasso di guarigione e spesso si calcola il suo reciproco ($1/\gamma$) che determina il periodo medio di infezione [16].

Le condizioni iniziali sono rappresentate dal numero iniziale di individui suscettibili $S(0) > 0$, dal numero di infetti $I(0) > 0$ che sono introdotti nella popolazione e dal numero di rimossi $R(0) = 0$ che risulta nullo in quanto ci si trova nello stadio iniziale della malattia.

Come discusso in [13], la proporzione iniziale di suscettibili nella popolazione deve superare una soglia critica affinché sia possibile che l'epidemia si diffonda.

A questo risultato, conosciuto come il "fenomeno soglia" [1], si giunge riscrivendo l'equazione della dinamica degli infetti nella forma:

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

da cui si evince che se $S(0) < \gamma/\beta$ allora $\frac{dI(t)}{dt} < 0$ e quindi l'invasione dell'infezione non si verifica.

γ/β viene indicato come il tasso relativo di rimozione e il suo inverso è il *basic reproductive ratio*, indicato con il simbolo R_0 [16]. R_0 è definito come il numero atteso di infetti generati da ciascun individuo infetto in una grande popolazione di suscettibili. In epidemiologia è una tra le più importanti quantità e essenzialmente misura il massimo potenziale riproduttivo per una malattia [1].

Nel caso in cui si abbia un'epidemia è importante saper stimare in anticipo qual è la percentuale totale della popolazione che sarà colpita, ossia il limite di $S(t)$ per grandi t , nonché quale sarà la percentuale massima di contagiati raggiunta nella popolazione [9]. Studiare lo stato a lungo termine del modello consente di avere una misura dell'impatto dell'epidemia.

Dividendo l'equazione differenziale che descrive l'evoluzione del gruppo dei suscettibili per quella dei rimossi risulta:

$$\frac{dS}{dR} = -\frac{\beta S}{\gamma} = -R_0 S.$$

Integrando rispetto ad R si ottiene:

$$S(t) = S(0)e^{-R(t)R_0}.$$

Quindi mentre l'epidemia si sviluppa, il numero di suscettibili declina e quindi, con un ritardo dovuto alla durata della malattia, il numero di rimossi cresce. Si nota che S rimane sopra lo zero perché $e^{-R(t)R_0}$ è una quantità sempre positiva. Da ciò emerge che, poiché ci sarà sempre un numero di suscettibili nella popolazione che sfugge l'infezione, la catena della trasmissione si interromperà per il declino degli infetti e non per la mancanza di suscettibili.

In questo modello sono presenti due assunzioni che lo rendono più semplice pur mantenendo la capacità di rendere conto bene dei dati epidemiologici in molti casi.

La prima è che il numero di suscettibili, infetti e rimossi è deterministico [1], infatti le equazioni del modello descrivono tutto come delle equazioni di bilancio. La seconda è l'assenza della dipendenza dalla storia passata: il periodo infettivo è privo di memoria e tutto dipende dallo stato presente in cui ci si trova; per esempio se due individui sono ammalati da tempo diverso possono passare nella classe dei rimossi entrambi con lo stesso tasso di guarigione. In conclusione lo stato è una statistica sufficiente [17] perché dalla conoscenza del numero dei suscettibili e degli infetti è possibile ricavare tutte le transizioni per la proprietà di assenza di memoria.

Si tratta ora nuovamente il modello SIR ma come esempio di processo stocastico [9]. Le equazioni differenziali del modello deterministico sono equazioni al primo ordine che delineano l'andamento di quantità medie. L'approccio stocastico ha come obiettivo lo studio della statistica completa.

Parecchie ragioni suggeriscono che il modello stocastico è preferibile. Innanzitutto la via più naturale per descrivere la diffusione della malattia è stocastica perché si definisce la probabilità della trasmissione della malattia tra due individui piuttosto che affermare con certezza se il contagio avrà luogo oppure no [18]. Secondo, ci sono fenomeni che sono naturalmente stocastici. Per esempio, in una grande comunità, molti modelli potrebbero portare o a uno scoppio minore dell'infezione che colpisce pochi individui, o a uno maggiore in cui il contagio si estende a una più larga porzione della comunità. Il calcolo della probabilità dei due eventi è possibile solo con un modello stocastico. Infine un altro importante vantaggio riguarda l'aspetto delle stime e della predizione. La conoscenza dell'incertezza delle stime richiede un modello stocastico, e una predizione non è di grande utilità senza alcuna conoscenza della sua incertezza.

Un modello stocastico, che ha origine con Bartlett nel 1949 [19], conosciuto come *general stochastic epidemic* denota con $X(t)$ e $Y(t)$ il numero di suscettibili e il numero di infetti al tempo t .

I periodi in cui perdura la malattia degli infetti sono variabili aleatorie indipendenti e identicamente distribuite tra loro secondo la distribuzione di probabilità di una variabile aleatoria I .

Il processo $(X, Y) = [(X(t), Y(t)); t \geq 0]$ è un processo di Markov e si assume perciò che I sia esponenzialmente distribuita con intensità γ . Questa scelta non è motivata in senso epidemiologico ma piuttosto con questa assunzione l'analisi matematica diventa più semplice [18].

Durante il suo periodo infettivo un malato può avere contatti con dei suscettibili secondo le leggi probabilistiche di un processo di Poisson di intensità λ/n . I nuovi infetti diventano istantaneamente contagiosi e quindi capaci di infettare altri individui. Una persona è considerata rimossa quando il suo periodo di infettività è terminato e diviene immune a nuove infezioni, non prendendo così più parte alla diffusione della malattia.

Il processo (X, Y) è governato dalle seguente tabella di transizione:

da	verso	con tasso
(i,j)	(i-1,j+1)	$\lambda ij/n$
	(i,j-1)	γj

che segue immediatamente dalla definizione del modello secondo Markov [18].

2.2 Lavori collegati

I modelli epidemiologici tradizionali presentati nella sezione 2.1 non considerano gli schemi della mobilità umana o le caratteristiche della rete sociale di ogni individuo e quindi hanno una limitata capacità di catturare la complessità del processo di diffusione della malattia. In [20] si propone un *agent-based model* che usa gli schemi delle interazioni sociali e della mobilità individuale estratti dai dettagli delle registrazioni delle chiamate per modellare accuratamente la diffusione di un virus. Questo *ABM* simula l'epidemia e modella i comportamenti umani individuali rappresentando ogni persona come un *software agent* caratterizzato da diversi parametri che sono considerati rilevanti per la propagazione del virus (schemi di mobilità, caratteristiche della rete sociale, stato socio-economico, stato di salute, ecc.). Il comportamento di ogni *software agent* è definito con tre modelli: modello di mobilità, modello della rete sociale e modello della malattia che caratterizza la progressione del contagio. Il sistema proposto è stato applicato allo studio dello scoppio nel 2009 in Messico del virus H1N1 per valutare l'impatto delle azioni governative [21]. Sono stati simulati due scenari: uno costruito usando la mobilità e il modello sociale ottenuto durante il periodo precedente la malattia, quando gli individui mostrava un comportamento normale in assenza di allerte mediche e uno costruito con i dati a seguito delle azioni governative. Dal confronto nello scenario degli interventi il picco di infetti è minore e ritardato nel tempo.

L'aggiunta di semplici modelli dei processi di mobilità nell'analisi della diffusione geografica dell'epidemia apre la strada a molteplici effetti che possono avere un impatto maggiore sia nella comprensione dei dati epidemiologici sia nella definizione di interventi pubblici. In particolare

[22] considera un modello di un gruppo di popolazioni spazialmente separate che interagiscono tra loro come una rete i cui collegamenti denotano la mobilità degli individui (*metapopulation model*). Vengono incorporati i cambiamenti di iniziativa personale nella mobilità in risposta a uno scoppio di un'epidemia e si studia come queste reazioni influenzino la diffusione della malattia infettiva. Ogni *sottopopolazione* consiste di un numero di individui che sono divisi in ulteriori classi in accordo con il loro stato di salute (susceptibili, infetti, rimossi). Le dinamiche compartimentali interne modellano il contagio considerando che le persone nella stessa popolazione sono in contatto e potrebbero cambiare il loro stato di salute. Infine, c'è uno scambio di individui a seconda della mobilità tra le sottopopolazioni. Gli individui hanno memoria del loro originale sottopopolazione, alla quale ritornano dopo aver raggiunto il luogo della loro destinazione. Se il tasso di mobilità degli individui dell'iniziale sottopopolazione infetta non è abbastanza elevato per assicurare il contagio in un'altra prima del declino della locale epidemia, lo scoppio non si propaga globalmente. Questo è equivalente all'esistenza di un secondo R_0 a livello di ogni sottopopolazione che dipende dai parametri della mobilità e definisce il numero medio di sottopopolazioni infettate da ogni altra sottopopolazione che ha sperimentato l'invasione. I cambiamenti dovuti alla presenza della malattia possono avere un impatto diretto sulla mobilità e sulle abitudini di viaggio, dall'evitare le aree infette cambiando il percorso abituale alla più drastica reazione di non viaggiare affatto. In quest'ultimo caso la frazione di sottopopolazioni infette alla fine dello scoppio decresce a causa della riduzione del tasso di mobilità degli individui. Nel primo scenario invece si verifica un effetto controintuitivo in quanto il flusso crescente di individui che vanno per percorsi alternativi porta l'infezione a nuovi sottopopolazioni.

Anche in [23] si fornisce una espressione esplicita per la soglia di invasione globale. In questo modello ogni sottopopolazione è rappresentata da un nodo e la dinamica della malattia in ogni sottopopolazione segue il modello SIR. Si usa un approccio markoviano per descrivere la mobilità da un nodo all'altro, così ad ogni individuo è applicata la stessa probabilità di viaggio senza avere memoria della sua origine. L'espressione di soglia a cui si giunge può essere usata per determinare il minimo tasso di mobilità che assicura che in media ogni sottopopolazione sia in grado di infettare più di un sottopopolazione vicina.

In [24] viene fornita un'altra dimostrazione di come la rete di mobilità può essere incorporata nella classe di modelli che studiano la propagazione delle malattie. Una scoperta che è emersa da questi studi è che gli individui tipicamente visitano frequentemente soltanto un limitato numero di posti, effettuando principalmente spostamenti tra casa, lavoro e altri pochi luoghi. Di conseguenza gruppi di individui mostrano schemi di movimenti spazialmente forzati rispetto all'elevato potenziale tasso di mobilità. Ciò può alterare lo scenario della diffusione e le predizioni della dinamica della malattia. Il modello proposto tiene conto esplicitamente della

mobilità bidirezionale degli individui tra il loro unico luogo base (ad esempio la loro casa) e un piccolo insieme di altri luoghi. Un'interessante caratteristica che si evidenzia è l'esistenza di una ulteriore soglia di invasione che è determinata dal tasso di ritorno o equivalentemente dal tempo che un individuo spende fuori dal luogo base. Per un basso tasso di ritorno la frazione dell'intera popolazione afflitta durante l'epidemia è vicino all'unità, tuttavia con la crescita del valore del tasso di ritorno, la frazione decresce fino a riflettere l'assenza di uno scoppio globale. Le simulazioni dimostrano che è il tasso di ritorno piuttosto che il tasso di flusso tra i diversi luoghi a limitare la propagazione globale dell'infezione e a determinarne quindi il fallimento. Possono perciò essere messe in atto strategie di contenimento che non coinvolgono l'intera mobilità ma piuttosto, modificano gli schemi di mobilità in modo asimmetrico.

Capitolo 3

Mobilità e compartimentazione

3.1 Catena di Markov

Il contributo di questa tesi è di introdurre nel modello epidemico due processi aleatori per studiare la diffusione di una malattia in uno scenario di mobilità, uno che modella la dinamica della malattia e l'altro la mobilità degli individui.

Sia $X(t)$ il processo aleatorio per la mobilità e si considerino casi per cui il sistema può trovarsi in n stati distinti $\in S$, numerati da un intero. Si ricorre a una catena di Markov a tempi discreti perché il numero di malati è discreto e si assume che la malattia non sia monitorata in tempo reale ma che le osservazioni mediche siano finite; inoltre si opta per un processo markoviano perché di tale tipo è anche il processo che modella la dinamica della malattia.

Entrambi i processi si svolgono nel tempo a passi discreti, $t \in \mathbf{Z}$, con la stessa granularità e si denota per convenienza con Δt il passo temporale; adottando un Δt piccolo si riduce l'errore di quantizzazione e si ottiene una buona approssimazione del caso continuo.

La probabilità di ogni stato al tempo $t + \Delta t$ dipende dallo stato del sistema al tempo t ma non dalla sua storia precedente e la probabilità di avere nello stesso istante più di due transizioni risulta trascurabile [26].

Riferendosi agli istanti di tempo discreti come “passi”, si definiscono le probabilità di transizione un passo dallo stato i allo stato j come:

$$p_{i,j} = P[X(t + \Delta t) = j | X(t) = i].$$

Analogamente, la matrice di transizione un passo, semplicemente conosciuta come *matrice di transizione* è così indicata [25]:

$$P = \begin{bmatrix} p_{11} & p_{12} & \cdots \\ p_{21} & p_{22} & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix}$$

Si nota che l' i -esima riga della matrice di transizione \mathbf{P} raccoglie tutte le possibili probabilità di transizione dallo stato i allo stato $j \in S$. Sommando tutti questi termini si ottiene:

$$\sum_{j=1}^n p_{i,j} = \sum_{j=1}^n P[X(t + \Delta t) = j | X(t) = i] = P[X(t + \Delta t) \in S | X(t) = i].$$

Poiché l'evento $X(t + \Delta t) \in S$ ha probabilità uno per ogni $t \geq 0$, si conclude che gli elementi di ogni riga di \mathbf{P} sommano a uno:

$$\sum_{j=1}^n p_{i,j} = 1.$$

Una matrice con questa proprietà, cioè con elementi non negativi e che sommano a uno in ogni riga, è chiamata *matrice stocastica* [26].

Il *sojourn time* s_j [25] in uno stato j è definito come il numero di passi consecutivi che un processo aleatorio spende nello stato j prima di muoversi verso un altro stato. A ogni passo, la probabilità che si verifichi una transizione dallo stato j verso se stesso è $p_{j,j}$, mentre la probabilità di lasciare lo stato j è $1 - p_{j,j}$, non tenendo conto del tempo già speso nello stato j per la proprietà di Markov. Perciò, il numero di passi che il processo spende nello stato j prima di muoversi verso un altro stato è una variabile aleatoria con una distribuzione geometrica di parametro $1 - p_{j,j}$.

Spesso è utile rappresentare la matrice di transizione \mathbf{P} con grafo chiamato *diagramma di transizione*, come in figura 3.1.

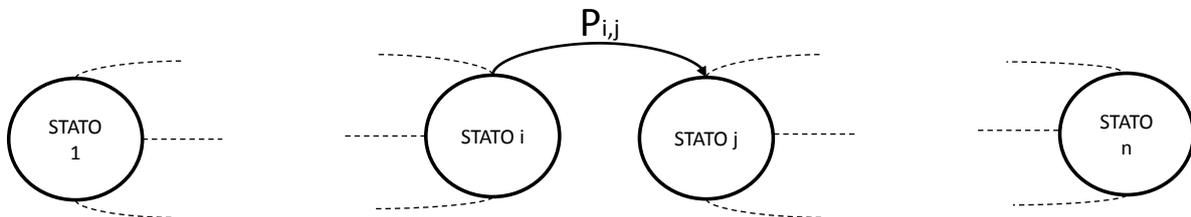


Figura 3.1: esempio di diagramma di transizione

I nodi del grafo sono in una corrispondenza uno a uno con gli stati dello spazio S della catena di Markov. Gli archi nel diagramma di transizione sono in corrispondenza a uno a uno con gli elementi positivi di \mathbf{P} . Nello specifico, per ogni elemento $p_{i,j} > 0$ di \mathbf{P} , c'è un arco che parte

dal nodo i e entra nel nodo j , che è etichettato come $p_{i,j}$. Si nota che i è connesso a j se lo stato della catena di Markov può cambiare da i a j in un passo solo con probabilità non nulla. La somma degli archi pesati che escono da ogni stato è sempre uguale a uno.

La probabilità assoluta che la catena di Markov sia in uno stato $j \in S$ al passo t -esimo è:

$$p_j(t) = P[X(t) = j].$$

che indica la probabilità di distribuzione di massa del processo aleatorio, che può essere rappresentata attraverso il *vettore della probabilità di stato*.

$$p(t) = [p_0(t), p_1(t), \dots, p_n(t)]$$

dove

$$\sum_{j=1}^n p_j(t) = 1$$

per la condizione di normalizzazione. Inoltre dal teorema della probabilità totale [17] si ottiene che il vettore della probabilità di stato ad ogni istante di tempo è interamente definito dalla conoscenza della matrice di transizione \mathbf{P} e dell'iniziale vettore delle probabilità di stato $p(0)$. Spesso è di interesse esaminare la distribuzione asintotica del processo. Il vettore asintotico della probabilità di stato $\boldsymbol{\pi} = [\pi_1, \pi_2, \dots, \pi_n]$, dove π_j è la probabilità che il processo sia nello stato j in qualche momento arbitrariamente lontano nel futuro, è definito come:

$$\pi_j = \lim_{t \rightarrow \infty} p_{i,j}(t)$$

quando il limite esiste e non dipende da i . Il vettore asintotico delle probabilità di stato, se esiste ed è indipendente dal vettore delle probabilità iniziali $p(0)$, descrive le caratteristiche del processo quando raggiunge il comportamento a regime, dopo che gli effetti delle condizioni iniziali sull'evoluzione del processo sono cessati.

Il vettore $\boldsymbol{\pi}$ si può ricavare quando la catena considerata è connessa, cioè da ogni stato si può raggiungerne un altro e dunque non si ha dipendenza dallo stato iniziale, e quando la matrice di transizione ha rango pieno e non comporta la formazione di anelli nella catena.

Dopo aver verificato queste condizioni, si possono adottare due differenti strategie per eseguire il calcolo [27]. La prima consiste nella diretta applicazione della definizione, che richiede l'analisi della matrice $\mathbf{P}(t) = \mathbf{P}^t$ con t che tende all'infinito. La seconda comporta la risoluzione del seguente sistema di equazioni:

$$\begin{cases} \boldsymbol{\pi}(\mathbf{I} - \mathbf{P}) = \mathbf{0} \\ \boldsymbol{\pi}\mathbf{1} = 1 \end{cases}$$

in cui la seconda equazione corrisponde alla condizione di normalizzazione e con $\mathbf{0}$ e $\mathbf{1}$ vettori di n elementi.

3.2 Il modello teorico

Il modello classico SIR è stato ampliato con lo scopo di studiare la dinamica dell'evoluzione di una malattia in una popolazione chiusa che si muove tra stati diversi. Questo modello esteso è stato conseguito ipotizzando che la mobilità e la diffusione dell'infezione siano due processi indipendenti.

Lo scenario prevede L stati diversi e una popolazione di N individui che, ad ogni passo, si spostano secondo la catena di Markov a tempi discreti precedentemente illustrata.

La presenza di aree diverse in cui si possono trovare gli individui modifica la dinamica del contagio. Per tener conto di tale effetto si è applicata l'ipotesi di *small world* [14] ad ogni stato indipendentemente, assumendo che la classe di infetti di uno stato può avere contatti e quindi eventualmente contagiare solo la classe di suscettibili che si trova nello stesso stato; inoltre si è ipotizzato che durante lo spostamento le persone non si infettino tra loro e ogni stato ha lo stesso coefficiente β di trasmissione e tasso γ di guarigione [1].

La mobilità è descritta dalla matrice di transizione \mathbf{P} :

$$P = \begin{bmatrix} p_{11} & p_{12} & \dots & \dots \\ p_{21} & p_{22} & \dots & \dots \\ \vdots & \vdots & \ddots & \\ p_{L1} & \dots & \dots & p_{LL} \end{bmatrix}$$

e le equazioni deterministiche, poiché tengono conto della mobilità tra le L aree, sono così definite:

$$\frac{dS_j(t)}{dt} = -\beta S_j(t)I_j(t) + \sum_{i \neq j} p_{ij}S_i(t) - \sum_{i \neq j} p_{ji}S_j(t) \quad (3.1)$$

$$\frac{dI_j(t)}{dt} = +\beta S_j(t)I_j(t) - \gamma I_j(t) + \sum_{i \neq j} p_{ij}I_i(t) - \sum_{i \neq j} p_{ji}I_j(t) \quad (3.2)$$

$$\frac{dR_j(t)}{dt} = +\gamma I_j(t) + \sum_{i \neq j} p_{ij}R_i(t) - \sum_{i \neq j} p_{ji}R_j(t) \quad (3.3)$$

Il parametro R_0 , definito nella sezione 2.1 come il numero medio di infezioni causate dal nucleo iniziale di infetti quando il resto della popolazione è suscettibile, nel semplice modello SIR, coincide con quello che si ottiene applicando la formula per determinare il tasso medio con cui aumentano gli infetti $R_0 = (N\beta)/\gamma$, che stabilisce la soglia dell'epidemia.

In questo modello più complicato, per R_0 è stato calcolato un valore ottenuto risolvendo il nuovo sistema di equazioni differenziali, intendendo R_0 come il valore di soglia.

Prima di procedere alla risoluzione del sistema è necessario calcolare il vettore asintotico della probabilità di stato $\boldsymbol{\pi} = [\pi_1, \pi_2, \dots, \pi_L]$, che descrive le caratteristiche del processo a regime, e stimare la percentuale di persone che si trovano nei diversi stati.

A questo punto si continua con il calcolo di R_0 sommando le equazioni differenziali che descrivono l'evoluzione della classe degli infetti:

$$\frac{dI_1(t)}{dt} + \frac{dI_2(t)}{dt} + \dots + \frac{dI_L(t)}{dt} = \beta(S_1I_1 + S_2I_2 + \dots + S_LI_L) - \gamma(I_1 + I_2 + \dots + I_L)$$

Supponendo che gli individui siano già separati negli L stati secondo queste equazioni:

$$\begin{cases} S_1 = \pi_1 S \\ S_2 = \pi_2 S \\ \dots \\ S_L = \pi_L S \end{cases} \quad \begin{cases} I_1 = \pi_1 I \\ I_2 = \pi_2 I \\ \dots \\ I_L = \pi_L I \end{cases}$$

e andando a sostituire tali valori si ricava:

$$\frac{dI}{dt} = \beta S I (\pi_1^2 + \pi_2^2 + \dots + \pi_L^2) - \gamma I$$

Da ciò si può dedurre che $\frac{dI}{dt} > 0$ quando è verificata la seguente disuguaglianza:

$$\frac{\beta S (\pi_1^2 + \pi_2^2 + \dots + \pi_L^2)}{\gamma} > 0$$

con S uguale al numero iniziale di suscettibili. Essendo il fattore $(\pi_1^2 + \pi_2^2 + \dots + \pi_L^2) < 1$, nel modello considerato l'effetto soglia di R_0 si abbassa.

3.3 Le implementazioni pratiche

Come per il modello SIR esistono due differenti approcci, anche per questo modello esteso sono state realizzate due implementazioni pratiche, una deterministica e una stocastica, con l'ausilio di Matlab [28].

Per le simulazioni si considera lo scenario di *small world* [14] e di contagio a compartimenti come discusso nella sezione 3.2 e si suppone che la discretizzazione della malattia corrisponda a quella della mobilità.

Si imposta il numero di aree a 3 e quindi in entrambi i programmi la mobilità è descritta dalla

seguinte matrice di transizione \mathbf{P} :

$$P = \begin{bmatrix} p_{11} & p_{12} & p_{13} \\ p_{21} & p_{22} & p_{23} \\ p_{31} & p_{32} & p_{33} \end{bmatrix}$$

che porta allo scenario rappresentato nel diagramma di transizione di figura 3.2.

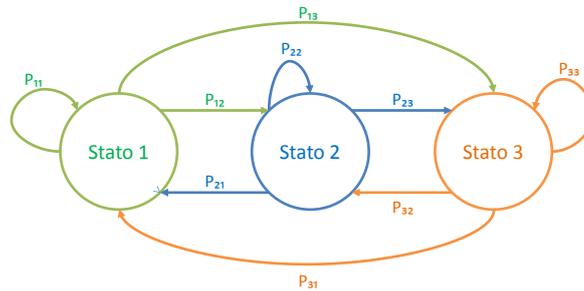


Figura 3.2: diagramma di transizione a tre stati delle implementazioni pratiche

Si sceglie un passo di campionamento Δt e un intervallo di durata T per cui si monitora la popolazione, effettuando un totale di $T/\Delta t$ rilevazioni.

Simulazione stocastica

Nella versione stocastica le costanti fondamentali del modello sono P_{cont} , la probabilità di contagio, e P_{guar} , la probabilità di guarigione.

Vengono creati nove vettori (`esseuno`, `iuno`, `erreuno`, `essedue`, `idue`, `erredue`, `essetre`, `itre`, `erretre`) per memorizzare le metriche di prestazione.

Con due vettori di N elementi, il vettore `posizione` e il vettore `contagio`, si tiene traccia dello stato in cui si trova ciascuna persona e della sua condizione di salute. La posizione di ogni utente è indicata con un numero intero corrispondente allo stato in cui si trova, mentre gli interi 1,2,3, elementi del vettore `contagio`, corrispondono rispettivamente all'essere suscettibile, infetto e rimosso di ciascuna persona. Per esempio se `contagio(j) = 2` e `posizione(j) = 3` allora l'individuo j -esimo è attualmente contagiato dall'infezione e si trova nello stato tre.

Nella fase di inizializzazione, in base alle condizioni iniziali che si desidera modellare, vengono impostati gli elementi dei vettori sopra descritti.

In questa versione mobilità e diffusione dell'infezione sono due processi entrambi aleatori ma

indipendenti e si assume che le frazioni di individui che si ammalano e che guariscono sono dei processi bernoulliani in cui ogni suscettibile ha una probabilità P_{cont} di ammalarsi a causa di un contatto con un individuo malato in una unità di tempo e ogni infetto ha una probabilità P_{guar} di guarire nella stessa unità di tempo, allora ciascuna delle due variabili risulta una variabile binomiale.

Per simulare lo spostamento della persona j -esima all'istante di tempo $t + \Delta t$ si effettua un ciclo su tutti gli N individui della popolazione. Per ogni persona j -esima si considera la corrispondente riga $posizione(j)$ -esima della matrice \mathbf{P} e, in base ai valori delle tre probabilità, si divide l'intervallo $I = [0, 1]$ in tre sotto intervalli; si estrae un numero casuale uniformemente distribuito tra 0 e 1 e si verifica a quale sotto intervallo appartiene per stabilire la posizione successiva occupata dall'individuo.

Per quanto riguarda il processo SIR si realizza un altro ciclo sempre su tutti gli individui. Per ogni persona prima si verifica che sia contagiosa controllando il valore dell'elemento corrispondente nel vettore contagio; in caso affermativo si individuano gli suscettibili che si trovano nella sua stessa area scorrendo il vettore posizione. Per ogni possibile contatto si estrae un numero casuale uniformemente distribuito tra 0 e 1 e se risulta minore della probabilità P_{cont} allora il contagio avviene con successo; in tal caso si aggiorna lo stato di salute del nuovo contagiato.

Dopo il contagio si simula la rimozione: ciascun infetto ha una probabilità P_{guar} di guarire e diventare immune al virus. Con un ciclo su tutti gli utenti si estrae un numero casuale uniformemente distribuito tra 0 e 1 per tutti gli individui infetti: se è minore di P_{guar} l'utente guarisce, in caso contrario permane nel suo stato contagioso. Si tiene traccia degli utenti che guariscono modificando correttamente il valore che indica il loro stato di salute nel vettore contagio.

Ad ogni rilevazione i vettori delle metriche di prestazione registrano il numero di suscettibili, infetti e rimossi in ognuno dei tre stati. Rappresentando graficamente questi vettori si può avere una visione della curva dell'epidemia in ciascun stato.

Valutazione secondo il modello deterministico

Nella versione deterministica non c'è nessuna aleatorietà perché si calcola la frazione di persone presenti in ogni area allo stesso tempo. La mobilità individuale non viene simulata ma si assume che la frazione di individui che si muovono da uno stato all'altro sia uguale alla media, così come avviene per il contagio.

Come nel modello SIR originale si considera che la trasmissione della malattia avvenga secondo il coefficiente di trasmissione β e la guarigione con un tasso γ . In questa simulazione si definiscono solo i nove vettori esseuno, iuno, erreuno, essedue, idue, erredue, essetre, itre, erretre.

A partire dalle nuove equazioni del modello teorico si ricavano le equazioni alle differenze che sono poi applicate in questo programma per simulare le dinamiche della mobilità e del contagio nei tre stati, con $j = 1, 2, 3$:

$$S_j(t + \Delta t) - S_j(t) = [-\beta S_j(t)I_j(t) + \sum_{i \neq j} p_{ij}S_i(t) - \sum_{i \neq j} p_{ji}S_j(t)]\Delta t \quad (3.4)$$

$$I_j(t + \Delta t) - I_j(t) = [+ \beta S_j(t)I_j(t) - \gamma I_j(t) + \sum_{i \neq j} p_{ij}I_i(t) - \sum_{i \neq j} p_{ji}I_j(t)]\Delta t \quad (3.5)$$

$$R_j(t + \Delta t) - R_j(t) = [+ \gamma I_j(t) + \sum_{i \neq j} p_{ij}R_i(t) - \sum_{i \neq j} p_{ji}R_j(t)]\Delta t \quad (3.6)$$

Capitolo 4

Risultati

In questo capitolo sono studiate le applicazioni pratiche del modello in casi concreti d'interesse. In ogni sottosezione si spiega e motiva lo studio svolto, si precisano lo scenario considerato e il valore dei parametri e, con l'ausilio di grafici, si commentano i risultati ottenuti. Nello specifico nella sezione 4.1 si valuta l'accuratezza dei due modelli costruiti; nella sezione 4.2 si mostra come la compartimentazione cambia il propagarsi dell'infezione attraverso due confronti con il modello statico: nel primo la rete unica non mobile ha gli stessi parametri di quella mobile invece nel secondo nella rete unica statica i parametri sono modificati con lo scopo di ottenere l'uguaglianza dei picchi; la sezione 4.3 è dedicata all'analisi delle metriche fondamentali del processo in quattro differenti condizioni di mobilità e poi si focalizza solo su una di queste metriche, esaminando come si modifica al variare della probabilità di guarigione combinato con il diverso flusso di spostamento degli individui; infine nella sezione 4.4 si confronta il valore di R_0 teorico con quello trovato sperimentalmente dalle simulazioni.

4.1 Deterministico vs Stocastico

Per verificare la correttezza dei due modelli proposti, deterministico e stocastico, si è deciso di confrontare i risultati ottenuti da entrambi.

Nelle simulazioni i dati sono stati impostati con i seguenti valori:

- numero totale di individui, $N = 1000$;
- numero iniziale di infetti, $I_0 = 3$;
- intervallo di monitoraggio, $T = 300$;
- coefficiente di trasmissione nella versione deterministica e probabilità di contagio in quella stocastica, rispettivamente $\beta = 0.0003$ e $P_{cont} = \beta\Delta t$, rispettivamente;

- tasso di guarigione nella versione deterministica e probabilità di guarigione in quella stocastica, rispettivamente $\gamma = 0.05$ e $P_{guar} = \gamma\Delta t$;

- la matrice di transizione, $P = \begin{bmatrix} 0.916 & 0.084 & 0 \\ 0.25 & 0.5 & 0.25 \\ 0 & 0.125 & 0.875 \end{bmatrix}$;

- vettore asintotico della probabilità di stato, $\boldsymbol{\pi} = [0.498, 0.1673, 0.3347]$.

I valori delle probabilità di transizione da uno stato all'altro sono stati assegnati in modo arbitrario sulla base di quanto tempo ciascun utente potrebbe trascorrere in ogni stato. Ad esempio in [29] si mostra come si possono ottenere analisi della mobilità usando i videogiochi. In questo modo è facile creare nuovi mondi virtuali e adattarli a bisogni specifici per predire e valutare l'impatto della mobilità nel sistema.

Si è ipotizzato che tutti gli utenti partano dallo stato 1 e di conseguenza sono stati inizializzati i vettori del programma.

Si mettono ora a confronto i due modelli realizzati dai grafici risultanti dalle simulazioni Matlab, in cui sono rappresentati i vettori delle metriche di prestazione. Sono stati ricavati quattro grafici: uno per la dinamica di contagio di ciascun stato, 4.1-4.3, e uno in cui è delineata la dinamica totale su tutta la popolazione, 4.4.

Per tutti i grafici sono riportati sia i risultati analitici previsti dal modello deterministico che quelli sperimentali ricavati con una singola simulazione del modello stocastico.

Si può evincere una buona corrispondenza dei risultati ottenuti con i due modelli. Nella simulazione, inoltre, è presente un ulteriore rumore causato dall'effetto di discretizzazione della popolazione. Infatti non si può avere un numero frazionario di suscettibili, infetti o rimossi perché gli individui sono atomici.

Figura 4.1: stato 1: modello deterministico vs stocastico

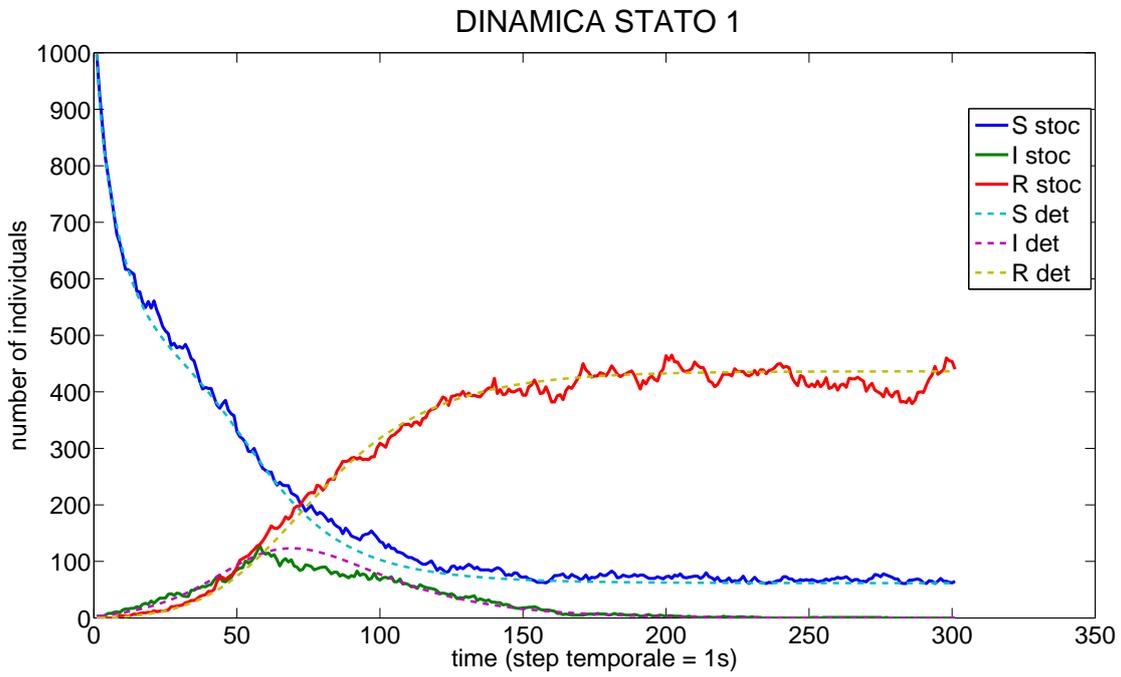


Figura 4.2: stato 2: modello deterministico vs stocastico

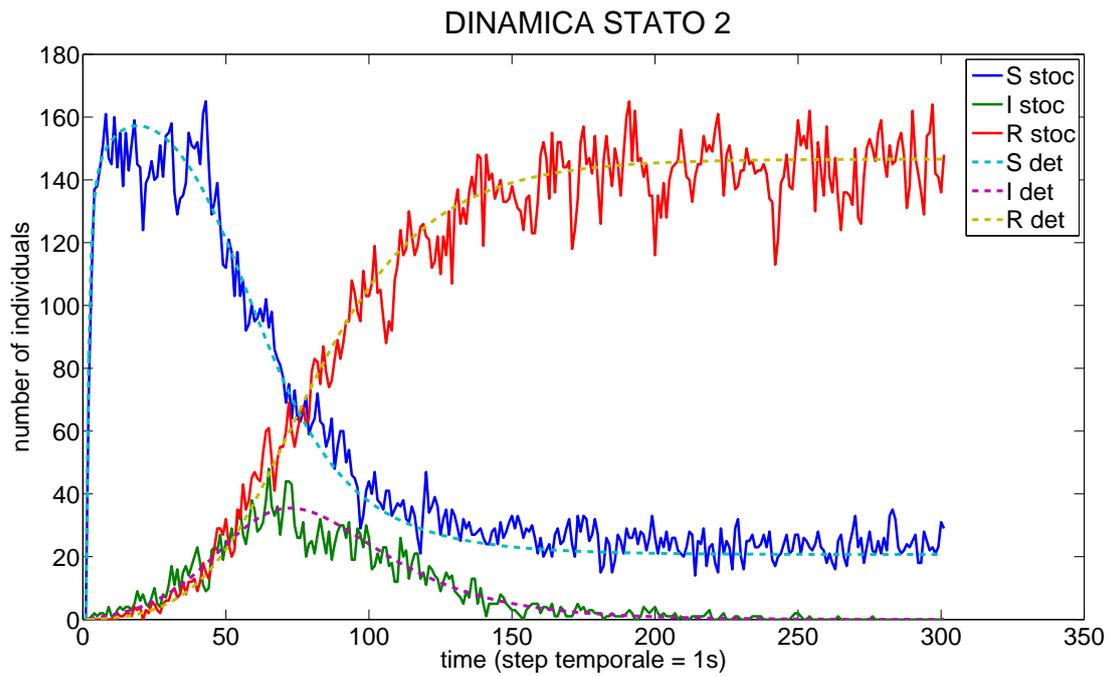


Figura 4.3: stato 3: modello deterministico vs stocastico

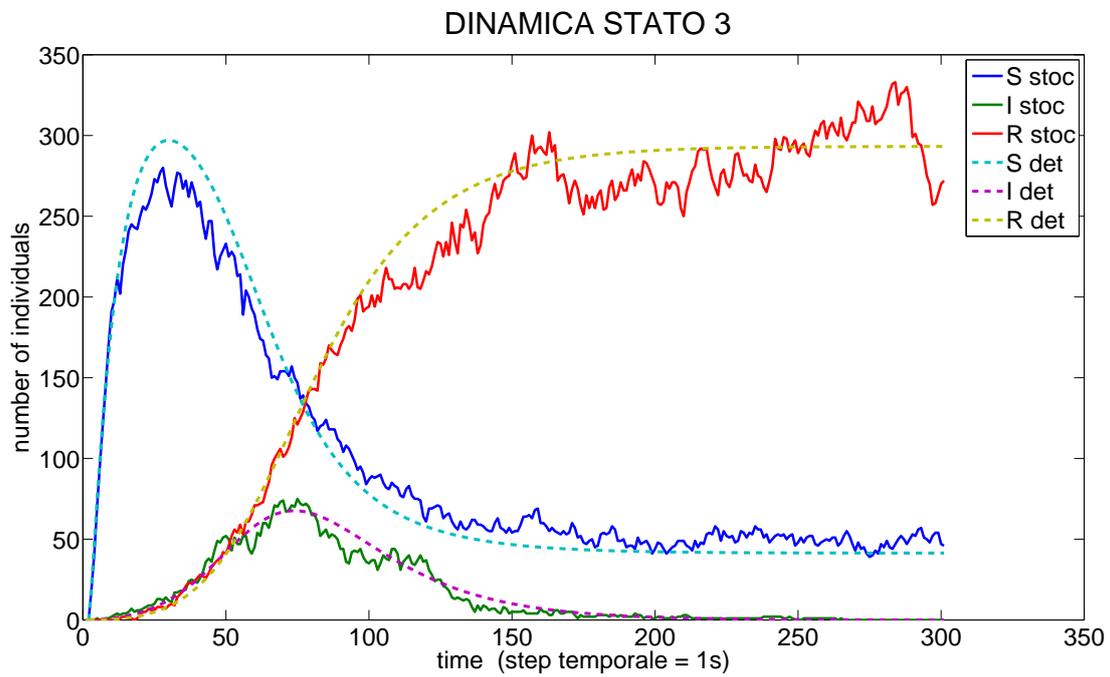
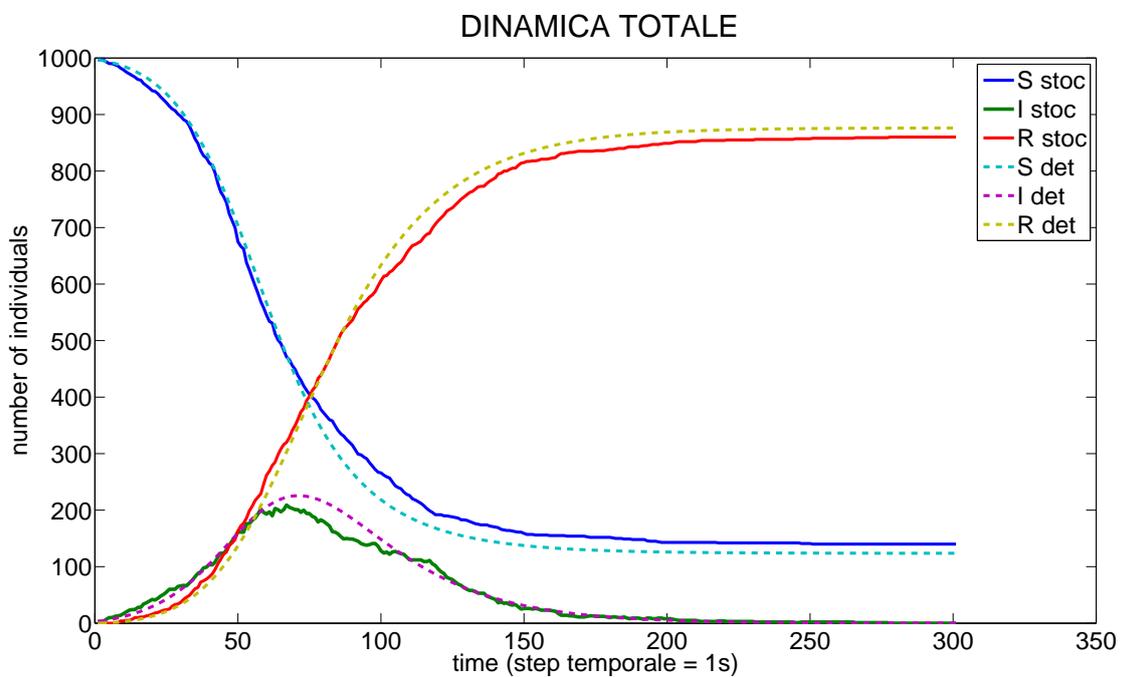


Figura 4.4: dinamica totale: modello deterministico vs stocastico



4.2 Effetto della compartimentazione

Per valutare l'effetto della compartimentazione sulla diffusione della malattia si è svolto un confronto tra i grafici di ogni stato, ottenuti con il modello deterministico, e il grafico ottenuto con il modello SIR base, cioè in cui gli utenti si trovano tutti in un unico stato e in assenza di mobilità.

Per il modello SIR base il numero di utenti dell'unico stato cambia a seconda dello stato con cui si vuole effettuare il confronto, rispettando sempre il numero atteso in base alle probabilità asintotiche:

- $N_1 = 500$ e $I_0 = 1$;
- $N_2 = 160$ e $I_0 = 1$;
- $N_3 = 300$ e $I_0 = 1$.

Per il passo di campionamento, l'intervallo di monitoraggio, β , γ e la matrice \mathbf{P} sono stati usati gli stessi valori del caso con i tre stati.

Per lo stato 1 (grafico 4.5) nel caso della mobilità il picco di infetti è anticipato nel tempo ed è di valore inferiore rispetto al caso con un unico stato.

Per lo stato 2 (grafico 4.6) nel caso in assenza di mobilità nell'intervallo di tempo considerato la malattia non si diffonde.

Per lo stato 3 (grafico 4.7) nel caso della mobilità il picco di infetti è anticipato nel tempo e di valore maggiore rispetto al caso di un unico stato.

Si è poi studiato come variano β e γ , lasciando invariati i restanti parametri, nel modello sir base affinché sia raggiunto lo stesso picco di infetti del modello nuovo.

Per ottenere la coincidenza di picchi per lo stato 2 e 3 (grafici 4.9 e 4.10) è stato necessario aumentare il valore del coefficiente di trasmissione e del tasso di guarigione, ottenendo che nel caso di mobilità il picco di infetti è ritardato nel tempo rispetto al caso di unico stato. Invece per lo stato 1 (grafico 4.8) i parametri β e γ sono stati diminuiti e il picco di infetti è anticipato nel tempo, contrariamente a quanto avviene negli altri stati.

Figura 4.5: stato 1: modello mobile vs statico con parametri uguali

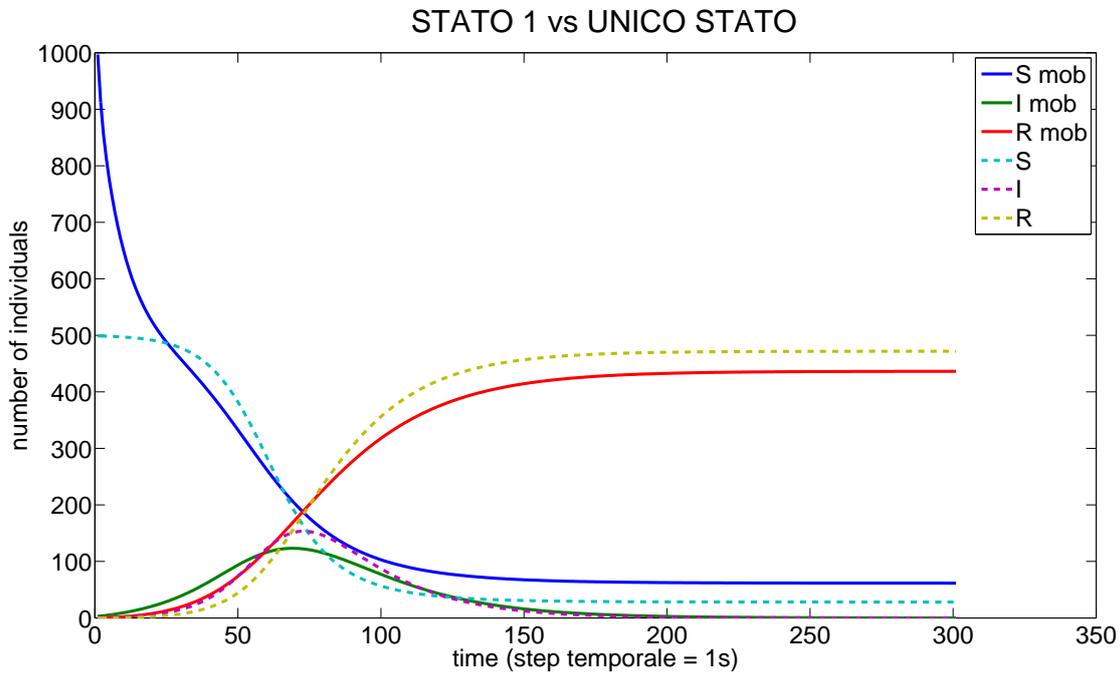


Figura 4.6: stato 2: modello mobile vs statico con parametri uguali

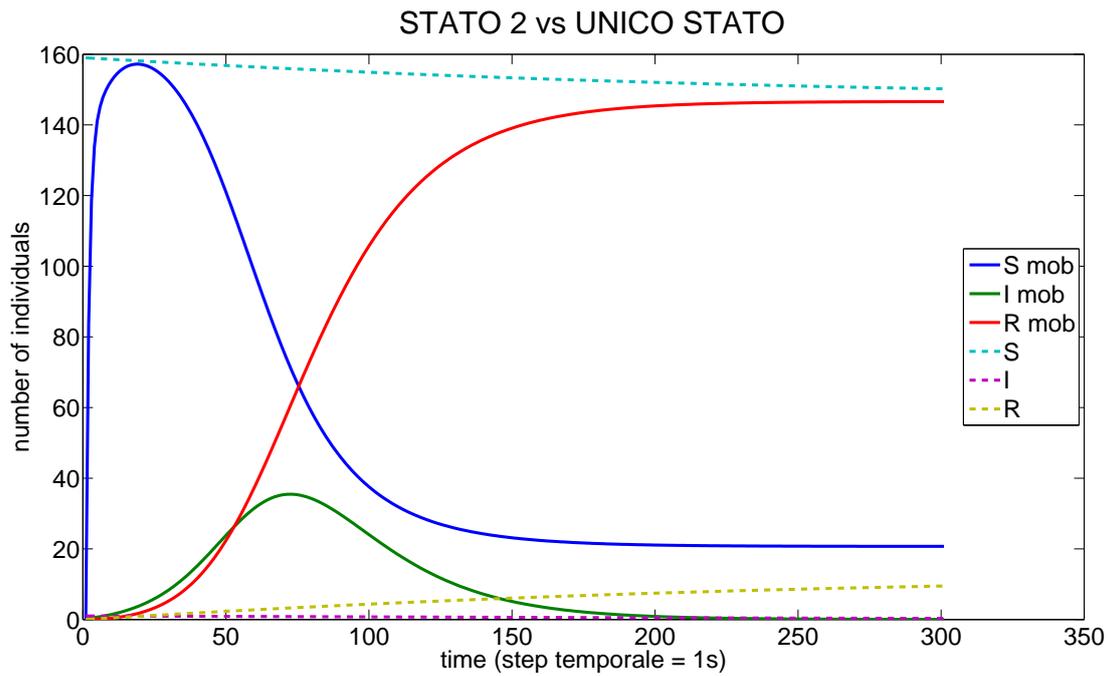


Figura 4.7: stato 3: modello mobile vs statico con parametri uguali

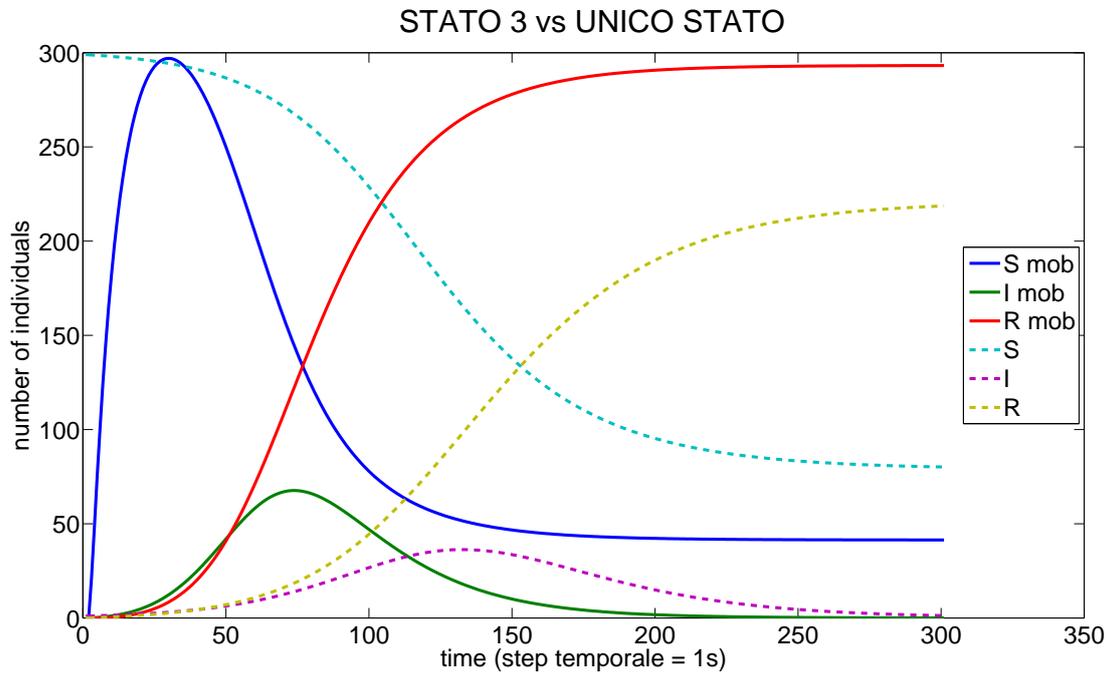


Figura 4.8: stato 1: modello mobile vs statico con picchi uguali con $\beta = 0.00021$ e $\gamma = 0.041$ nel modello statico

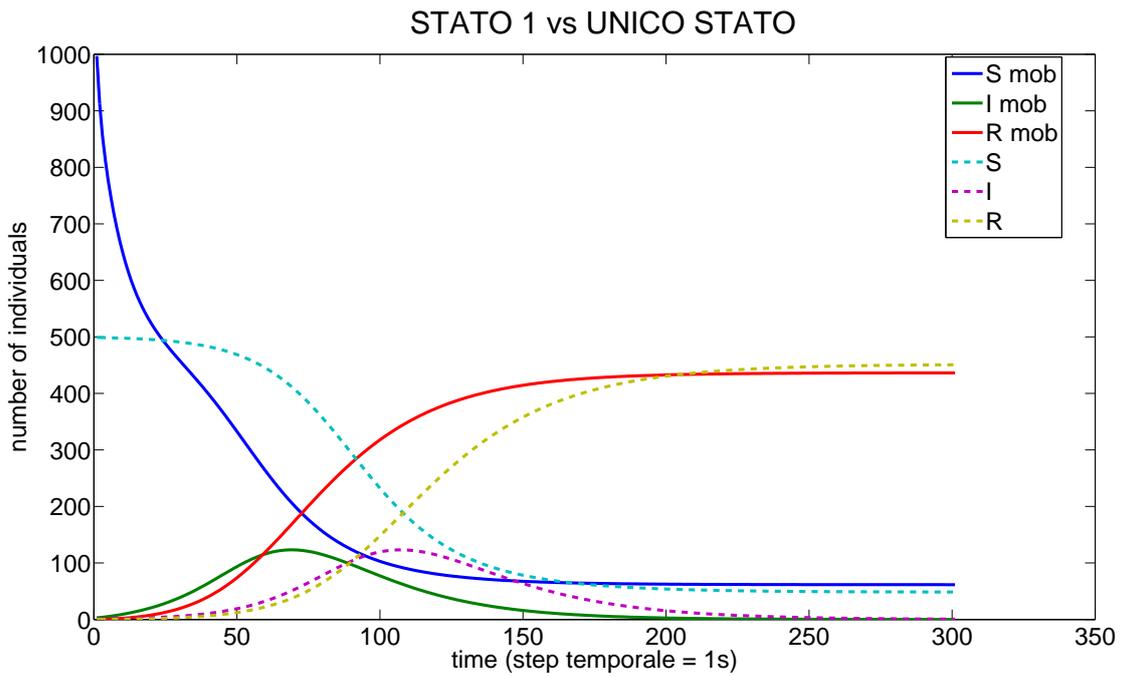


Figura 4.9: stato 2: modello mobile vs statico con picchi uguali con $\beta = 0.0015$ e $\gamma = 0.1$ nel modello statico

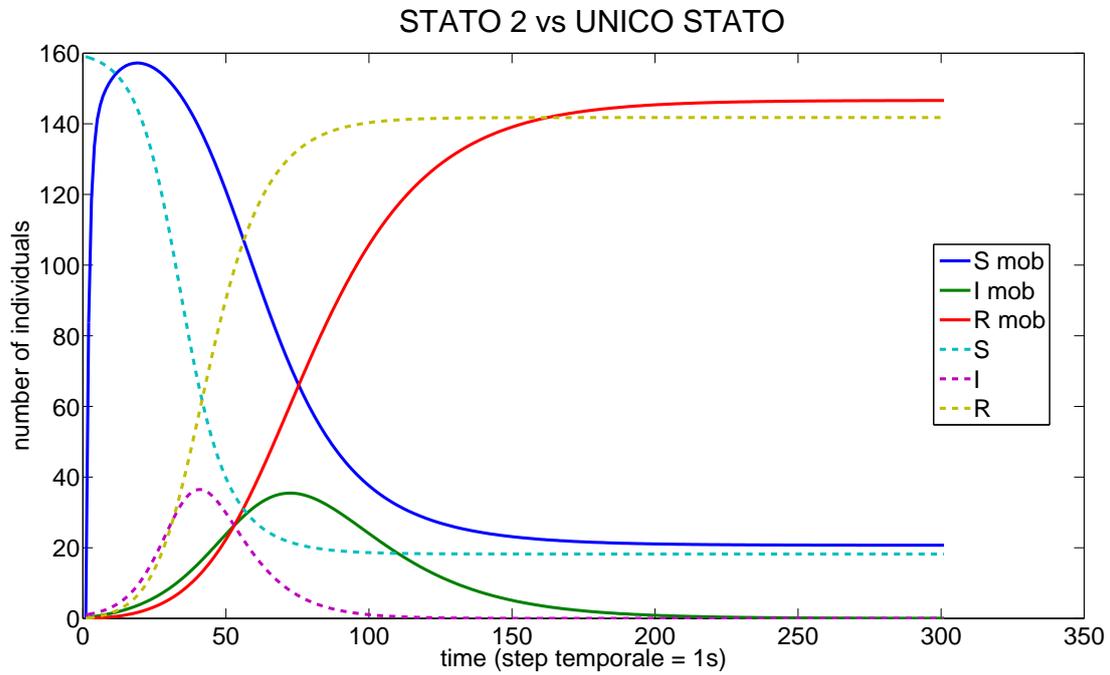
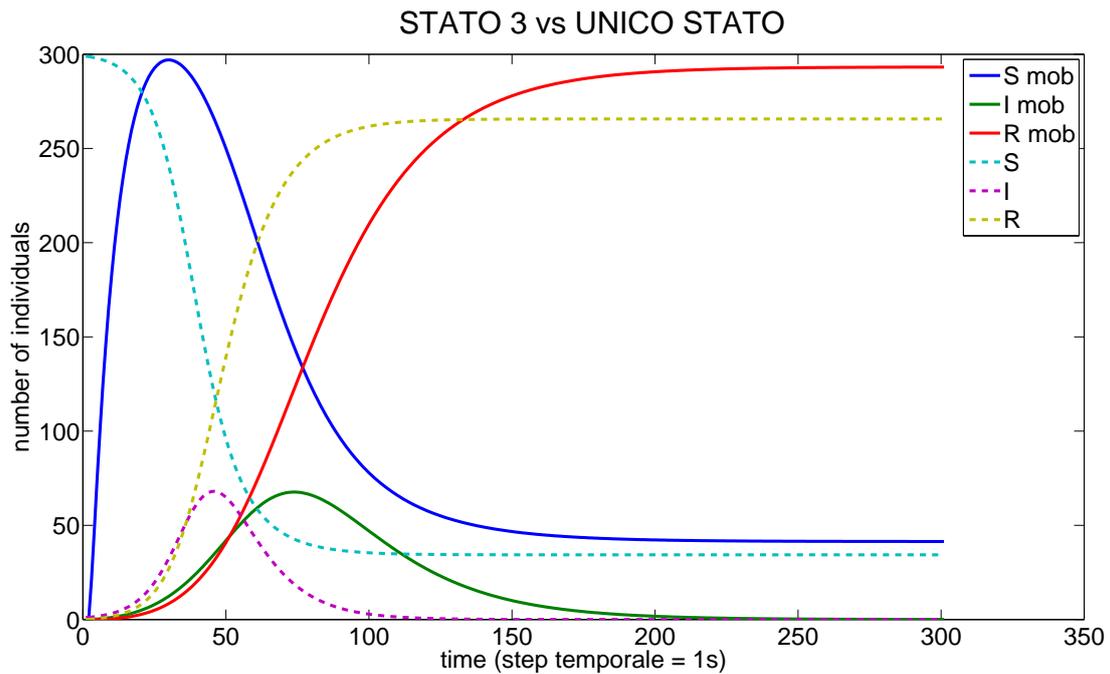


Figura 4.10: stato 3: modello mobile vs statico con picchi uguali con $\beta = 0.0008$ e $\gamma = 0.1$ nel modello statico



4.3 Correlazione del movimento

Per un modello di propagazione di una malattia in uno scenario di mobilità è interessante valutare l'impatto che può avere la diversa correlazione del movimento sul processo stesso e sulle metriche considerate [26]. A tal fine a partire dalla matrice \mathbf{P} , introdotta nella sezione 4.1, sono state ricavate altre tre matrici che hanno le stesse probabilità asintotiche ma un diverso grado di correlazione:

- la matrice indipendente e identicamente distribuita, le cui righe sono gli elementi del vettore asintotico delle probabilità di stato

$$P_{iid} = \begin{bmatrix} 0.498 & 0.1673 & 0.3347 \\ 0.498 & 0.1673 & 0.3347 \\ 0.498 & 0.1673 & 0.3347 \end{bmatrix}$$

- una matrice più correlata

$$P_C = \begin{bmatrix} 0.95 & 0.033584337 & 0.016415663 \\ 0.148834429 & 0.8 & 0.051165571 \\ 0 & 0.05 & 0.95 \end{bmatrix}$$

- una matrice molto più correlata

$$P_{MC} = \begin{bmatrix} 0.97 & 0.013431726 & 0.016568273 \\ 0.089300657 & 0.9 & 0.010699342 \\ 0 & 0.03 & 0.97 \end{bmatrix}$$

Per ogni matrice si calcola la media della variabile aleatoria geometrica *sojourn time* dei tre stati con la formula $m_{s_j} = \frac{1}{1-p_{j,j}}$:

	P_{iid}	P	P_C	P_{MC}
m_{s_1}	2	12	20	33
m_{s_2}	1	2	5	10
m_{s_3}	1.5	8	20	33

Si definisce poi il parametro σ che descrive il diverso grado di correlazione della mobilità. Esso quantifica per le quattro matrici la media del numero di istanti di tempo consecutivi che il processo spende nei tre stati:

- $\sigma_{iid} = 1.5$;

- $\sigma_P = 7.3$;
- $\sigma_C = 15$;
- $\sigma_{MC} = 25.3$.

Il parametro quindi indica la durata media della permanenza in una posizione e perciò più correlata è la matrice, più alto è σ .

Da ciò si deduce che la matrice indipendente e identicamente distribuita è poco correlata, cioè il tasso di spostamento tra stati è maggiore perché gli individui stazionano per un tempo minore nello stesso stato. Le altre matrici invece presentano un grado di correlazione crescente e di conseguenza la mobilità tra stati è ridotta.

Nelle simulazioni di questa sezione gli individui sono partiti divisi nei tre stati, rispettando il numero atteso in base alle probabilità asintotiche.

Per ogni matrice sono state eseguite 100 simulazioni per calcolare la media e la varianza del massimo numero di infetti in ogni stato e per calcolare il numero di volte in cui la malattia non si propaga.

Quando il movimento è poco correlato il contagio si diffonde nei tre stati in modo più uniforme, invece se la probabilità di spostamento da uno stato all'altro è bassa ci sono zone molto contagiate e altre meno.

Come si può constatare dal grafico 4.11:

- nello stato 1 nonostante il diverso grado di correlazione del movimento la media del massimo numero di infetti converge all'incirca allo stesso valore;
- nello stato 2, che è la posizione più transitoria, la media cala con l'aumentare della correlazione;
- nello stato 3 la differenza maggiore è tra caso poco correlato e caso correlato, indipendentemente dal grado di correlazione, mostrando la stessa tendenza che si registra per lo stato 2;

Per quanto riguarda la varianza (grafico 4.12), a parte nel caso dello stato 2, tende ad aumentare con l'aumentare della correlazione. Anche il numero di volte in cui la malattia non si propaga cresce quando è più elevato il grado di correlazione (grafico 4.13).

Quindi in ambiente di mobilità oltre al calcolo della media del massimo numero di infetti è interessante anche calcolare la deviazione standard in ogni stato poiché mette in evidenza l'oscillazione rispetto alla media. Infatti a seconda della correlazione del movimento lo sviluppo del contagio può cambiare e diffondersi diversamente e questo si può evincere dalla varianza.

Figura 4.11: effetto della correlazione: media del massimo numero di infetti

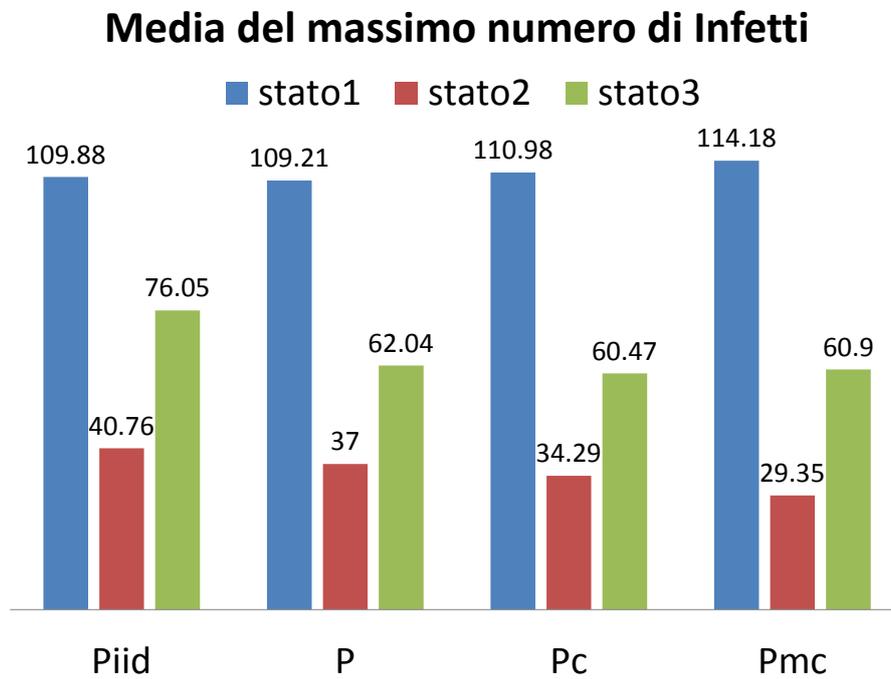


Figura 4.12: effetto della correlazione: varianza del massimo numero di infetti

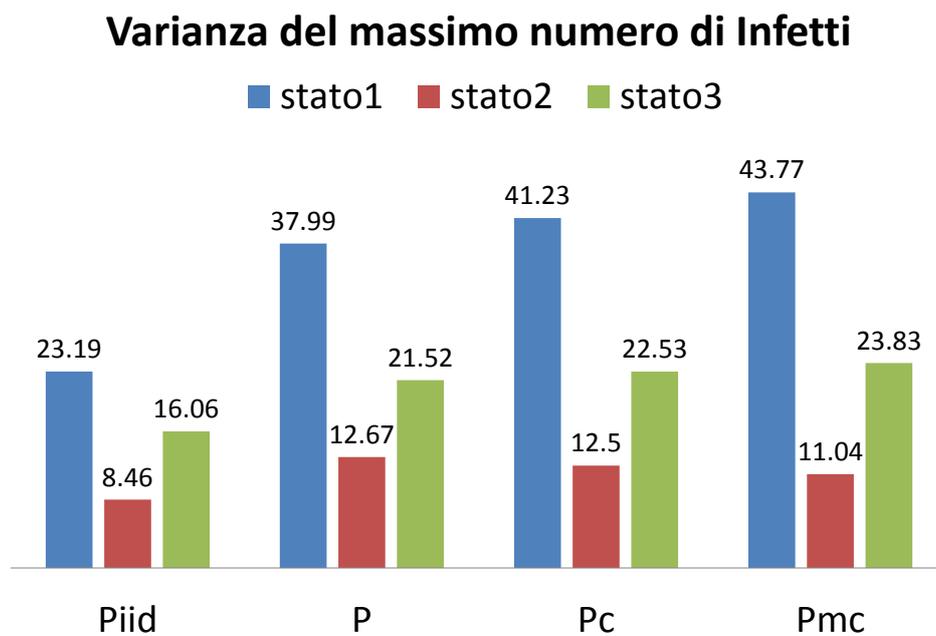
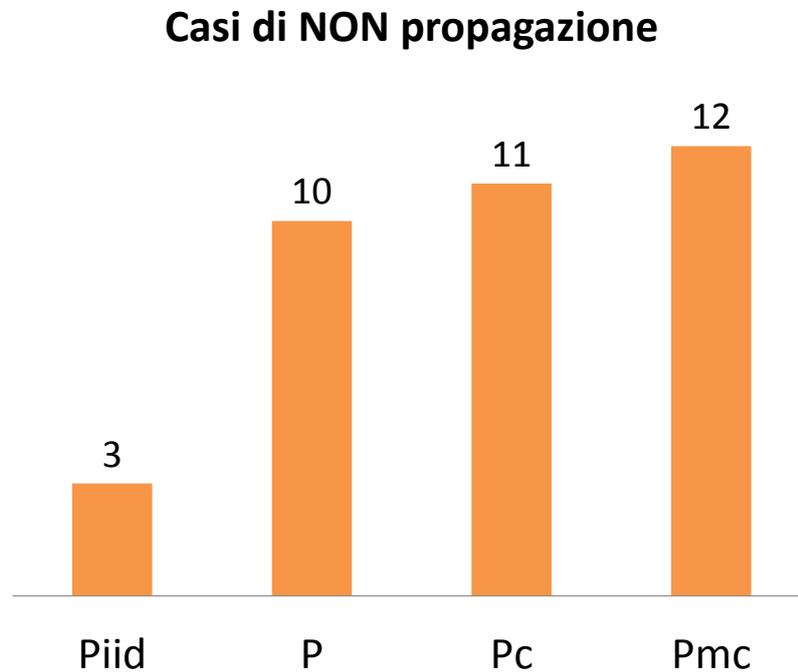


Figura 4.13: effetto della correlazione: casi in cui la malattia non si diffonde



Per ognuna delle quattro matrici si è anche registrata la variazione del massimo valore di infetti in ogni stato, variando la probabilità di contagio nell'intervallo $[0.0001; 0.001]$, con passo di campionamento di 0.0001.

I grafici 4.14-4.16 sono il risultato della media di 10 esecuzioni per ogni valore di probabilità considerato.

La differenza in caso di mobilità è maggiore tanto più piccola è la probabilità asintotica dello stato considerato. In particolare nello stato 2 che è lo stato più transitorio, cioè con π_2 minore, si riscontra una differenza maggiore per la diversa correlazione del movimento mentre nello stato 1, che è quello con la probabilità asintotica più grande, non si ha quasi nessuna differenza. L'aumentare della mobilità tra stati comporta un aumento del numero degli infetti negli ambienti più mobili. Apparentemente sembrerebbe un risultato controintuitivo perché quando la probabilità di restare nello stato 2 è più elevata, caso di maggior correlazione, il contagio dovrebbe diffondersi più velocemente perché gli infetti rimangono più a lungo sempre nello stesso stato. Però anche gli infetti degli altri stati si comportano di conseguenza. Quindi nel caso di matrice poco correlata c'è un'influenza non solo dei malati che vengono generati nello stato 2 ma anche dei malati che vengono dagli altri stati, che porta ad un aumento del massimo numero di infetti.

Anche dalle equazioni 3.5 si deduce che all'aumentare della correlazione i contagiati "in loco"

aumentano però diminuiscono i contagiati che arrivano da fuori:

Figura 4.14: stato 1: massimo numero di infetti al variare di β

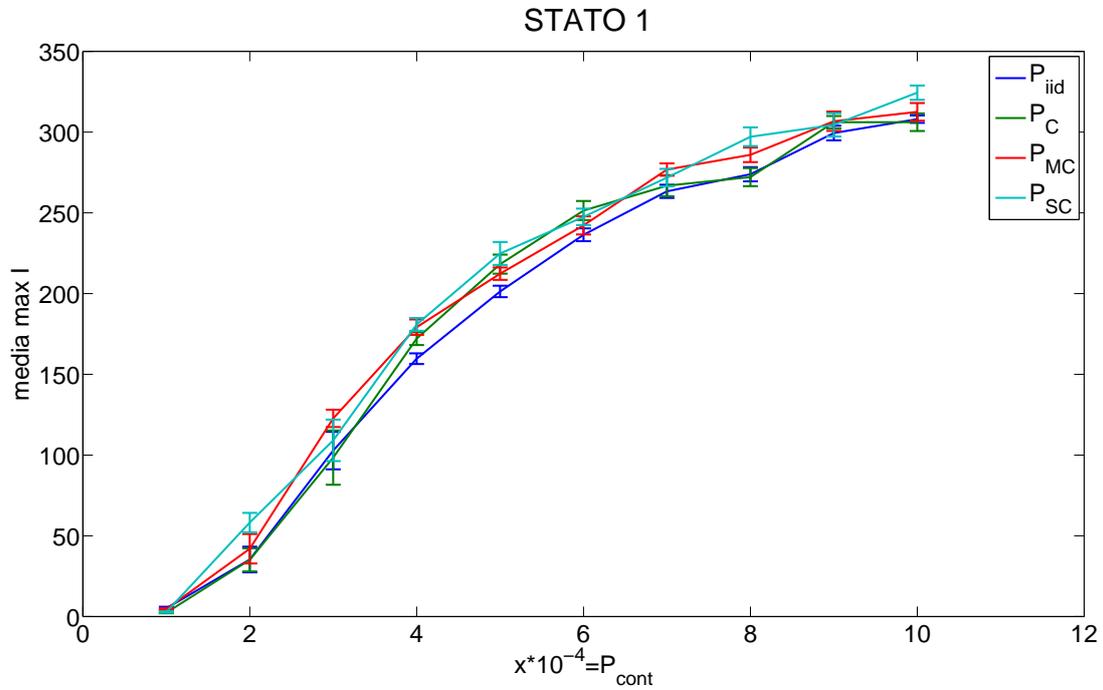


Figura 4.15: stato 2: massimo numero di infetti al variare di β

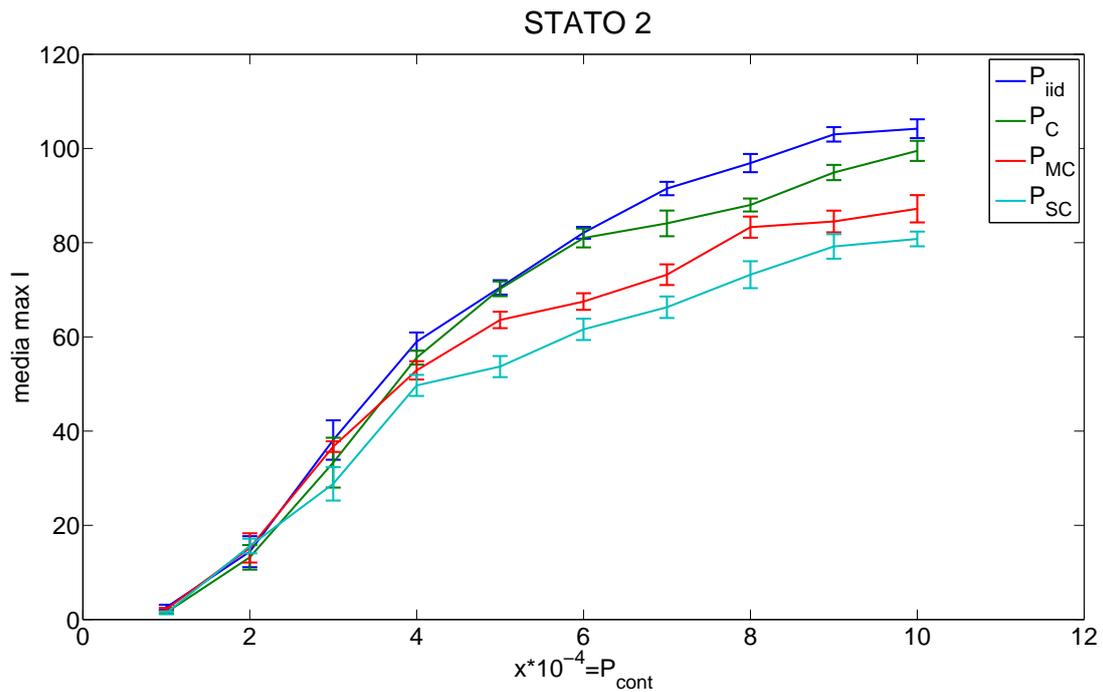
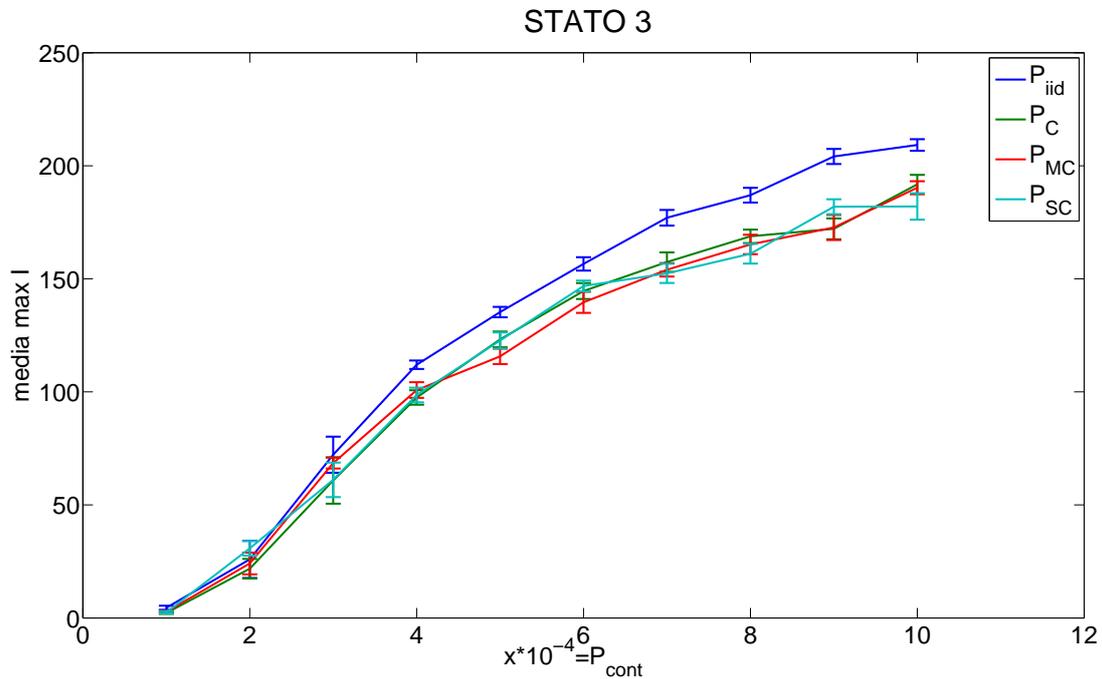


Figura 4.16: stato 3: massimo numero di infetti al variare di β 

4.4 Calcolo di R_0

Da quanto scritto nella sezione 3.2 in merito al parametro R_0 , con l'ipotesi di tre stati, si ricava che la nuova formula per il calcolo teorico è:

$$R_0 = \frac{(N - I_0)\beta(\pi_1^2 + \pi_2^2 + \pi_3^2)}{\gamma}$$

e sostituendo i valori, precisati in 4.1, si ottiene $R_0 = 2.32$.

Per R_0 è stato calcolato anche un valore sperimentale ricavato dalla simulazione stocastica, applicando in modo diretto la definizione: numero medio di infezioni causate dal nucleo iniziale di infetti quando il resto della popolazione è suscettibile.

A tale scopo nel programma è stato introdotto il vettore `contagiati` di N elementi con il quale si è tenuto traccia del numero di individui che sono contagiati dagli infetti iniziali I_0 e al termine delle rilevazioni si calcola $R_0 = \text{sum}(\text{contagiati}(1 : I_0))/I_0$.

Per ottenere un valore confrontabile con quello teorico nella simulazione all'inizio gli utenti partono divisi secondo le probabilità asintotiche del vettore di stato.

Per poter stimare una media attendibile sono state effettuate 100 esecuzioni ed è risultato $R_0 = 2.38$ con una deviazione di 1.5, che avvalorava quanto trovato in teoria.

Il massimo potenziale riproduttivo di ogni singolo stato in condizioni statiche:

- stato 1: $R_0 = 2.99$ con 499 suscettibili iniziali;
- stato 2: $R_0 = 0.95$ con 159 suscettibili iniziali;
- stato 3: $R_0 = 1.79$ con 299 suscettibili iniziali;

è confrontato con quello dei tre stati, determinato in condizioni di mobilità.

La mobilità produce un aumento del massimo potenziale riproduttivo negli stati con una minore probabilità asintotica, addirittura determinando lo scoppio epidemico nello stato 2 in cui, per il fenomeno soglia, non si sarebbe verificato.

In conclusione dai risultati del capitolo emerge che la compartimentazione modifica le curve epidemiche rispetto a quelle che si ottengono con un unico stato e il suo effetto principale è quello di incrementare e velocizzare il contagio negli stati più transitori.

Trattandosi di uno studio in uno scenario di mobilità è stato osservato l'impatto della correlazione del movimento. Dall'analisi è emerso che il contagio può cambiare a seconda della mobilità degli individui, diffondendosi in modo diverso da quanto atteso. Inoltre con l'intensificarsi del grado di correlazione cresce il numero di infetti nello stato però diminuisce l'afflusso di contagiati da altri stati. Di conseguenza l'aumentare della mobilità tra stati favorisce l'incremento del numero degli infetti negli ambienti più mobili.

Capitolo 5

Conclusioni e sviluppi futuri

In questa tesi si è discussa la diffusione di contenuti epidemici in ambienti di mobilità e si è scoperto che il movimento degli individui cambia il modo in cui avviene la propagazione dei contenuti.

Si è espanso il modello teorico introducendo un processo di mobilità markoviano nella caratterizzazione SIR delle epidemie in una popolazione di individui che si muovono tra stati diversi. Si è poi costruita un'implementazione duplice in Matlab, ideando una versione deterministica e una stocastica.

Si è verificato che negli stati più transitori e con un numero atteso di individui più piccolo, a parità di parametri, se non c'è mobilità la diffusione non solo è inferiore ma anche non sarebbe avvenuta. Infatti per raggiungere lo stesso livello di propagazione nel caso statico è necessario aumentare la contagiosità.

Questo consente di studiare scenari più realistici ed è interessante perché fornisce un altro aspetto su cui agire per controllare la trasmissione.

Nel modello presentato tutti gli individui si muovono indipendentemente, però in realtà alcune persone nelle reti si muovono assieme. Uno sviluppo futuro potrebbe essere studiare cosa succede se la mobilità ha un schema di questo tipo che causa un'aggregazione dei movimenti. Con questo scopo nell'articolo [30] si discute come sviluppare un modello che tenga conto di questo ulteriore aspetto della mobilità e che sia facilmente adattabile a differenti scenari. Inoltre affinché sia un utile strumento per le simulazioni si sono cercati parametri che presentino un significato intuitivo e una facile impostazione proprio perché spesso si desidera valutare la dipendenza dei risultati dai parametri.

Il modello SIR è principalmente adoperato per lo studio di un'epidemia, che si vorrebbe non si diffondesse ma può essere riadattato con l'obiettivo di capire meglio la propagazione, per esempio, di messaggi utili, per i quali si vorrebbe aumentare e migliorare la diffusione. Un ulteriore studio allora potrebbe essere quello di sfruttare la mobilità per incrementare l'efficienza

della trasmissione delle informazioni come in [31] o degli interventi per contrastare le epidemie. Infine a partire dalle consapevolezza della mobilità di gruppo si apre la futura possibilità per fare predizioni e migliorare la gestione di processi, non solo riguardo a strategie di contenimento per le epidemie ma anche per sfruttare dinamiche analoghe in ambito di reti ICT, come le strategie di routing per utenti mobili in reti wireless [32].

Bibliografia

- [1] M. J. Keeling, P. Rohani, *Modeling Infectious Diseases in Humans and Animals*, Princeton University Press, 2007.
- [2] C. Cobelli, E. Carson, *Introduzione alla modellistica in fisiologia e medicina*, Pàtron Editore, Bologna 2012.
- [3] A. Khelil, C. Becker, J. Tian, K. Rothermel, “*An epidemic model for information diffusion in MANETs*”, Proceedings of the 5th ACM international workshop on Modeling analysis and simulation of wireless and mobile systems, 2002.
- [4] G. M. de Souza Dias, R. M. Salles, “*Epidemic SIR model applied to Delay-Tolerant Networks*”, Brazilian Telecommunication Symposium, Settembre 2012.
- [5] M. Opuszko, J. Ruhland “*Impact of the network structure on the SIR model spreading phenomena in online networks*”, Proceedings of International Multi-Conference on Computing in the Global Information Technology, Luglio 2013.
- [6] D. J. Daley, D. G. Kendall, “*Epidemics and Rumours*”, Nature vol.204, 12 Dicembre 1964.
- [7] D. P. Maki, M. Thompson, *Mathematical models and applications: with emphasis on the social, life, and management sciences*, Prentice-Hall, 1973.
- [8] F. Jin, E. Dougherty, P. Saraf, Y. Cao, N. Ramakrishnan, “*Epidemiological modeling of news and rumors on Twitter*”, Proceedings of the 7th Workshop on Social Network Mining and Analysis, 2013.
- [9] G. Gaeta, *Modelli matematici in biologia*, Springer, 2007.
- [10] A. Jafarabadi, M. A. Azgomi, “*An SIR model for the propagation of topology-aware active worms considering the join and leave of hosts*”, Proceedings of International Conference on Information Assurance and Security (IAS), 2011.
- [11] S. Nousiainen, K. Kordybach, P. Kemppi, “*User distribution and mobility model framework for cellular network simulation*”, Proceedings of IST Mobile and Wireless Communications Summit 2002.
- [12] L. Badia, M. Zorzi, A. Gazzini, *On the impact of user mobility on call admission control in WCDMA systems*, Proceedings VTC 2002 Fall, Vancouver, BC, vol. 1, pp. 121–126, Settembre 2002.

- [13] W.O. Kermack , A.G. McKendrick, “*A Contribution to the Mathematical Theory of Epidemics*”, Proc. Roy. Soc. Lond. A vol. 115, pp. 700-721, 1927.
- [14] D. J. Watts, S. H. Strogatz, “*Collective dynamics of “small-world” network*”, Nature vol. 393, pp. 440-442,4 June 1998.
- [15] A.V. Oppenheim, A.S. Willsky, *Signals and systems*, Prentice-Hall, 1997.
- [16] R. M. Anderson, R. M. May, *Infectious Diseases of Humans: Dynamics and Control*, Oxford University Press, Oxford, 1991
- [17] J.A. Gubner, *Probability and Random Processes for Electrical and Computer Engineering*, Cambridge University Press, 2006.
- [18] Hakan Andersson, Tom Britton, *Stochastic epidemic models and their statistical analysis*,Febbraio 2000.
- [19] M. S. Bartlett “*Some Evolutionary Stochastic Processes*”, Journal of the Royal Statistical Society, Series B (Methodological), Vol. 11, No. 2 (1949), pp. 211-229, Wiley.
- [20] E. Frias-Martinez, G. Williamson, V. Frias-Martinez, “*An agent-based model of epidemic spread using human mobility and social network information*”, Proceedings of IEEE International Conference on Privacy,Security, and Trust, 2011.
- [21] R. Oliveira, A. Karatzoglou, P. Concejero, A. Armenta, N. Oliver, “*Towards a psychographic user model from mobile phone usage*”, Proceedings of ACM Int. Conf. on Human Factors in Computing Systems, Vancouver, Canada, 2011.
- [22] S. Meloni, N. Perra, A. Arenas, S. Gomez, Y. Moreno e A. Vespignani, “*Modeling human mobility responses to the large-scale spreading of infectious diseases*”, Scientific Reports, Agosto 2011.
- [23] V. Colizza, A. Vespignani, “*Invasion threshold in heterogeneous metapopulation networks*”, Febbraio 2008.
- [24] V. Belik, T. Geisel, D. Brockmann, “*Natural human mobility patterns and spatial spread of infectious diseases*”, Physical Review X 1, 2011.
- [25] N. Benvenuto e M. Zorzi (editori), *Principles of communications networks and systems*, Wiley, 2011.
- [26] L. Badia, “*On the impact of correlated arrivals and errors on ARQ delay terms*”, IEEE Transactions on Communications, vol. 57, pp. 334–338, Febbraio 2009.
- [27] M. Rossi, L. Badia, M. Zorzi, “*SR ARQ delay statistics on N-state Markov channels with finite round trip delay*”, IEEE Transactions on Wireless Communications, vol. 5, pp. 1526-1536, Giugno 2006.
- [28] Matlab, versione 7.12.0.635 (R20011a), 64-bit (win64), Marzo 2011

- [29] F. Fitzek, L. Badia, M. Zorzi, G. Schulte, P. Seeling, T. Henderson, “*Mobility and stability evaluation in wireless multi-hop networks using multiplayer games*”, Proceedings ACM NetGames 2003, San Francisco, CA, pp. 177-185, Maggio 2003.
- [30] L. Badia, N. Bui, “*A group mobility model based on nodes’ attraction for next generation wireless networks*”, Proceedings IEE Mobility 2006, Bangkok, Thailand, Ottobre 2006.
- [31] M. Rossi, L. Badia, P. Giacom, M. Zorzi, “*Energy and connectivity performance of routing groups in multi-radio multi-hop networks*”, Wireless Communications and Mobile Computing, vol.8, pp.327-342, Marzo 2008.
- [32] L. Badia, N. Bui, M. Miozzo, M. Rossi, M. Zorzi, “*Mobility-aided routing in multi-hop heterogeneous networks with group mobility*”, Proceedings IEEE Globecom, Washington (DC), USA, Novembre 2007.