



**Università degli studi “Magna Græcia”
di Catanzaro**

SCUOLA DI MEDICINA E CHIRURGIA

**CORSO DI LAUREA MAGISTRALE IN
INGEGNERIA BIOMEDICA**

Tesi di Laurea Sperimentale

*Disegno e Sviluppo di un ambiente software per la
simulazione di Processi su Reti*

Relatore
Chiar.mo Prof.
Pietro Hiram Guzzi

Candidata
Martina De Salazar
Matr.244393

Anno Accademico 2023/2024

INDICE

INTRODUZIONE	4
---------------------------	----------

CAPITOLO 1

“PROCESSI SU RETI”

1.1 PROPRIETA' GENERALI	7
1.1.1 APPLICAZIONI DELLA BIOINFORMATICA.....	10
1.2 TEORIA DELLE RETI	11
1.2.1 CENNI STORICI	12
1.2.2 CONCETTI CHIAVE DELLE RETI	14
1.3 MODELLI SU RETI	21
1.3.1 MODELLI COMPARTIMENTALI	26

CAPITOLO 2

“DIFFUSIONE DELL'INFORMAZIONE”

2.1 DINAMICA SU RETI	30
2.2 MODELLI A CASCATA	32
2.3 CENTRALITA' DEI NODI	40
2.4 ATTORI CHIAVE NELLA DIFFUSIONE	45

2.5 CENTRO E PERIFERIA: GUIDA ALLA DIFFUSIONE DELLE INFORMAZIONI.....	48
2.5.1 SIMULAZIONI RETI	
MODELLO & REALI.....	49
2.5.2 ANALISI DI STUDI PRECEDENTI	51
2.5.2.1 MASSIMIZZAZIONE DELL'INFLUENZA : “ IDENTIFICARE I NODI CHIAVE ”.....	62
2.6 SICUREZZA E GESTIONE DEI RISCHI NELLE RETI COMPLESSE	65

CAPITOLO 3

“PIANO DI STUDIO”

3.1 PANORAMICA SUL MORBILLO	69
3.1.1 IL MORBILLO IN ITALIA	74
3.1.2 SINTOMI E PREVENZIONE	75
3.2 MONITORAGGIO E ANALISI	79
3.3 STUDIO PROPOSTO	81
3.3.1 FASI DI SVILUPPO	83
3.3.2 SIMULAZIONI	87
3.4 DISCUSSIONI	98

CONCLUSIONI	99
BIBLIOGRAFIA E SITOGRAFIA	101

INTRODUZIONE

La *bioinformatica* rappresenta un campo interdisciplinare cruciale nell'era moderna, che integra la biologia con l'informatica per analizzare dati biologici complessi.

Uno degli ambiti più significativi di applicazione della bioinformatica è l'analisi delle reti biologiche, che verrà illustrato nel *primo capitolo*.

Queste reti sono essenziali per comprendere i modelli, le proprietà emergenti e i fenomeni biologici che sottostanno a processi vitali fondamentali.

Nel contesto della bioinformatica, l'attenzione si concentra sull'identificazione di interazioni molecolari, processi metabolici e regolatori che si manifestano come reti interconnesse di elementi biologici.

Nel secondo capitolo si approfondiscono le dinamiche di diffusione delle informazioni all'interno di reti complesse. Queste reti possono essere modellate come strutture di nodi interconnessi da link che rappresentano varie forme di interazione.

Un concetto chiave in questo contesto è quello di "*information cascade*", un processo attraverso il quale

un'informazione, un'idea o una malattia si diffonde attraverso la rete.

L'analisi delle reti complesse studia come la topologia e la centralità dei nodi influenzino la propagazione di fenomeni come l'informazione genetica, le opinioni nelle reti sociali o la diffusione di malattie infettive.

Nel terzo capitolo, si presenterà uno studio focalizzato sul monitoraggio delle *epidemie di Morbillo*, utilizzando approcci bioinformatici e di analisi delle reti complesse.

L'obiettivo principale sarà quello di esplorare l'impatto della copertura vaccinale e di altri fattori epidemiologici sulla diffusione di questa malattia altamente contagiosa. Particolare attenzione sarà data alla raccolta, all'analisi e all'interpretazione dei dati relativi alla sorveglianza epidemiologica, al fine di identificare pattern di diffusione e strategie di intervento efficaci.

Questa tesi si propone di integrare conoscenze teoriche e metodologiche provenienti dalla bioinformatica e dall'analisi delle reti complesse per affrontare questioni cruciali nel campo della salute pubblica. Attraverso un approccio interdisciplinare, si mira a fornire nuove

prospettive e strumenti per migliorare la comprensione e la gestione delle epidemie, con un focus specifico sul Morbillo come caso di studio.

CAPITOLO 1.

PROCESSI SU RETI

In questo capitolo si discuterà della bioinformatica e delle sue applicazioni. I processi su reti in bioinformatica si concentrano sull'analisi di tali reti per identificare modelli, proprietà emergenti e fenomeni biologici significativi. Illustrando inoltre, dei concetti chiave e delle applicazioni .

1.1 PROPRIETA' GENERALI

La *bioinformatica* è un campo interdisciplinare che unisce la biologia e l'informatica per comprendere e analizzare i dati biologici. Con l'avvento delle tecnologie genomiche e delle piattaforme di sequenziamento ad alto rendimento, la quantità di dati biologici generati è esplosa, richiedendo strumenti informatici avanzati per interpretare e analizzare questa vasta mole di informazioni.

Uno dei principali obiettivi della bioinformatica è quello di analizzare le sequenze genetiche, che sono le istruzioni fondamentali per la vita di un organismo.

Utilizzando algoritmi e tecniche di analisi computazionale, i bioinformatici possono identificare geni, predire le funzioni delle proteine codificate da questi geni e studiare le relazioni tra geni e malattie.

Inoltre, la bioinformatica è fondamentale per lo studio della genomica comparativa, che coinvolge il confronto e l'analisi delle sequenze genetiche di diverse specie. Questo approccio consente di identificare regioni genomiche conservate, che possono svolgere ruoli cruciali nell'evoluzione e nella funzione biologica.

Oltre all'analisi delle sequenze genetiche, la bioinformatica è coinvolta nella modellizzazione e simulazione di processi biologici complessi, come le interazioni proteina-proteina, le reti metaboliche e la struttura tridimensionale delle proteine.

Questi modelli computazionali forniscono preziose informazioni sulla funzione biologica e possono essere utilizzati per sviluppare nuovi farmaci e terapie. [1] [2]

Utilizza metodi di matematica, informatica, biologia, medicina e fisica allo scopo di migliorare la comprensione dei fenomeni biologici.

Di seguito alcuni *obiettivi* della *bioinformatica*:

- Sviluppo di strumenti per la generazione e il mantenimento dell'informazione proveniente dalle varie fonti: mappa fisica, mappa genetica, mappa cromosomica, mappa citogenetica, polimorfismi, e le informazioni relative alle sequenze genomiche e proteiche ecc.;
- Raccolta e organizzazioni delle informazioni genetiche associate alle patologie mediche;
- Sviluppo di programmi di calcolo per l'analisi delle sequenze;
- Sviluppo di Interfacce grafiche in grado di visualizzare in maniera efficace l'informazione richiesta;
- Sviluppo di strumenti informatici, includendo software, hardware e algoritmi per l'organizzazione e l'analisi dei dati;
- Realizzazione di standard per lo scambio e la descrizione dei dati;
- Realizzazione di una rete dati per la raccolta, la distribuzione e l'aggiornamento costante di tutta l'informazione prodotta;

- Raccolta della bibliografia, dei brevetti e di altri database di supporto all'informazione specifica;
- Predizione dei geni nelle sequenze di DNA;
- Predizione delle strutture tridimensionali delle proteine partendo dalle sequenze primarie;
- Predizione delle funzioni biologiche e biofisiche sia dalle sequenze e sia dalle strutture;
- Simulazione dei processi metabolici e cellulari basati su queste funzioni.

1.1.1 APPLICAZIONI DELLA BIOINFORMATICA

Negli ultimi anni la comunità scientifica ha rivolto una particolare attenzione allo studio delle popolazioni isolate per identificare geni responsabili di malattie complesse.

La scelta delle popolazioni isolate si basa sulla possibilità di semplificare il modello genetico alla base di tali malattie. Infatti, le caratteristiche di queste popolazioni quali il ridotto numero di fondatori e l'elevata consanguineità, causerebbero una riduzione del numero di geni implicati nello sviluppo di queste malattie. D'altra parte, l'uniformità dello stile di vita e dell'ambiente, permetterebbero di ridurre la variabilità dei fattori di rischio non genetici.

Recentemente sono stati sviluppati nuovi strumenti bioinformatici per l'identificazione e la ricerca nuovi geni e varianti geniche trovate nella popolazione e ritenute responsabili delle malattie complesse del sistema cardiovascolare e nervoso. [3] [4]

1.2 TEORIA DELLE RETI

Qualsiasi sistema reale, naturale o sociale, può essere rappresentato in questo modo: i rapporti interpersonali nella rete sociale, le interazioni tra molecole nella cellula, le reti ecologiche, internet, la rete neurale e molte altre.

Come affermano Newman, Barabási e Watts, i principali studiosi di questo campo, “le reti sono ovunque”.

Studiando i sistemi complessi da questa prospettiva, si può avere una maggiore comprensione della loro struttura, funzionamento e comportamento. La *teoria delle reti* si pone come superamento del riduzionismo, che in epoche precedenti aveva portato a scomporre i sistemi nelle loro componenti elementari, che venivano studiate singolarmente.

Una volta descritte tutte le parti, ci si aspettava, assemblandole, di poter comprendere il tutto: in realtà questo non è possibile, in quanto la complessità dei sistemi reali non è riconducibile alla semplice somma delle loro parti.

Si è capito, infatti, che la complessità deriva dal fatto che i costituenti dei sistemi reali interagiscono e si uniscono in modi molto diversi. Attraverso la prospettiva delle reti si vuole quindi riassemblare queste parti ricostruendo le mappe delle loro interazioni, in modo da comprendere i sistemi reali nella loro totalità. Per fare questo, dapprima si studia l'architettura di base dei sistemi, osservando solo i nodi e i link che formano la rete, per poi considerare le dinamiche che hanno luogo nei collegamenti, andando oltre la semplice struttura [5].

1.2.1 CENNI STORICI

Lo studio delle reti ha una lunga storia, con radici nella matematica, a partire dalla teoria dei grafi.

La teoria dei grafi è un ramo della matematica che studia i diversi modi in cui determinati oggetti si connettono.

Per **grafo** si intende un insieme di vertici, o nodi, connessi da archi, o link.

Si è soliti far risalire l'origine della teoria dei grafi al 1736, con *Leonard Eulero* e la sua risoluzione del cosiddetto problema dei sette ponti di Königsber. Eulero risolse il problema dimostrando matematicamente l'impossibilità di una sua soluzione, con quello che venne poi considerato da molti il primo teorema della teoria dei grafi: “su un grafo che abbia più di due nodi con un numero dispari di link un percorso che attraversi tutti i nodi una sola volta non può esistere”.

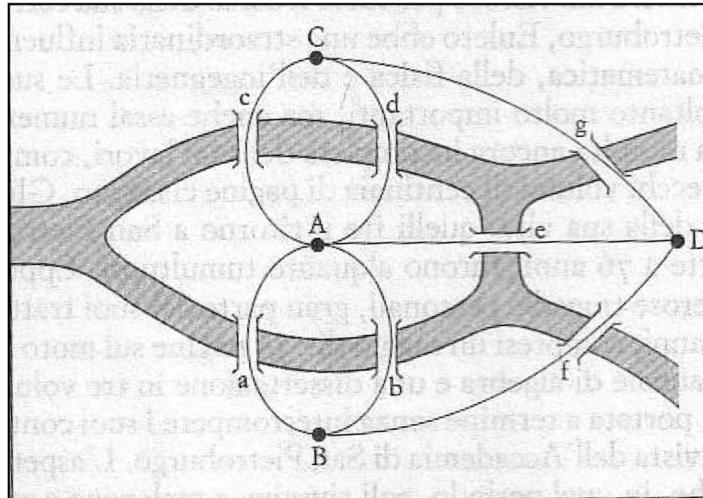


Fig.1 : Pianta schematica di Königsber [1]

Questa dimostrazione di Eulero è importante in quanto egli ha concepito i ponti come un grafo. Da questo momento si comprende, quindi, che la realtà può essere non solo rappresentata da grafi, o reti, ma anche essere spiegata attraverso le proprietà di questi grafi.

Le reti possono perciò essere usate come modello per la comprensione dei sistemi reali.

Dopo Eulero, la teoria dei grafi ha avuto un grande sviluppo, vertendo soprattutto sullo studio dei grafi regolari, cioè grafi in cui tutti i nodi hanno lo stesso numero di collegamenti, come ad esempio il reticolo degli atomi in un cristallo o quello delle cellette esagonali nei favi delle api.

Fino alla metà del XX secolo l'obiettivo principale di questa teoria è stato quello di catalogare le proprietà dei vari grafi. Solo a partire dagli anni '50 del secolo scorso, con gli studi di *Erdős* e *Rényi* sui grafi casuali, si passò dallo studio delle proprietà dei grafi a quello della loro nascita e formazione.

1.2.2 CONCETTI CHIAVE DELLE RETI

Una rete può esser rappresentata sotto forma di un grafo, ed infatti nello specifico:

- La rete è il *conceitto del mondo reale*.
- Un grafo è la *rappresentazione matematica* della rete.

Mediante le reti possiamo rappresentare con un formalismo quantitativo più entità collegate fra loro. Che permettono di modellare i “sistemi complessi”.

I grafi, invece, sono composti da oggetti semplici, detti NODI; e collegamenti tra vertici, detti archi o link;

I collegamenti possono essere **orientati** (l'arco ha un verso) e **non orientati**. E eventualmente dati associati a nodi e/o collegamenti.

Rappresentazione fondamentale e utile nelle reti complesse, facente parte delle proprietà generali delle reti è la **matrice di adiacenza**.

Essenzialmente, è una tabella rettangolare che rappresenta una rete attraverso le connessioni tra i suoi nodi.

Nella matrice di adiacenza:

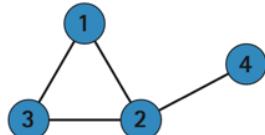
- Le righe e le colonne rappresentano i nodi della rete.
- Se esiste un collegamento (o un arco) diretto tra due nodi, l'elemento corrispondente nella matrice è solitamente rappresentato da un valore diverso da zero (a seconda del tipo di rete, può essere un 1, un valore di peso o altro).

- Se non esiste un collegamento tra due nodi, l'elemento nella matrice sarà 0 o un valore che indica l'assenza di collegamento.

Adjacency matrix

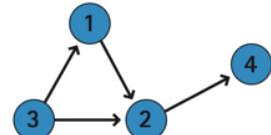
$$A_{ij} = \begin{pmatrix} A_{11} & A_{12} & A_{13} & A_{14} \\ A_{21} & A_{22} & A_{23} & A_{24} \\ A_{31} & A_{32} & A_{33} & A_{34} \\ A_{41} & A_{42} & A_{43} & A_{44} \end{pmatrix}$$

Undirected network



$$A_{ij} = \begin{pmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}$$

Directed network



$$A_{ij} = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}$$

Fig. 2 Rappresentazione della Matrice di Adiacenza

$A_{ij} = 1$ se esiste un collegamento che punta dal nodo j al nodo i;

$A_{ij} = 0$ se i nodi i e j non sono collegati fra loro.

Se il grafo è orientato avremo +1 o -1 a seconda dell'orientamento del grafo.

La matrice di adiacenza per grafi non orientati è simmetrica ovvero $A_{ij}=A_{ji}$.

La matrice di adiacenza è quindi uno strumento potente per visualizzare e analizzare la struttura delle reti, poiché fornisce un'indicazione chiara delle connessioni tra i nodi. Può essere utilizzata per calcolare varie proprietà delle reti, come il numero di archi, il grado di ciascun nodo, le distanze tra i nodi e molto altro.

Di seguito i concetti chiave della teoria delle reti :

- **Nodi e link:** I nodi rappresentano gli elementi del sistema, mentre i link rappresentano le relazioni tra i nodi. Ad esempio, nei social network, i nodi possono rappresentare individui, mentre i link rappresentano le connessioni di amicizia o le interazioni.
- **Grado:** Il grado di un nodo in una rete rappresenta il numero di link che sono collegati a quel nodo. Questa misura è importante per comprendere l'importanza o l'influenza di un nodo all'interno della rete. Il grado (detto anche ordine o valenza) di un nodo del grafo, che è uguale al numero di archi che incidono sul nodo stesso. Si può dire, che per ogni nodo definiamo il grado dei nodi come il numero dei nodi direttamente connessi al nodo in esame. Indichiamo con K_i il grado dell' i -esimo nodo della rete. In una rete non orientata il numero totale di collegamenti, L , può essere espresso come la somma

dei gradi del nodo. In base al grado che hanno si possono definire quelli più importanti rispetto ad altri. Nel caso di quelli orientati si considerano il numero di nodi uscenti ed entranti dal nodo. Infine si dice grado massimo di un grafo G , il grado massimo dei suoi nodi, e analogamente, grado minimo del grafo il grado minimo dei suoi nodi.

- **Centralità:** La centralità di un nodo misura l'importanza relativa di quel nodo all'interno della rete. Ci sono diverse metriche di centralità, come la centralità di grado, la centralità di vicinanza e la centralità di intermediazione, che forniscono diverse prospettive sull'importanza di un nodo.
- **Clustering:** Il clustering in una rete si riferisce alla tendenza dei nodi a formare gruppi o cluster fortemente interconnessi. Questa proprietà è importante per comprendere la struttura modulare delle reti e l'organizzazione dei loro sottogruppi.
- **Modelli di rete:** I modelli di rete sono rappresentazioni matematiche o computazionali che catturano le proprietà e le dinamiche delle reti reali. Questi modelli possono essere utilizzati per generare reti artificiali che riproducono determinate caratteristiche delle reti reali e per studiarne le proprietà emergenti.

Comprensione e applicazione della teoria delle reti sono ampiamente utilizzate in molti campi, tra cui la biologia, la sociologia,

l'informatica, l'economia e altro ancora. La teoria delle reti fornisce uno strumento potente per analizzare la complessità dei sistemi interconnessi e per estrarre conoscenze significative dalle loro strutture e dinamiche.

Infine, nel contesto delle reti, sia che si tratti di reti sociali, reti di trasporto, reti di comunicazione o reti biologiche, i concetti di percorsi e distanze giocano un ruolo centrale nell'analisi e nella comprensione della struttura e del funzionamento di tali sistemi.

Un path è una sequenza di nodi in modo tale che ogni nodo sia collegato al nodo successivo lungo il percorso di un arco.

Un percorso è sempre costituito da n nodi e $n - 1$ link. La lunghezza di un percorso è definita come il numero dei suoi collegamenti, contando più spigoli più volte.

La lunghezza media del percorso ($\langle d \rangle$) è la media dei percorsi più brevi tra tutte le coppie di nodi.

Diametro di una rete è la massima distanza fra due nodi, più è grande più è allungata la rete. Se il diametro è piccolo allora la rete sarà compatta.

Introducendo la “*Network Science*”. Si utilizzano concetti che abbracciano per la loro natura molte discipline dalla matematica con le matrici e il calcolo del laplaciano, l’informatica perché ha bisogno di elaboratori per essere simulata e rappresentata. È empirica in quanto derivata dai dati. I modelli non sono nulla se i dati non sono rilevati da osservazioni reali. Dunque, i modelli non sono nulla se non ci sono dati.

Mappe e dati sono indispensabili. Tuttavia, nessuna mappa è perfetta e ogni mappa che realizzeremo ammetterà degli errori di cui non si può fare a meno.

Diversi concetti chiave della scienza delle reti hanno le loro radici nella teoria dei grafi, un campo fertile della matematica. Quindi ciò che distingue la scienza delle reti dalla teoria dei grafi è la sua natura empirica, ovvero il suo focus su dati, funzione e utilità.

Ogni strumento che sviluppiamo è testato su dati reali e il suo valore è giudicato dalle informazioni che offre sulle proprietà e sul comportamento di un sistema.

Quantitativa e di natura matematica. Quando si fa una simulazione siamo sicuri del risultato perché segue leggi

matematiche e anche quantitative in quanto è possibile trovare sia il minimo che il massimo. È di natura computazionale [6].

1.3 MODELLI SU RETI

I modelli su reti sono strumenti potenti e versatili per comprendere la complessità dei sistemi basati su reti. Questi forniscono un quadro concettuale per esplorare le reti in vari contesti applicativi e costituiscono la base per una vasta gamma di studi e applicazioni interdisciplinari.

I modelli su reti consentono di studiare le proprietà strutturali delle reti, come la distribuzione dei gradi dei nodi, la centralità e la modularità. Inoltre, possono essere utilizzati per predire il comportamento futuro della rete e simulare l'effetto di cambiamenti o perturbazioni sulla rete.

I modelli dinamici consentono di studiare processi come la diffusione delle informazioni, la propagazione delle malattie e la dinamica delle reti sociali nel tempo.

Esistono diversi modelli su reti, utilizzati in varie discipline, i modelli più comuni sono riportati di seguito.

Il primo modello è quello di **Erdős-Rényi** (*Random Graph*). Questo è uno dei modelli più fondamentali per reti

casuali. Si basa su un insieme di n nodi, dove ciascuna coppia di nodi è connessa da un arco con una probabilità p . Utile per studiare proprietà statistiche delle reti casuali, come la soglia di connettività e la dimensione delle componenti connesse.

Esistono due varianti di questo modello:

1. Modello $G(n,p)$
2. Modello $G(n,m)$

Nel primo si inizia con n nodi isolati, per ogni coppia di nodi, si aggiunge un arco tra di essi con una probabilità p . La probabilità p controlla la densità della rete. Man mano che p aumenta, la rete diventa più densa.

Nel secondo , si inizia sempre con n nodi isolati, si aggiunge in arco tra coppie di nodi in modo casuale, finché il numero totale di archi raggiunge m .

La densità della rete è controllata dal numero totale di archi m .

Il secondo modello è **Barabasi-Albert**, questo modello risolve il problema che si ha nel modello random, poiché quest'ultimo non riproduce fedelmente le caratteristiche delle reti reali e in particolare non riproduce la presenza di

pochi nodi di grado elevato che compaiono in molti modelli di domini molto differenti. Con Barabasi si riesce a riprodurre i nodi di grado elevato, ovvero pochi nodi con un grado elevato (gli **HUB**).

È un modello *scale - free*, ovvero un modello di crescita di rete senza scala. [7]

Il funzionamento del modello Barabasi, si basa su 3 punti principali:

- 1) Inizializzazione, inizia con un piccolo numero di nodi collegati insieme;
- 2) Crescita progressiva, si aggiungono nuovi nodi uno alla volta;
- 3) Collegamento preferenziale, ogni nuovo nodo si collega a nodi esistenti con una probabilità proporzionale al loro grado attuale . Questo principio è noto come “*preferential attachment*”, in altre parole i nodi che hanno già un alto grado hanno una maggiore probabilità di ricevere nuovi collegamenti rispetto ai nodi con un grado più basso.

L’attaccamento preferenziale può essere *lineare o non lineare*:

- Lineare, significa che si ha una crescita in cui gli hub sono introdotti in maniera delicata e quindi lineare;
- Non lineare, significa che esisteranno dei super hub che andranno a dominare tutto il resto;

Questo processo continuerà fin quando la rete raggiunge la dimensione desiderata. Questo modello riflette alcune caratteristiche delle reti reali osservate in vari contesti, come la World Wide Web, le reti sociali e molte altre.

Poi il **Stochastic Block Model (SBM)**, è un modello probabilistico per generare reti complesse che include la struttura a blocchi o comunità. In questo modello la probabilità di connessioni tra nodi dipendono dai blocchi a cui appartengono.

Ed infine, i **Modelli di Diffusione e Contagio**. I modelli computazionali di diffusione e contagio simulano la propagazione di informazioni, opinioni, malattie o comportamenti attraverso le reti. Ad esempio, i modelli SIR (*Susceptible-Infected-Recovered*) vengono utilizzati per simulare l'epidemia di malattie infettive su reti sociali. La modellazione computazionale comporta l'uso di algoritmi e

software per descrivere e simulare sistemi naturali sfruttando la matematica e l'informatica. L'uso di modelli computazionali consente lo studio della dinamica di tali sistemi attraverso l'osservazione di variabili matematiche che riassumono le loro caratteristiche. La simulazione è uno strumento prezioso per rappresentare l'evoluzione dei sistemi dinamici. L'uso di modelli computazionali efficaci è una strategia chiave per progettare e implementare interventi a tutti i livelli con l'obiettivo di bloccare la diffusione delle malattie, salvando così vite umane e riducendo lo stress sul sistema sanitario.

Modelli deterministici sono modelli semplicisti per modellare ad esempio una epidemia anche se dipende da malattia, da tipo di contagio, da come si diffonde la malattia. I primi modelli nascono prima della verifica sul campo del covid, come il *modello di diffusione della Peste* del 1300.

Citando i *modelli compartmentali* (come il nome stesso suggerisce) sono modelli a compartimenti cioè la divisione delle persone in compartimenti etichettati con il loro stato di malattia, cioè suscettibili, infetti o recuperati, ad esempio, basati su equazioni differenziali ordinarie. Quindi, il

processo di diffusione della malattia può essere facilmente modellato utilizzando una struttura di dati grafici, i cui nodi rappresentano gli individui e i bordi dei loro contatti.

I modelli semplici non tengono conto di altre informazioni.

I modelli raffinati integrano nei nodi informazioni aggiuntive, come le caratteristiche degli individui e nei bordi informazioni relative ai contatti (ad esempio il tempo di esposizione, l'ambiente in cui si svolgono i contatti, ecc.). In particolare, le informazioni relative al tempo vengono utilizzate per costruire reti temporali in cui i bordi memorizzano la durata o i tempi del contatto. È quindi possibile indagare alcune politiche ad hoc di contenimento (ad esempio l'uso di mascherine, lockdown localizzato o strategie di vaccinazione su misura) con questi modelli. [8]

1.3.1 MODELLI COMPARTIMENTALI

Le caratteristiche principali dei modelli compartmentali suddividono la popolazione in diverse "compartimenti" o "stati" in base al loro stato rispetto alla malattia. I compartimenti più comuni includono:

- **Susceptible (S):** Individui che sono suscettibili all'infezione e possono contrarre la malattia.

- **Infected (I)**: Individui che sono infetti e possono trasmettere la malattia ad altri.
- **Recovered (R)**: Individui che sono guariti dalla malattia e sono immunizzati contro di essa.

A seconda della complessità del modello, possono essere inclusi altri compartimenti come individui esposti (Exposed), individui in quarantena, individui portatori asintomatici, ecc.

Alcuni esempi di modelli compartmentali sono:

- a) **Modello SIS** : Ha due compartimenti, Suscettibile e Infetto. Un individuo può essere suscettibile poi infetto e successivamente suscettibile in quanto non diventa immune una volta contratta la prima volta la malattia come per esempio il raffreddore.

The SIS model

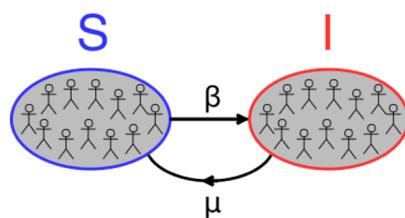


Fig.3 : SIS Model

β è la probabilità con cui uno suscettibile diventa infetto mentre μ è il contrario ovvero che da infetto diventa di nuovo suscettibile.

Modello che si applica a tutte le malattie che non conferiscono immunità.

Si definisce un *modello di contatto*, dove i nodi hanno una **label** (etichetta) S o I.

b) Modello SIR : In questo modello abbiamo probabilità β che uno diventa infetto e con probabilità μ che da infetto diventi recovery ovvero che non si può più infettare. In altri modelli più complessi non si passa direttamente da suscettibile a infetto ma c'è un altro stato chiamato exposed, ovvero uno suscettibile può essere exposed e può o sviluppare la malattia perché ha avuto un contatto o ritornare ad essere suscettibile.

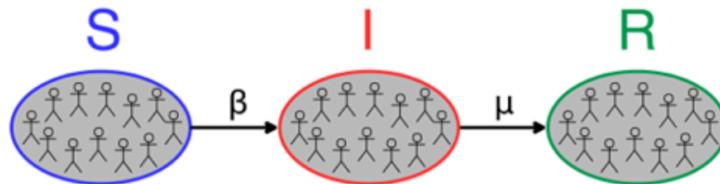


Fig.4 : SIR Model

I modelli compartmentali possono essere implementati utilizzando simulazioni computazionali o equazioni differenziali per descrivere il cambiamento nel tempo della distribuzione dei compartimenti all'interno della popolazione. Le implementazioni computazionali

consentono di studiare il diffondersi della malattia al variare di diversi parametri come il tasso di contatto, il tasso di guarigione, il tasso di mortalità, ecc.

Le simulazioni computazionali dei modelli compartmentali possono essere eseguite utilizzando linguaggi di programmazione come Python, MATLAB, R, o tramite l'uso di software specifici per l'epidemiologia computazionale. Queste simulazioni consentono agli epidemiologi e agli studiosi di testare strategie di controllo delle malattie e di prevedere il possibile impatto di interventi come la vaccinazione o il distanziamento sociale.

CAPITOLO 2.

DIFFUSIONE DELL'INFORMAZIONE

In questo capitolo si descrivono le dinamiche di diffusione delle informazioni all'interno di reti complesse. In questo contesto, una "information cascade" indica il processo attraverso il quale un'informazione, un'idea, una malattia o altro si diffondono da un nodo o da un insieme di nodi all'interno di una rete complessa. Le reti complesse possono essere costituite da nodi (o individui) e dai collegamenti (o relazioni) tra di essi, che possono essere rappresentate da reti sociali, reti di comunicazione, reti di trasmissione di malattie, ecc. L'analisi in reti complesse mira a comprendere come le caratteristiche della rete, come la topologia e la centralità dei nodi, influenzino la diffusione delle informazioni.

2.1 DINAMICA SU RETI

I modelli basati su reti rappresentano uno strumento fondamentale per analizzare tali fenomeni complessi. Essi si differenziano significativamente dai modelli più tradizionali, consentendo di esplorare ipotesi e dinamiche

altrimenti difficilmente accessibili con altri approcci. Una premessa chiave di molti di questi modelli è l'ipotesi del "*mixing perfetto*", che postula una distribuzione uniforme delle interazioni tra individui, indipendentemente dalla distanza fisica.

Nell'ambito della diffusione di virus, il concetto di mixing perfetto implica che ogni individuo ha la stessa probabilità di trasmettere un'infezione simultaneamente a un vicino rispetto a una persona molto distante. Tuttavia, tali modelli non tengono conto delle complesse variazioni nei contatti reali che si verificano tra le persone.

La diffusione dell'informazione, anche nota come trasmissione o cascata dell'informazione, costituisce un argomento di notevole interesse all'interno della comunità scientifica delle reti, data la sua rilevanza e le potenziali applicazioni interdisciplinari. Un esempio di questo fenomeno è rappresentato dalla strategia di marketing virale, in cui un'azienda offre gratuitamente un determinato prodotto a individui influenti, con l'obiettivo che essi possano consigliarlo ai propri contatti in caso di soddisfazione [9].

L'immunizzazione target è un altro esempio in cui un piccolo sottoinsieme di nodi influenti viene selezionato e immunizzato [10] [11].

2.2 MODELLI A CASCATA

I primi modelli introdotti sono i Modelli a Soglia, *Threshold Models*. Si basano sull'idea che gli individui abbiano una soglia personale che deve essere superata affinché essi adottino un certo comportamento o opinione.

In questo contesto, gli individui sono rappresentati come nodi in una rete, e i legami tra di loro indicano la possibilità di influenzarsi reciprocamente. Ogni individuo ha un livello di soglia personale che rappresenta la quantità di adozioni di comportamento o opinione necessaria tra i suoi contatti per convincerlo ad adottare lo stesso comportamento.

I modelli a soglia possono essere applicati a vari contesti, inclusi quelli legati alla diffusione di nuove tecnologie, prodotti, idee, o comportamenti sociali. Alcuni concetti chiave associati ai *threshold models* includono:

- **Soglia di adozione:** Rappresenta il numero minimo di adozioni che un individuo deve osservare tra i suoi

contatti per essere influenzato e adottare a sua volta il comportamento.

- **Dinamica di diffusione:** I modelli a soglia descrivono come l'adozione si diffonda attraverso la rete in base alle soglie individuali e ai legami tra gli individui.
- **Processo di contagio:** In questo contesto, il "contagio" si riferisce alla diffusione del comportamento o dell'opinione attraverso la rete in base alle interazioni sociali e alle soglie di adozione.

Uno dei modelli a soglia più noti è il modello di cascata di adozione (*cascade model*), che descrive la diffusione di informazioni o comportamenti attraverso una serie di passaggi successivi, dove l'adozione di un individuo può innescare l'adozione di altri nelle fasi successive.

La costruzione di questi modelli inizia con la definizione di una rete di nodi interconnessi, dove ogni nodo rappresenta un individuo e i collegamenti tra i nodi rappresentano le relazioni o le influenze sociali tra di loro. Inizialmente, nessun nodo è attivo, il che significa che nessun individuo ha adottato il comportamento o l'opinione iniziale. Successivamente, vengono selezionati in modo casuale alcuni nodi per diventare attivi o adottare il comportamento desiderato. Questi nodi inizialmente attivi possono essere

considerati come "seminali" che iniziano il processo di diffusione.

La simulazione della diffusione avviene attraverso l'applicazione di regole di propagazione del comportamento o dell'opinione attraverso la rete. Questa propagazione può avvenire in modo sincrono o asincrono. Nel caso sincrono, tutti i nodi nella rete aggiornano il loro stato contemporaneamente in intervalli di tempo discreti, mentre nel caso asincrono, i nodi si aggiornano in tempi diversi in base a specifiche regole.

Un concetto chiave introdotto successivamente è quello della soglia frazionaria.

In questo contesto, la norma di adozione è basata sul cambiamento nell'ambiente sociale locale di un individuo. Ad esempio, un individuo può decidere di adottare un comportamento se una percentuale specifica dei suoi vicini immediati (connessi nella rete) ha già adottato lo stesso comportamento. Questo tipo di norma consente di mantenere una soglia unica, cioè uguale per tutti i nodi della rete, ma al contempo agisce in modo locale, considerando solo le interazioni dirette tra i vicini di ciascun individuo.

Dunque, i modelli predittivi a cascata possono essere generalmente divisi in due classi: modelli basati su grafi e modelli non basati su grafi. I modelli non basati su grafi sono *modelli epidemici*, come precedentemente analizzato (Capitolo 1) che studiano matematicamente la diffusione utilizzando dinamiche basate sulla popolazione [12] [13]. Modelli ben noti in questo contesto sono il modello suscettibile-infetto-rimosso (SIR) [14]. In questi metodi, spesso si desidera studiare l'evoluzione nel tempo del rapporto tra nodi suscettibili, infetti e rimossi. Questi modelli epidemici sono stati frequentemente utilizzati per studiare come un virus si diffonde all'interno di una popolazione e anche su altri processi di contagio come la diffusione dell'innovazione, la diffusione delle informazioni, la diffusione dei rumori e la diffusione dei movimenti politici.

Nei modelli di previsione a cascata che si basano sull'analisi dei grafi, si esamina spesso la dinamica a livello individuale dei nodi. In queste situazioni, un certo tipo di informazione parte da un gruppo iniziale di nodi e si diffonde attraverso la rete seguendo un processo a cascata. Due dei modelli più noti che si basano su questo approccio sono il **modello a**

soglia lineare e il modello a cascata indipendente [15] [16].

Il modello di soglia lineare è stato sviluppato per descrivere il processo decisionale binario ed è stato ampiamente impiegato nell'ambito dell'economia e della sociologia per modellare questo tipo di processo. Questo modello si propone di simulare comportamenti simili al fenomeno di "seguire la massa", dove un individuo prende una decisione basandosi sulle azioni dei suoi vicini più immediati .

Nel contesto di questo modello, ogni nodo di una rete può assumere diversi stati. Generalmente, i nodi possono essere attivi (ovvero influenzanti) o silenziosi (non influenti). All'inizio del processo, tutti i nodi sono considerati in modalità silenziosa. Successivamente, alcuni nodi vengono selezionati come "*early adopters*" (precursori), per i quali lo stato viene cambiato in attivo. Questi precursori possono a loro volta influenzare i loro vicini, a patto che si verifichi una condizione di soglia. In generale, un nodo silenzioso diventa attivo se almeno una certa frazione dei suoi vicini è già attiva. Questa soglia può essere uniforme per tutti i nodi (come avviene di solito) oppure può essere assegnata in base a una distribuzione specifica, definendo soglie diverse per ciascun nodo. Il processo di diffusione iterativa continua finché non si raggiunge uno stato stazionario, cioè

quando non si verificano più cambiamenti nello stato dei nodi e nessun altro nodo viene attivato. In determinate condizioni, questo processo può portare a una cascata completa nella rete, dove tutti i nodi diventano attivi. Il modello di soglia lineare rappresenta un approccio semplice ma efficace, ampiamente utilizzato in numerosi studi di diffusione e contagio all'interno di reti complesse [17].

Mentre, nel modello a cascata indipendente, come il precedente, i nodi possono essere attivi o silenziosi [18]. Spesso, si fanno due assunzioni principali per il modello a cascata indipendente:

- Ogni nodo attivo **i** ha una sola possibilità di attivare il suo vicino silenzioso **j** e, se il processo di attivazione non riesce, non ci sarà nessuna influenza del nodo **i** sul nodo **j**;
- La probabilità che un nodo silenzioso **j** venga influenzato da un nodo attivo **i** (che si trova nel suo vicinato) è indipendente dall'influenza degli altri nodi attivi sul nodo **j**.

In questo modello, l'attivazione di un nodo attivo può propagarsi a un nodo silenzioso solo una volta, e il risultato dell'attivazione è determinato dalla probabilità di influenza specifica tra quei due nodi, indipendentemente dalle altre

attivazioni che possono avvenire contemporaneamente nella rete.

Volendo fare un sviluppo di questo modello, si svolge nel seguente modo. Inizialmente, tutti i nodi della rete vengono considerati silenziosi. Successivamente, una piccola percentuale di nodi diventa attiva. Denotiamo i nodi che diventano attivi in uno specifico passaggio del processo con $AC(t)$.

Ad ogni passaggio dell'algoritmo, ciascun nodo attivo nell'iterazione precedente ($AC(t - 1)$) attiva uno dei suoi nodi silenziosi con una certa probabilità, rappresentata da Q_{ij} , che indica la probabilità che un nodo attivo i attivi un vicino silenzioso j. Questa probabilità, specifica per ciascun collegamento, è anche nota come probabilità di diffusione nella letteratura. Tali valori di probabilità devono essere definiti prima di avviare la simulazione numerica. Se le probabilità di diffusione non sono conosciute, è possibile utilizzare un metodo di massima verosimiglianza per stimarle, basandosi sulle osservazioni delle sequenze di cascata. Nel caso in cui un nodo silenzioso j abbia più di un vicino attivato nel passaggio precedente, i vicini attivi tenderanno ad influenzare il nodo j in un ordine arbitrario. Questo procedimento continua fino a raggiungere uno stato

stazionario, dove non ci sono ulteriori cambiamenti nello stato dei nodi [19].

Infine, il numero o la proporzione di nodi attivati partendo dall'attivazione iniziale di uno o più nodi viene definito come la sua influenza di diffusione o raggio di influenza. Un nodo con un'alta influenza di diffusione è cruciale per la diffusione delle informazioni nella rete.

Questa illustrazione schematica (Fig.5) rappresenta una cascata di informazioni in una rete composta da 15 nodi e 17 collegamenti, dove ogni collegamento ha una probabilità uniforme di diffusione dell'informazione del 50% .

Il processo di diffusione inizia attivando il nodo 1 (a).

Nel secondo passaggio (b), uno dei vicini del nodo 1 (il nodo 3) viene attivato.

Nel terzo passaggio (c), il nodo 3, attivato nel passaggio precedente, a sua volta attiva due dei suoi vicini, i nodi 4 e 5. Il nodo 4 non ha vicini silenziosi da attivare, ma il nodo 5 attiva il nodo 9 nel quarto passaggio (d), il quale successivamente attiva uno dei suoi vicini nel passaggio seguente (e).

Infine, poiché nessun nodo viene attivato nel sesto passaggio, il processo si arresta, risultando nell'attivazione

totale di 5 nodi partendo dall'attivazione iniziale del nodo 1 (f).

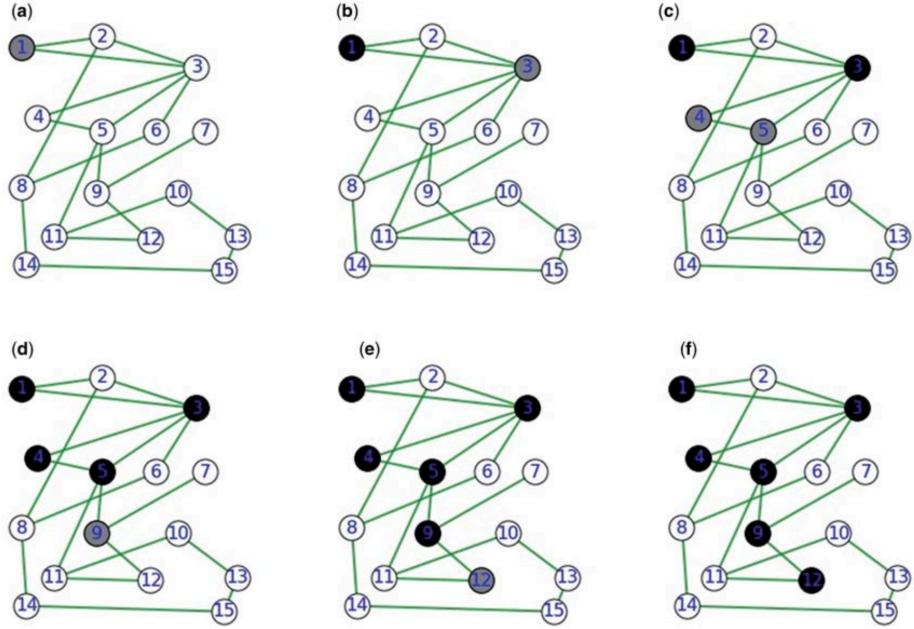


Fig. 5 - L'illustrazione mostra graficamente come un'informazione si diffonda all'interno di questa rete specifica, dove ogni connessione ha la stessa probabilità del 50% di trasmettere l'informazione al suo vicino.

2.3 CENTRALITA' DEI NODI

La centralità di un nodo è una misura che indica l'importanza o l'influenza di un nodo all'interno di una rete. Questa importanza dipende dalla posizione strutturale del nodo nella rete e dalla sua capacità di influenzare il flusso di informazioni o di connessioni [20]. Le misure di centralità dei nodi sono utilizzate per identificare i nodi più

importanti o centrali in una rete, e possono essere calcolate in vari modi, tenendo conto di diversi aspetti della struttura della rete.

Queste misure sono fondamentali per analizzare le proprietà dei sistemi complessi e per comprendere il ruolo specifico dei nodi all'interno di tali sistemi.

Alcuni esempi di misure di centralità dei nodi includono:

- ▶ Centralità di grado (**degree centrality**): misura il numero di connessioni di un nodo con gli altri nodi della rete.
- ▶ Centralità di vicinanza (**closeness centrality**): misura la distanza media di un nodo da tutti gli altri nodi nella rete.
- ▶ Centralità di intermediarietà (**betweenness centrality**): misura il numero di volte che un nodo si trova lungo il percorso più breve tra tutte le coppie di nodi nella rete.

Riprendendo la centralità del grado con la definizione di nodo, si riferisce al grado e si indica dal punto di vista matematico con il concetto di centralità ovvero quanto un nodo è rilevante.

Molto importante è che influenza e centralità non sono la medesima cosa. Dunque, per centralità si intende il grado del nodo e più esso è alto, più è centrale (*degree centrality*); Per comparare due reti si adotta un valore di degree centrality normalizzato cioè il grado del nodo diviso un valore standard che può essere il numero massimo dei nodi o il numero dei nodi effettivi. Abbiamo così un valore tra 0 e 1. La centralità è quindi definita come il grado dei nodi normalizzato.

Un'altra misura della centralità da analizzare è un nodo che ha molti cammini minimi che passano da esso indica che il nodo è un nodo centrale (*betweenness centrality*).

In altre parole è il numero di cammini minimi che passano da un nodo;

Spesso *Betweenness* può essere collegata con *closeness* e *degree* ma non sempre.

Le misure di centralità dei nodi svolgono un ruolo significativo nello studio delle proprietà dei sistemi reali. Ad esempio, Mantzaris et al. hanno utilizzato misure dinamiche di centralità per caratterizzare l'attività aggregata in diverse regioni cerebrali, così come per riprodurre

informazioni legate all'apprendimento nel cervello [21]. Considerando una rete non diretta e non pesata $\mathbf{G(VE)}$, dove V è l'insieme dei nodi e E è l'insieme degli archi. La rete è completamente descritta dalla sua matrice di adiacenza $A = [a_{ij}]$, dove $a_{ij} = 1$ quando c'è un collegamento tra i nodi i e j , e $a_{ij} = 0$ altrimenti. Inoltre, si denota, il collegamento tra i nodi i e j con e_{ij} . La misura di centralità più semplice di un nodo è il suo grado (*in- e out-degree per reti dirette*). Come definito (Capitolo1), il grado di un nodo è il numero totale di collegamenti che il nodo ha con gli altri. Il grado \mathbf{k}_i del nodo i è calcolato come:

$$k_i = \sum_j a_{ij}.$$

Il grado è una misura di centralità importante, e molti processi dinamici sono stati correlati al grado dei nodi [22]. Il grado tiene conto solo del numero di vicini, tuttavia a volte l'importanza di un particolare vicino può essere rilevante.

Ad esempio, connettersi con un individuo importante può conferire maggiore importanza (o prestigio) a un nodo rispetto a essere connesso con più vicini non importanti. Per tener conto di ciò, l'indice di Hirsch (noto come *H-index*) è stato esteso per la centralità dei nodi. L'H-index, proposto

da Hirsch come un indice per quantificare l'impatto della ricerca di un individuo [23], è il numero di pubblicazioni di un individuo che ha ricevuto almeno h citazioni.

Lo stesso concetto è stato esteso alle reti [24][25].

Si definisce un **operatore H** che, applicato a un insieme di variabili reali $[y_1, \dots, y_n]$, restituisce il massimo intero h tale che tra i membri di questo insieme ci siano almeno h membri con un valore non inferiore a h .

$$h_i = H(k_j), j \in N_i.$$

Come precedentemente detto, la **centralità betweenness**, tiene conto della centralità dei nodi nella navigazione attraverso la rete. La centralità B_i del nodo i in un grafico, che mostra il numero di percorsi più brevi che utilizzano il nodo i (eccetto quelli tra il nodo i -esimo con gli altri nodi), è calcolata come:

$$B_i = \sum_{j \neq i \neq k} \frac{\Gamma_{jk}(i)}{\Gamma_{jk}},$$

dove T_{jk} è il numero di cammini minimi tra i nodi j e k e $\Gamma_{jk}(i)$ è il numero di questi cammini minimi che fanno uso del nodo i .

Altre misura importante è la **Closeness**, indicante la distanza media che c'è da uno nodo ad altri nodi. Calcolata come:

$$C_i = \frac{1}{N-1} \sum_j \frac{1}{p_{ij}}.$$

Questo spiega il perchè i nodi che sono più centrali nella rete, cioè quelli che sono più vicini a molti altri nodi, potrebbero avere un migliore accesso alle informazioni e quindi una maggiore influenza sugli altri.

Per fare questo, si valuta quanto un nodo contribuisce alla quantità complessiva di informazioni che possono passare attraverso tutti i possibili percorsi nella rete. Più un nodo è importante per il trasferimento di informazioni tra i nodi della rete, più alto sarà il suo indice informativo. Questo approccio differisce dalla centralità della vicinanza, che considera solo la distanza fisica dei nodi senza valutare il loro ruolo nel trasferimento di informazioni.

2.4 ATTORI CHIAVE NELLA DIFFUSIONE

In una rete, alcuni nodi hanno una maggiore capacità di diffondere le informazioni rispetto ad altri perché occupano

posizioni strategiche. Molte ricerche sono state condotte per comprendere meglio questo fenomeno e individuare i nodi più influenti, anche chiamati "*super diffusori*", nelle reti complesse. [26][27]

Il problema della massimizzazione dell'influenza, introdotto per la prima volta da Kempe e colleghi, si concentra su come trovare un insieme di nodi che massimizzi la diffusione delle informazioni nella rete.[27]

La ricerca di soluzioni ottimali è spesso complessa e richiede l'impiego di algoritmi sofisticati. Inoltre, le tradizionali misure di centralità, come il grado e il betweenness, non sempre identificano i nodi più influenti.

Infatti, la posizione strategica di un nodo nella rete, risulta essere più significativa nel valutare la sua influenza sulla diffusione delle informazioni. Questo concetto è stato approfondito attraverso l'utilizzo di algoritmi di decomposizione *k-sell* per identificare i nodi chiave nella diffusione all'interno della rete. [28][29]

Gli algoritmi *k-sell* sono utilizzati per identificare i nodi chiave o influenti all'interno di una rete. Questi algoritmi si basano sulla decomposizione della rete in componenti strutturali, chiamate “*k-sell*”.

In sostanza, un k-sell è una parte della rete che, una volta rimossa, porta alla frammentazione della rete in k o più componenti connesse. Gli algoritmi k-sell cercano di identificare le parti della rete che, se rimosse, possono influenzare significativamente la sua struttura o la diffusione delle informazioni. L'idea alla base degli algoritmi k-sell è che i nodi all'interno di queste componenti sono cruciali per la connettività o la diffusione all'interno della rete. Pertanto, identificare e analizzare questi nodi può fornire preziose informazioni sulla struttura e sul funzionamento della rete. Questi algoritmi possono variare nella loro complessità e nelle strategie utilizzate per identificare queste componenti critiche nella rete. Tuttavia, in generale, mirano a individuare i nodi o i gruppi di nodi il cui ruolo è cruciale per la coesione o l'efficienza della rete stessa.

Seguendo un procedimento ben definito. Si definisce la "*coreness*" di una rete , è una misura che indica il livello di importanza o centralità di ciascun nodo all'interno della rete stessa.

Per calcolare la coreness di una rete, si segue il seguente procedimento:

1. Inizialmente, si assegna una coreness di zero a tutti i nodi isolati della rete, ovvero quei nodi che non sono collegati ad altri nodi. Questi nodi vengono poi rimossi dalla rete.
2. Successivamente, si applica l'algoritmo di decomposizione k-sell alla rete. In questo algoritmo, vengono rimossi tutti i nodi che hanno un grado (ovvero il numero di archi che collegano un nodo ad altri nodi) pari a 1. La rimozione di questi nodi riduce il numero totale di archi nella rete.
3. I nodi rimossi durante questo passaggio vengono assegnati una coreness di 1.
4. Si identificano quindi i nodi con un grado inferiore o uguale a 1 nel grafo rimanente e li si rimuove ulteriormente dalla rete. Questi nodi ricevono una coreness di 2.
5. Il processo continua fino a quando si identificano i nodi con la coreness massima.

La coreness può essere calcolata in modo efficiente per reti di grandi dimensioni. [30]

Tuttavia, non può essere applicata a molti tipi di reti modello, come ad esempio le reti ad albero e le reti scale-

free costruite utilizzando l'algoritmo di Barabasi-Albert. Questo perché in queste reti la coreness di tutti i nodi ha valori simili, rendendo difficile distinguere i nodi più importanti o centrali. Sono stati proposti approcci alternativi, come il *mixed degree* e il *generalized degree discount*, per risolvere questo problema [31] [32]

2.5 CENTRO E PERIFERIA: GUIDA ALLA DIFFUSIONE DELLE INFORMAZIONI

Numerose ricerche hanno evidenziato che la configurazione di una rete influisce notevolmente sulla sua capacità di diffondere informazioni. Ciò tuttavia dipende anche dal processo dinamico oggetto di studio.

Ad esempio, un nodo potrebbe essere un diffusore chiave nella dinamica dell'informazione, ma non in quella delle epidemie. Nonostante ciò, alcune misure generali di centralità si sono dimostrate piuttosto vitali per molte dinamiche di diffusione [33] De Arruda et al. [34] hanno studiato come le diverse misure di centralità siano correlate con le loro capacità di diffusione dell'epidemia. Hanno scoperto che il grado e il k-core mostrano la più alta correlazione per la diffusione dell'epidemia nelle reti non spaziali, mentre il grado di vicinanza e il grado medio di

vicinato sono quelli più legati alla dinamica delle voci. L'identificazione dei nodi con un ruolo centrale nella diffusione delle informazioni avrà molte potenziali applicazioni in vari campi. Ad esempio, determinando gli *spreader* più influenti nei gruppi criminali, l'organismo investigativo competente può gestire meglio la criminalità e attuare strategie preventive [35].

2.5.1 SIMULAZIONI RETI MODELLO & REALI

In questa sezione vengono presentati i risultati delle simulazioni numeriche condotte su una serie di reti sintetiche e reali. Durante le simulazioni, viene considerata una probabilità uniforme di diffusione su tutti i collegamenti, espressa come $Q_{ij} = Q$. Per un dato valore di Q , per ogni nodo i otteniamo il suo raggio di influenza nella diffusione, S_i , che rappresenta il rapporto tra i nodi attivati dalla attivazione iniziale del nodo i . Valori più elevati di S_i indicano una maggiore capacità di diffusione per il nodo i . Successivamente, questi valori di influenza nella diffusione vengono correlati con i valori di centralità.

Vengono considerate reti modello, tra cui reti scale-free con preferenziale attaccamento, reti casuali e reti small-world. Le reti casuali sono costruite secondo il modello proposto

da Erdős e Renyi. Nelle reti Erdős-Rényi, c'è un collegamento tra ogni coppia di nodi con una certa probabilità P .

Le *reti small-world* sono costruite secondo l'algoritmo proposto da Watts e Strogatz nel loro lavoro fondamentale. Per costruire le reti Watts-Strogatz, si considera innanzitutto un grafo ad anello con N nodi, ognuno connesso ai suoi m vicini più vicini. Successivamente, i collegamenti vengono ricollegati uno per uno. Per diversi valori della probabilità di ricollegamento, si ottengono reti che vanno da regolari (per $P = 0$) a disordine completo (per $P = 1$). Le reti **Erdős-Rényi** e **Watts-Strogatz** hanno una distribuzione dei gradi quasi omogenea. Tuttavia, molti network reali mostrano gradi nodali eterogenei. Il modello di Barabási-Albert genera reti di questo tipo. Il modello è il seguente:

Innanzitutto, si considera una rete connessa “tutti con tutti” con m nodi.

Successivamente, ad ogni passaggio successivo, viene aggiunto un nuovo nodo alla rete e crea m collegamenti con i vecchi nodi. La probabilità di collegare il nuovo nodo a un vecchio nodo è proporzionale al grado del vecchio nodo; maggiore è il grado di un vecchio nodo, maggiore è la probabilità di connettersi al nuovo nodo.

Anche se lo studio delle reti modello può fornire informazioni utili su come si comportano i sistemi di reti reali, i modelli non riescono tuttavia a imitare tutte le proprietà delle reti reali. Pertanto, si considerano anche diverse reti reali e si studiano le relazioni tra i valori di centralità dei loro nodi e l'influenza nella diffusione.

Le reti reali prese in considerazione in questo lavoro includono la rete di co-autori di scienziati attivi nel campo della Scienza delle Reti (Net Sci), reti sociali tra dottorandi in Informatica (CS PhD), simili a Facebook, una rete del Dizionario Online di Biblioteconomia e Scienza dell'Informazione (ODLIS), gli aeroporti degli Stati Uniti, Internet al livello di Sistema Autonomo (AS), la rete elettrica degli Stati Uniti e quella dell'Unione Europea. [36]

2.5.2 ANALISI DI STUDI PRECEDENTI

Si mostra come vengono esaminate le relazioni tra l'influenza nella diffusione dei nodi e i loro livelli di centralità nelle reti scale-free, caratterizzate da un grado medio e una eterogeneità variabile.

I risultati sono basati su medie calcolate da 20 esecuzioni indipendenti dalle simulazioni. Si utilizzano i coefficienti di

correlazione di *Spearman* (misura statistica non parametrica utilizzata per valutare la forza e la direzione della relazione tra due variabili ordinali o classificate) per valutare queste relazioni, includendo misure di centralità come grado, betweenness, vicinanza, eccentricità, indice H, indice di informazione e centralità degli autovalori.

Una *correlazione positiva* indica che, all'aumentare dei livelli di centralità di un nodo, aumenta anche la sua influenza nella diffusione, implicando che attivando questi nodi inizialmente si attivino anche un maggior numero di altri nodi. Al contrario, una *correlazione negativa* suggerisce il contrario, cioè che un aumento nei livelli di centralità sia associato a una minore influenza nella diffusione. [36]

Si osserva che, man mano che le reti diventano più dense, i valori assoluti della correlazione tra l'influenza nella diffusione e i valori di centralità diminuiscono (Figura 6).

Mentre grado, betweenness, vicinanza, indice H e centralità degli autovalori sono positivamente correlati con l'influenza nella diffusione dei nodi, l'indice di informazione e l'eccentricità mostrano valori di correlazione negativi. L'indice di informazione ha la più alta correlazione negativa, mentre il grado ha spesso la correlazione positiva più significativa.

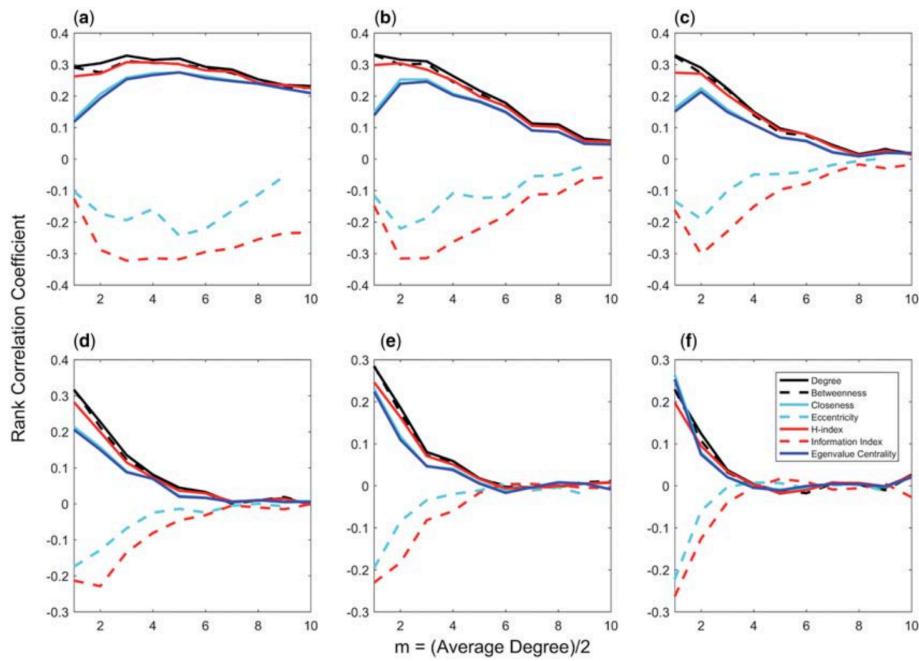


Fig. 6 : Mostra i valori di correlazione di Spearman tra la centralità dei nodi e la loro influenza nella diffusione, in funzione del grado medio ($2m$) nelle reti scale-free con dimensione $N = 500$ e $B = 0$ (vedere il testo per la spiegazione di questo parametro). La soglia per il modello a cascata indipendente è fissata rispettivamente a 0,1, 0,2, 0,3, 0,4, 0,5 e 0,6 per (a)-(f).

Ciò indica che scegliere i nodi con il grado più alto probabilmente porta all'attivazione di un insieme più ampio di nodi silenziosi, mentre scegliere quelli con l'indice di informazione più alto comporterà il minor numero di nodi attivati.

Il profilo dei modelli di correlazione è quasi indipendente dal livello di eterogeneità della rete, espresso da B (cioè maggiore è il B , minore è il livello di eterogeneità), e le

correlazioni svaniscono per grandi valori del parametro di soglia Q (Figura 7).

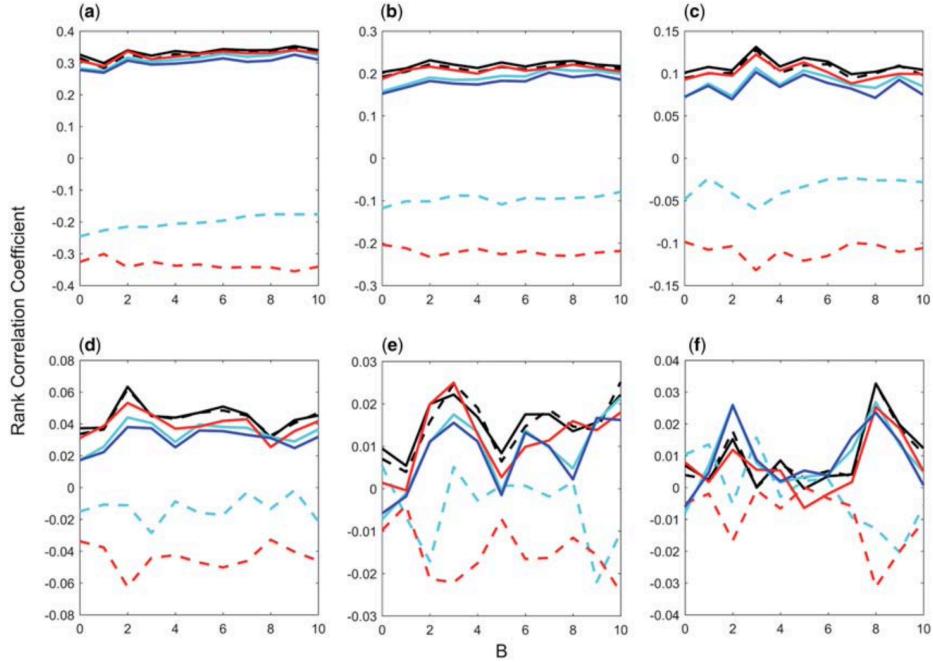


Fig.7: mostra i valori di correlazione di Spearman tra la centralità dei nodi e la loro influenza nella diffusione, in funzione di B nelle reti scale-free con $N = 500$ e grado medio di 10. B indica il parametro che controlla l'eterogeneità della rete, dove un valore più alto di B corrisponde a una minore eterogeneità della rete. I valori di soglia per i diversi pannelli e la legenda per i diversi stili di linea sono gli stessi della Figura 6.

Le reti scale-free sono esempi tipici di reti con gradi nodali eterogenei. Come esempi di reti con distribuzione quasi omogenea dei gradi, vengono considerate le reti small-world di Watts-Strogatz e le reti Erdős-Rényi.

La Figura 8 mostra la correlazione dell'influenza nella diffusione con le misure di centralità dei nodi per le reti di Watts-Strogatz con $N = 500$, $P = 0.2$ e diverso grado medio (le reti hanno grado medio pari a $2m$).

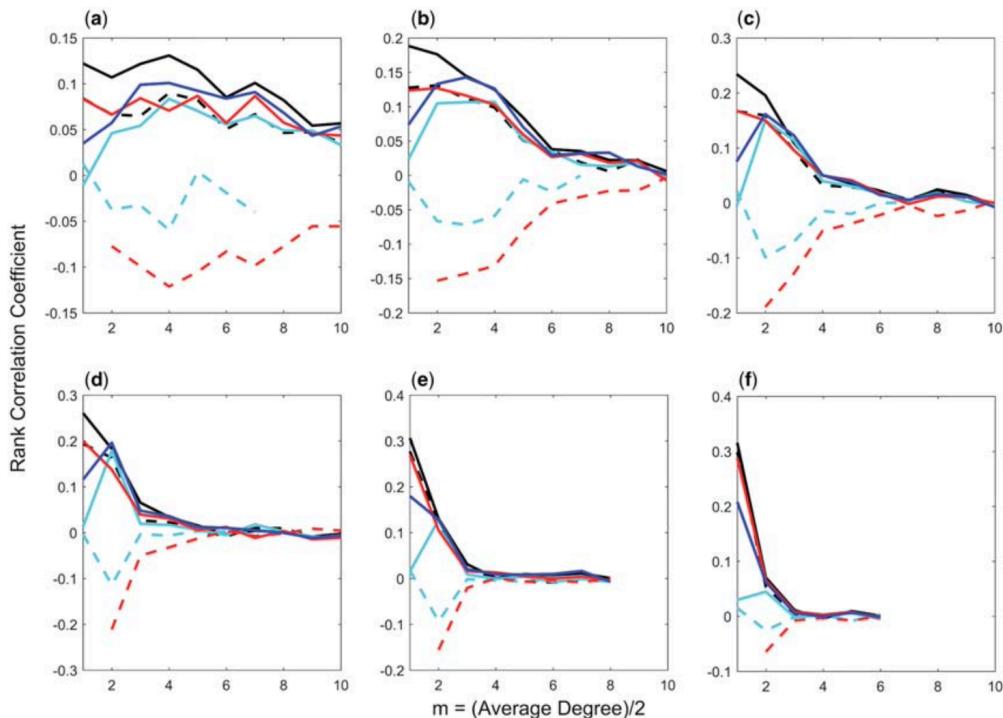


Fig. 8 : Mostra i valori di correlazione di Spearman tra la centralità dei nodi e la loro influenza nella diffusione, in funzione del grado medio ($2m$) nelle reti small-world di Watts-Strogatz con $N = 500$ e probabilità di ricollegamento $P = 0.2$. I valori di soglia per i diversi pannelli e la legenda per i diversi stili di linea sono gli stessi della Figura 6.

Si osserva un pattern quasi simile a quello delle reti scale-free (Figura 6), dove aumentando la densità della rete (cioè aumentando m), il valore assoluto delle correlazioni diminuisce. Di nuovo, l'indice di informazione e

l'eccentricità mostrano una correlazione negativa, con l'indice di informazione che ha la correlazione indiretta più forte. Altre misure di centralità hanno una correlazione positiva con l'influenza nella diffusione, con il grado che ha la correlazione diretta più forte.

La Figura 9 mostra i valori di correlazione per le reti di Watts-Strogatz con ri-collegamento variabile.

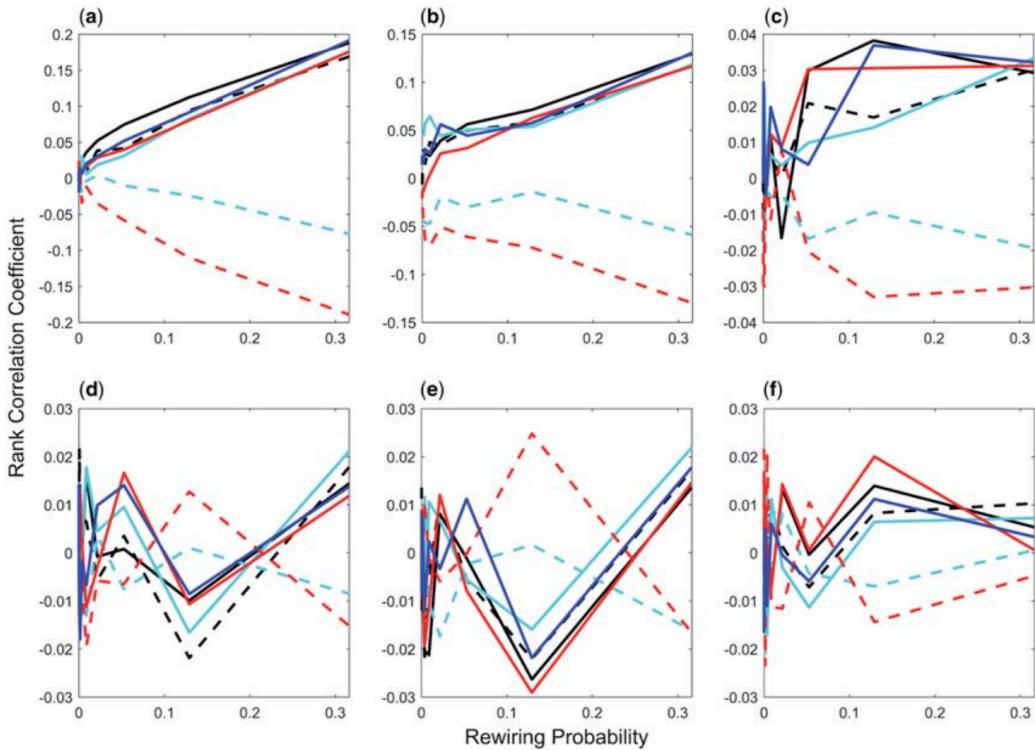


Fig.9: Mostra i valori di correlazione di Spearman tra la centralità dei nodi e la loro influenza nella diffusione, in funzione della probabilità di ricollegamento, nelle reti small-world di Watts-Strogatz con $N = 500$ e grado medio di 10. I valori di soglia per i diversi pannelli e la legenda per i diversi stili di linea sono gli stessi della Figura 6

Per evitare la diffusione di malattie contagiose e lo sviluppo di epidemie, si lavora sulla prevenzione e il controllo.

Le reti con elementi hub, come le reti scale-free con nodi ad alto grado, sono più suscettibili ai virus e alle malattie contagiose [37][38].

Queste reti non hanno una soglia epidemica, il che significa che una malattia può diffondersi molto rapidamente, mettendo così a rischio una grande percentuale dei nodi. Anche i virus informatici seguono schemi simili di diffusione e possono propagarsi velocemente sulla rete grazie alla sua topologia scale-free. Nelle reti con agenti mobili, come i social network, i nodi si spostano e creano nuove connessioni, il che può essere un fattore determinante nelle epidemie globali, come le pandemie di influenza H1N1 del 2009 e le epidemie di SARS del 2003 [39][40] [41].

È stato dimostrato che è possibile individuare un "distanza efficace" che predice approssimativamente il tempo di arrivo della malattia. Analizzando i modelli di mobilità su larga scala e i dati di traffico individuali, *Eubank et al.* hanno scoperto che le connessioni sono fortemente di tipo small-world, con una distribuzione dei gradi scale-free ben definita. Ciò consente una rilevazione efficiente degli

scoppio posizionando i sensori nei luoghi hub, dove sono coinvolti molti individui ad alto grado.

I nodi centrali (o hub) svolgono un ruolo significativo nel facilitare o bloccare la diffusione di malattie contagiose.

Studi come quello condotto da *Christley et al.* hanno dimostrato che una semplice misura di centralità, come il grado, può essere utilizzata efficacemente per identificare il rischio di infezione nelle popolazioni esposte a un virus [42].

Questo passaggio evidenzia che individuare i diffusori influenti è cruciale, poiché sono probabili diffusori veloci a causa della loro posizione centrale nella rete di contatti. D'altra parte, bloccandoli tramite vaccinazione o isolandoli temporaneamente fino a quando la diffusione non è controllata e il rischio di epidemie scompare, si può prevenire la rapida diffusione della malattia.

È stato dimostrato più volte che la vaccinazione mirata delle reti complesse è molto più efficace rispetto alla vaccinazione casuale[43][44].

Salathé et al. hanno raccolto reti sociali di interazioni in prossimità e analizzato le relazioni tra la diffusione della malattia e la struttura della rete. Hanno anche dimostrato che è possibile prevedere con precisione i casi reali di

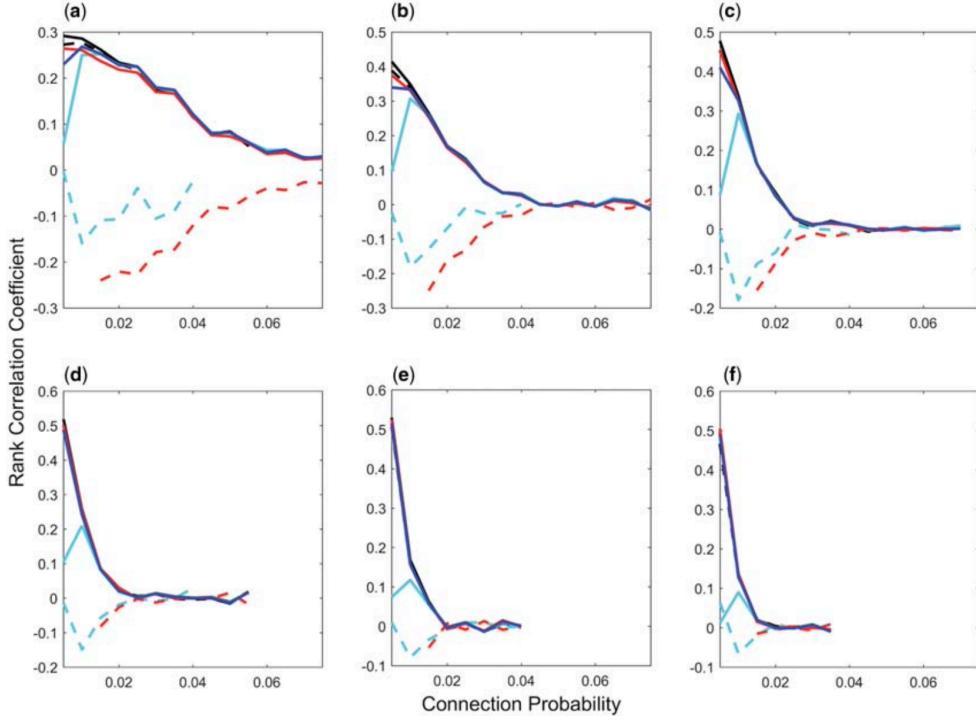


Fig.10: Mostra i valori di correlazione di Spearman tra la centralità dei nodi e la loro influenza sulla diffusione, in funzione della probabilità di connessione in reti casuali di Erdos-Renyi con $N = 500$ (dove N è il numero di nodi nella rete). La correlazione di Spearman è una misura di quanto bene i ranghi di due variabili si allineano tra loro.

influenza analizzando efficientemente la struttura della rete di prossimità [45].

Le informazioni di una tale rete possono essere utilizzate per progettare strategie di immunizzazione mirate ed efficaci. È possibile progettare strategie di immunizzazione più efficaci ottimizzando il costo della vaccinazione. *Mirzasoleiman et al.* hanno proposto un algoritmo semplice per progettare strategie di immunizzazione con un budget limitato [46].

Hanno dimostrato che, in condizioni di budget limitato, è possibile utilizzare una semplice strategia di sconto per decidere l'importo dello sconto da dare ai nodi centrali, dove l'importo dello sconto (per ricevere la vaccinazione) dipende dai valori di centralità.

La disposizione e la connettività delle relazioni all'interno di una rete sociale influenzano notevolmente la diffusione di idee, trend o comportamenti all'interno di quella rete.

Questo concetto è fondamentale nel marketing virale, dove il passaparola e la condivisione di contenuti da parte degli individui all'interno delle loro reti sociali sono utilizzati per promuovere un prodotto, un servizio o un'idea. Comprendere la struttura della rete sociale aiuta i marketer a identificare chi sono gli individui più influenti o centrali all'interno della rete, i cosiddetti "*influencer*", e a pianificare strategie mirate per massimizzare la diffusione del messaggio virale [47] [48].

In *figura 11*, i grafici riportano:

- (a) Rete di co-autori di scienziati attivi nel campo della Scienza delle reti (Net Sci),
- (b) Reti sociali tra dottorandi in Informatica (CS PhD),
- (c) Rete simile a Facebook,

- (d) Rete del Dizionario Online di Biblioteconomia e Scienza dell'Informazione (ODLIS),
- (e) Aeroporti degli Stati Uniti, Internet a livello di Sistema Autonomo (AS),
- (f) Rete elettrica degli Stati Uniti,
- (g) Rete elettrica dell'Unione Europea.

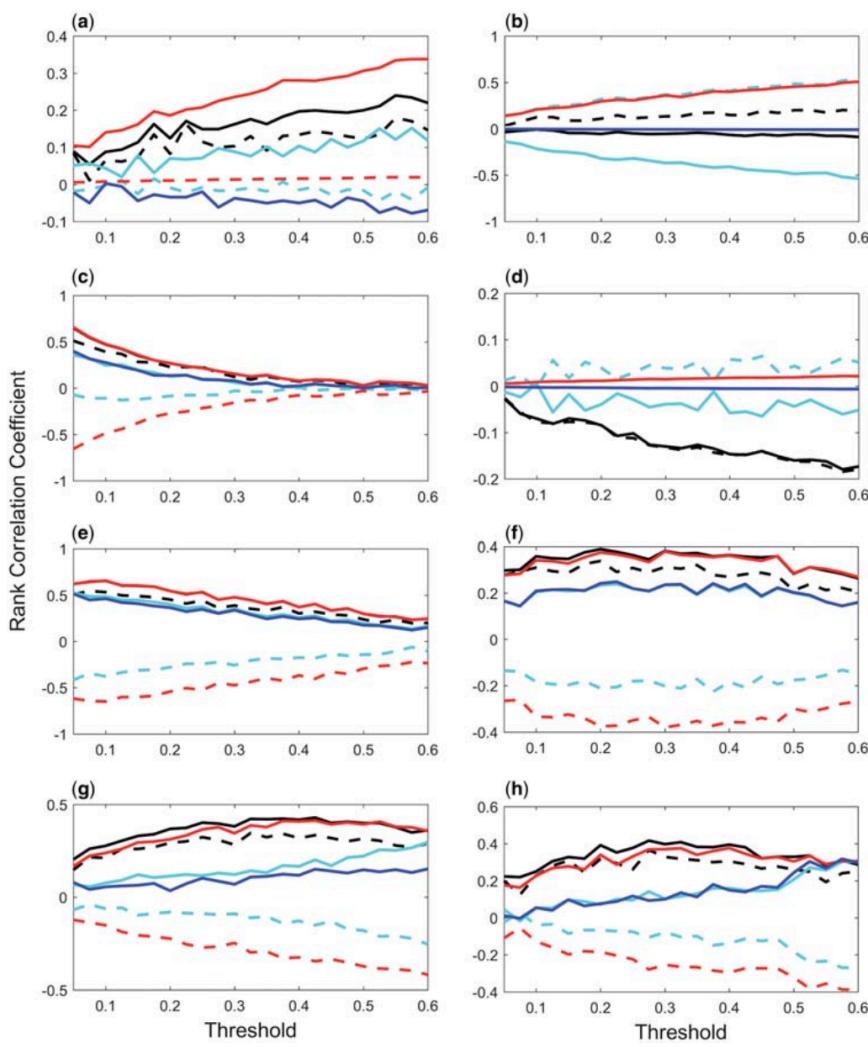


Fig.11: Mostra i valori di correlazione di Spearman tra la centralità dei nodi e la loro influenza sulla diffusione, in funzione della soglia dei modelli di cascata indipendente in diversi network reali.

La legenda per i diversi stili di linea è la stessa della Figura 6, che

indica vari tipi di centralità dei nodi o altre caratteristiche rilevanti.

2.5.2.1 MASSIMIZZAZIONE DELL'INFLUENZA :

“ IDENTIFICARE I NODI CHIAVE ”

In questo paragrafo si discute di come identificare i nodi in una rete sociale che possono avviare una grande diffusione di informazioni o comportamenti [49]. I "**nodi top-k**" sono quelli considerati più importanti in questo processo e spesso vengono selezionati per ricevere un prodotto gratuito o scontato nelle campagne di marketing, con la speranza che influenzino altri a fare lo stesso.

Il problema di trovare questi nodi cruciali è piuttosto complesso e richiede tempo computazionale. Tuttavia, esistono approcci approssimati che possono aiutare a trovare una soluzione abbastanza buona.

Alcuni studi hanno dimostrato che la struttura della rete, in particolare la presenza di comunità dense e vicinato clusterizzato, può influenzare la diffusione delle

informazioni. Alcuni algoritmi sono stati sviluppati per trovare questi nodi cruciali in modo efficiente, ma rimangono ancora più lenti rispetto ai metodi che si basano sull'individuazione dei nodi più centrali nella rete.

In sostanza, individuare i nodi chiave in una rete sociale è cruciale per massimizzare la diffusione di informazioni o comportamenti, e ci sono vari approcci per farlo, anche se non sempre sono veloci o precisi al cento per cento.

Dunque, gli individui che interagiscono in una struttura di rete possono influenzare i loro vicini e/o essere influenzati da loro, cambiando di conseguenza i loro valori di opinione.

Esistono vari modelli proposti per studiare l'evoluzione delle opinioni nelle reti complesse, che possono essere generalmente categorizzati in due ampie classi: *modelli con opinioni a tempo continuo* e quelli *con valori di opinione a tempo discreto* [50] [51].

Nella maggior parte degli studi sulla formazione delle opinioni, gli agenti vengono inizialmente assegnati con un certo valore di opinione e poi adattano i loro valori di opinione considerando quelli dei loro vicini che possono influenzarli. Ad esempio, nel noto modello di "confidenza limitata", si sceglie prima un bordo casuale in ogni iterazione del processo evolutivo. Successivamente, i nodi

terminali del bordo selezionato cambiano le loro opinioni (rendendole più vicine) se le loro opinioni sono abbastanza vicine, cioè sono negoziabili. Questa procedura continua fino a ottenere una soluzione di stato stazionario e non avvengono ulteriori cambiamenti di opinione. Se determinate condizioni sono soddisfatte, tutti (o la maggior parte) degli agenti raggiungono un consenso sui valori delle loro opinioni.[52]

Nelle reti sociali reali, i diversi nodi hanno diversi livelli di influenza sui loro amici. Ad esempio, i leader della società spesso hanno un ruolo più significativo nel plasmare l'opinione della folla rispetto agli individui normali. Allo stesso tempo, i leader potrebbero essere meno influenzati dagli altri rispetto agli individui normali. Infatti, gli individui hanno diversi livelli di potere sociale. Studi hanno mostrato che l'introduzione del potere sociale nel processo di formazione delle opinioni spesso facilita il consenso.

Un piccolo numero di agenti informati con una posizione strategica nella rete può plasmare con successo l'opinione dell'intera società. Collegando gli agenti informati agli agenti con gradi piuttosto bassi ma con vicini di alto grado, si può massimizzare l'influenza [53]. Anche se le misure di centralità generali sono state utilizzate con successo per

facilitare la formazione delle opinioni nelle reti complesse, la costruzione di misure di centralità specifiche per le proprietà del processo di formazione delle opinioni è ancora un argomento molto discusso in questo campo.

2.6 SICUREZZA E GESTIONE DEI RISCHI NELLE RETI COMPLESSE

In ambito delle reti, il fallimento può esser legato anche alla diffusione dei guasti a catena .

Le strutture di rete del mondo reale, come quelle delle reti di distribuzione dell'acqua o delle reti elettriche, possono essere soggette a guasti nei loro componenti (nodi e archi), che possono avvenire in due forme: *guasto casuale* (errore) o *guasto intenzionale* (attacco).

Mentre molte reti dimostrano una sorprendente resilienza contro gli errori casuali, sono fragili contro gli attacchi intenzionali. In alcuni casi, un guasto in un componente può avere conseguenze più drastiche, innescando una cascata di guasti in altri componenti.

Lo studio delle cascata di guasti e la comprensione del loro comportamento ha un ruolo significativo nella salvaguardia

delle infrastrutture critiche in rete, come le reti di distribuzione dell'acqua e le reti elettriche.

Ad esempio, si è scoperto che i guasti dei componenti possono spesso innescare una cascata di guasti nelle reti di distribuzione dell'acqua. Le reti elettriche sono anche soggette a guasti in cascata, responsabili di alcuni black-out su larga scala [54].

La posizione del guasto iniziale ha un ruolo significativo nell'esito della cascata, ovvero il numero di componenti guasti. Alcuni studi hanno scoperto che, in molti tipi di reti, il grado di un nodo è correlato negativamente con la profondità della cascata, il che indica che un guasto iniziale nei nodi ad alto grado spesso non ha un effetto significativo in cascata [55]. Altri studi hanno identificato misure di centralità con correlazione positiva con la profondità della cascata.

Infine, anche le istituzioni finanziarie spesso hanno interconnessioni intense, e un rischio in una determinata istituzione finanziaria può diffondersi ad altre, portando a risultati finanziari drastici. La struttura della rete finanziaria ha un ruolo significativo nello sviluppo dei rischi nelle istituzioni finanziarie partecipanti [56].

Le reti sociali hanno varie applicazioni e possono propagare anomalie e attività illegali. La rilevazione di anomalie nelle reti sociali viene spesso utilizzata per identificare spammer, predatori sessuali, utenti malintenzionati, attività terroristiche, riciclaggio di denaro e truffatori online. Il *comportamento anomalo* è definito come qualcosa che è significativamente diverso dal comportamento normale delle reti [57][58].

Ad esempio, i nodi spuri spesso hanno comportamenti (e caratteristiche) significativamente diversi rispetto ai nodi normali. I collegamenti spuri sono quelli che non dovrebbero esistere se la rete si evolve in modo naturale. Non è sempre un compito facile individuare i criminali nella rete poiché gli individui che intendono compiere attività criminali attraverso le reti sono intelligenti e spesso trovano nuovi modi per non essere riconosciuti. Anche se i leader delle attività illegali hanno una posizione strategica nella rete, in molti casi non hanno molte connessioni, ma solo alcune critiche. Le attività di riciclaggio di denaro spesso coinvolgono un certo numero di individui collaborativi, e le prove possono essere scoperte solo quando viene considerato il comportamento collettivo di

questi individui. Per avere una rilevazione efficace delle attività di riciclaggio di denaro, è possibile utilizzare efficientemente l'analisi delle reti e l'apprendimento supervisionato.

Sebbene ci siano stati molti lavori sulla previsione dei collegamenti futuri nelle reti sociali, ci sono stati pochi sforzi per prevedere i collegamenti anomali, in parte a causa della mancanza di set di dati affidabili. *Magalingam et al.* hanno proposto un metodo basato sui percorsi più brevi per trovare un piccolo sottoinsieme di nodi sospetti di attività criminali [59]. *Yasami e Safaei* hanno proposto un metodo a più passaggi per la rilevazione di anomalie nelle reti sociali. Nel loro metodo proposto, il primo passo è identificare il comportamento normale della rete, e le anomalie vengono identificate nel passo successivo [60]. *Das e Sinha* hanno utilizzato misure di centralità dei nodi per rilevare utenti malintenzionati nelle reti sociali [61]. *Kaur et al.* hanno utilizzato misure di centralità per identificare utenti criminali e hanno dimostrato che combinando diverse misure di centralità è possibile identificare meglio tali utenti rispetto al caso di utilizzo delle misure di centralità in modo indipendente [62].

CAPITOLO 3.

PIANO DI STUDIO

In questo capitolo si illustrerà lo studio proposto, il monitoraggio delle epidemie di Morbillo. Con un focus particolare sui dati relativi alle casistiche e sorveglianza dell'epidemie.

3.1 PANORAMICA SUL MORBILLO

Il **morbillo** è una malattia infettiva esantematica acuta causata da un *virus a RNA del genere Morbillivirus* (famiglia dei Paramyxoviridae), di cui esiste un unico sierotipo. Il morbillo è diffuso in tutto il mondo, e l'uomo è l'unico ospite naturale.

Prima dell'introduzione del vaccino contro il morbillo negli anni 60, e dei programmi estesi di vaccinazione, si verificavano epidemie all'incirca ogni due o tre anni che, a livello globale, causavano un numero stimato di 2,6 milioni di morti ogni anno. Grazie alla vaccinazione, si è verificata una significativa riduzione dell'incidenza del morbillo nel mondo.

Tuttavia, la malattia rimane un'importante causa di morbilità e mortalità: si stima che nel 2022 siano morte di morbillo 136.200 persone, per lo più bambini di età inferiore ai cinque anni. La maggior parte dei decessi avviene in Paesi sottosviluppati [63].

Tutte e sei le regioni dell'*Organizzazione Mondiale della Sanità (OMS)* hanno fissato obiettivi di eliminazione del morbillo (interruzione della trasmissione endemica in una determinata area geografica). L'eliminazione del morbillo richiede livelli di immunità molto elevati nella popolazione. Ad oggi, la Regione delle Americhe è l'unica ad avere raggiunto, nel 2016, questo traguardo.

Dunque è una malattia virale altamente contagiosa causata da virus.

Ecco alcune informazioni chiave sul morbillo:

1. *Modalità di trasmissione:* Il morbillo si trasmette principalmente attraverso il contatto diretto con le goccioline respiratorie di una persona infetta. Questo può avvenire tramite la tosse, gli starnuti o il contatto

diretto con le secrezioni nasali o della gola di una persona infetta.

- 2. Sintomi:** I sintomi del morbillo di solito iniziano circa 10-12 giorni dopo l'esposizione al virus. Inizialmente, compaiono sintomi simili a quelli dell'influenza, come febbre, tosse, mal di gola, congestione nasale e malessere generale. Successivamente, si sviluppa una eruzione cutanea rossa, che inizia di solito sul viso e si diffonde al resto del corpo.
- 3. Complicazioni:** Sebbene la maggior parte delle persone si riprenda completamente dal morbillo senza complicazioni, la malattia può essere grave, specialmente nei bambini piccoli, negli adulti e nelle persone con sistemi immunitari compromessi. Le complicazioni possono includere infezioni dell'orecchio, polmonite, encefalite (infiammazione del cervello) e, in casi estremamente rari, panencefalite sclerosante subacuta, una forma letale di malattia del cervello.
- 4. Prevenzione:** La prevenzione del morbillo si basa principalmente sull'immunizzazione attraverso il vaccino MPR (morbillo-parotite-rosolia). Questo vaccino è altamente efficace nel prevenire il morbillo e deve essere somministrato in due dosi, generalmente

intorno all'età di 12-15 mesi e poi tra i 4 e i 6 anni. L'immunità di gregge, cioè la protezione fornita dalla vaccinazione di una grande percentuale della popolazione, è fondamentale per controllare la diffusione del morbillo.

5. *Eradicazione*: Prima dell'introduzione del vaccino contro il morbillo, questa malattia era diffusa in tutto il mondo. Tuttavia, grazie agli sforzi di vaccinazione su larga scala, il morbillo è stato eliminato da molti paesi e regioni. Tuttavia, sono ancora presenti focolai e occasionali epidemie in aree dove la copertura vaccinale è bassa.

In ambito diagnostico non è sempre facile diagnosticarlo, soprattutto nei periodi di bassa incidenza della malattia [63].

La conferma della diagnosi si ottiene generalmente attraverso la ricerca nel siero di **anticorpi IgM** specifici diretti contro il virus del morbillo, oppure osservando un aumento significativo degli **anticorpi IgG** tra la fase acuta e la fase convalescente.

Un altro metodo si basa sull'identificazione di particelle virali mediante *amplificazione genica* (RT-PCR).

Colpisce tutti gli individui non vaccinati, e che non hanno mai contratto la malattia in vita in loro, sono potenzialmente a rischio di ammalarsi di morbillo. Una seconda infezione è infatti possibile, ma assai improbabile.

Il morbillo è considerato una malattia infettiva infantile, come la rosolia, la varicella, la pertosse e la parotite, in quanto colpisce soprattutto i bambini tra i 12 mesi e i 4 anni (si sta parlando, ovviamente, di soggetti a rischio, cioè non vaccinati e che non si sono mai ammalati di morbillo in vita loro).

Oggi, la sua diffusione (specie nei paesi industrializzati, Italia compresa) è calata notevolmente grazie alla realizzazione di un vaccino efficace.

Ciononostante, si verificano ancora dei casi mortali: alcune fonti autorevoli, infatti, riportano che la morte dovuta alle complicazioni del morbillo coinvolge dai 30 ai 100 individui ogni 100.000 persone malate.

3.1.1 IL MORBILLO IN ITALIA

Da molti decenni, ormai, nel nostro paese ogni caso di morbillo va notificato obbligatoriamente alle autorità sanitarie presenti sul territorio.

Inoltre, in seguito all'epidemia scoppiata tra il 2002 e il 2003, è stato istituito il ***sistema di sorveglianza speciale del morbillo***: in base a questo protocollo, il medico deve segnalare ogni caso sospetto all'Asl più vicina, la quale, ricevuta la segnalazione, ha il compito di svolgere tutte le indagini di laboratorio più appropriate e comunicarne gli esiti al Ministero della salute e al Centro nazionale di epidemiologia, sorveglianza e promozione della salute (**Cnesps**) [63].

Come precedentemente detto, il morbillo è causato da un virus appartenente al genere *morbillivirus*, il quale fa parte, a sua volta, dell'ordine *mononegavirales* e della famiglia *paramyxoviridae*.

I morbillivirus sono dotati di **capside** (un involucro esterno che protegge il virus dalle insidie dell'ambiente esterno) e possiedono, per materiale genetico, un solo filamento di RNA.

Esistono varie specie di morbillivirus; tali specie, che possono infettare non solo l'essere umano ma anche altri mammiferi, sono:

- Il virus del cimurro canino
- Il morbillivirus dei cetacei
- Il virus del morbillo
- Il virus della peste dei piccoli ruminanti
- Il virus del cimurro delle foche
- Il virus della peste bovina

3.1.2 SINTOMI E PREVENZIONE

La comparsa dei primi sintomi avviene dopo una decina di giorni dal contagio (periodo di incubazione).

Tali sintomi consistono in:

- Manifestazioni tipiche del raffreddore, ovvero: naso che cola, occhio che lacrima, palpebre gonfie, starnuti ecc.
- Occhi arrossati e sensibilità alla luce
- Febbre alta, anche a 40°C
- Stanchezza, irritabilità e senso di mancanza di forze
- Dolori e malessere

- Tosse secca
- Formazione di piccole macchie grigio-biancastre a livello della bocca e della gola (segno di Köplik)
- Perdita di appetito
- La durata di tale sintomatologia è di circa 7-10 giorni.

Il segno distintivo del morbillo è il **rash cutaneo** (o eruzione cutanea), il quale determina la comparsa, su tutto il corpo, di macchie rossastre, leggermente in rilievo e di grandezza variabile.

La diffusione sul resto del corpo richiede uno o due giorni, mentre, per la completa scomparsa, bisogna attendere in genere dai 4 ai 7 giorni.

In ambito preventivo, il morbillo può essere “combattuto” con il **vaccino MPR** (dove *M* sta per *Morbillo*, *P* per *Parotite* e *R* per *Rosolia*).

Tale vaccinazione va eseguita durante l'infanzia, con due iniezioni: una a 12-13 mesi circa e un'altra a 5-6 anni (di solito poco prima di iniziare la scuola elementare).

Caso particolare:

Per gli individui adulti non vaccinati e i bambini al di sotto di un anno di vita, il vaccino non è indispensabile/previsto. Lo diventa, però, qualora fosse concreto il rischio di un contagio.

Due classiche situazioni che richiedono una vaccinazione, per così dire, straordinaria sono: un viaggio in zone geografiche in cui il virus è altamente diffuso oppure l'infezione capitata a un membro molto stretto della famiglia.

I bambini al di sotto dei 6 mesi di vita nati da madre vaccinata o immune godono di un'immunità di origine materna che li protegge temporaneamente; pertanto, non hanno bisogno di vaccinarsi.

Al contrario, i bambini della stessa età che non hanno una madre vaccinata o immune, sono sprovvisti di protezione e più predisposti a contrarre la malattia. Per questi soggetti, quindi, potrebbe essere presa in considerazione l'ipotesi di eseguire un'iniezione di immunoglobuline protettive.

Molto importante è ricordare che le immunoglobuline, dette anche anticorpi, sono delle proteine del sistema immunitario

che si occupano di combattere gli agenti infettivi, e di formare la cosiddetta memoria immunologica. Una loro iniezione extra, come nel suddetto caso, non equivale a un vaccino, ma può rappresentare comunque una valida soluzione antinfettiva.

In fase di guarigione, bisogna chiarire che il sistema immunitario rappresenta il sistema difensivo contro le minacce provenienti dall'ambiente esterno (in primis, gli agenti infettivi di natura virale o batterica).

Esso, quando combatte gli agenti infettivi come i virus, prepara anche delle cellule di difesa particolari, capaci di riconoscere in anticipo la medesima minaccia e impedire una seconda infezione.

Questo prodigioso meccanismo si chiama memoria immunitaria e le cellule che lo mettono in pratica (che sono degli anticorpi particolari) sono dette cellule della memoria.

Le donne non vaccinate e non immuni, che vogliono avere un figlio, possono rivolgersi al proprio medico e chiedere delucidazioni sul da farsi. La vaccinazione contro il morbillo, in questi frangenti, non è indispensabile e può essere sostituita da un'iniezione di immunoglobuline [64].

I vaccini antivirali sono realizzati sul concetto di memoria immunitaria.

3.2 MONITORAGGIO E ANALISI

Il *monitoraggio* e l'*analisi del morbillo* sono fondamentali per controllare e prevenire le epidemie. Questo processo coinvolge diverse attività.

Studi recenti e articoli dimostrano l'importanza di un monitoraggio e di un'analisi accurati per comprendere e affrontare le disparità regionali nella copertura vaccinale e per migliorare le strategie di immunizzazione a livello locale e nazionale.

A livello nazionale dopo l'approvazione del *decreto-legge n. 73/2017* sulla vaccinazione obbligatoria, sono stati osservati miglioramenti nella copertura vaccinale, sebbene permangano sfide legate all'implementazione a livello regionale. In Africa Sub-Sahariana, la prevalenza dell'utilizzo della seconda dose di vaccino contro il morbillo (MCV2) varia tra i paesi, con una media del 44,77% [65].

Prima di tutto si effettua una raccolta dati.

La *raccolta dati* è un elemento fondamentale nel monitoraggio e controllo delle epidemie di morbillo. L'obiettivo principale di questa attività è ottenere informazioni accurate e tempestive sull'incidenza della malattia, sulla copertura vaccinale e sulla diffusione geografica. Questi dati sono essenziali per analizzare l'andamento dell'epidemia e per prendere decisioni informate sulle misure di intervento.

Nello specifico *raccolta sistematica* di dati sulle infezioni, compresi i nuovi casi, i decessi e la copertura vaccinale.

Provenienti da strutture sanitarie, studi di sorveglianza e rapporti di vaccinazione.

I dati raccolti sono utilizzati per:

- *Identificare Focolai*: Rilevamento precoce di focolai per intervenire tempestivamente.
- *Valutare la Copertura Vaccinale*: Monitoraggio dell'efficacia delle campagne di vaccinazione.
- *Pianificare Interventi*: Sviluppo di strategie mirate per controllare la diffusione del morbillo.

- *Informare le Politiche Sanitarie:* Supporto alla formulazione di politiche basate sull'evidenza per migliorare la salute pubblica.

Si procede con la sorveglianza epidemiologica, analisi dei dati per identificare tendenze, focolai, e modelli di trasmissione. L'uso di strumenti analitici avanzati e software specifici. Quest'ultimi aiutano a rilevare focolai emergenti e a valutare l'efficacia delle campagne di vaccinazione. Includendo la raccolta di dati attraverso notifiche obbligatorie di casi, indagini epidemiologiche sul campo e studi di sieroprevalenza per valutare l'immunità della popolazione.

Andando a rappresentare la visualizzazione dei dati , creando dei report e grafici per presentare i risultati delle analisi. Strumenti di visualizzazione dei dati, come mappe e grafici interattivi, facilitano la comprensione delle tendenze epidemiologiche e supportano la comunicazione con le autorità sanitarie e il pubblico.

Sviluppo e implementazione di strategie basate sui dati raccolti e analizzati per contenere e prevenire la diffusione

del morbillo. Questi interventi possono includere campagne di vaccinazione, sensibilizzazione pubblica e misure di quarantena.

3.3 STUDIO PROPOSTO

Nel contesto di sviluppo software, un *repository* contiene il codice sorgente di un progetto, insieme alla sua cronologia di versioni, documentazione e altri materiali rilevanti. I repository possono essere ospitati su piattaforme come GitHub, GitLab o Bitbucket, che facilitano la collaborazione tra sviluppatori, il controllo delle versioni e la gestione delle modifiche.

Tramite **GitHub**, una piattaforma di hosting per il controllo delle versioni e la collaborazione, che ospita milioni di *repository* di progetti open-source e privati, diventando una risorsa essenziale per sviluppatori e team di tutto il mondo, si è studiato il *repository GitHub "measles"*.

Fornendo una serie di strumenti e dati per il monitoraggio dell'epidemia di morbillo.

Esso include vari file CSV contenenti dati sulle epidemie e sulla copertura vaccinale a livello nazionale, europeo e globale.

Inoltre, il repository offre il pacchetto "*EpiSignalDetection*" per rilevare possibili focolai epidemici attraverso l'analisi dei dati di sorveglianza delle malattie infettive. Questo pacchetto permette di generare un rapporto automatico in formato HTML, facilitando la presentazione e l'analisi dei risultati.

Nel dettaglio, ci sono:

- 1)**File CSV:** Contengono dati sulle epidemie di morbillo e sulla copertura vaccinale.
- 2)**Pacchetto EpiSignalDetection:** Utilizzato per identificare focolai di morbillo tramite dati di sorveglianza.
- 3)**Applicazione Interattiva:** Permette l'analisi dinamica delle serie temporali.
- 4)**Rapporto Automatico:** Genera report in HTML per una facile condivisione e visualizzazione dei risultati.

Il repository può essere utilizzato da epidemiologi, ricercatori e autorità sanitarie per monitorare l'andamento delle epidemie di morbillo, identificare rapidamente focolai emergenti e valutare l'efficacia delle campagne di vaccinazione. Inoltre, i dati e gli strumenti forniti possono

supportare decisioni informate e strategie di intervento per contenere la diffusione della malattia.

3.3.1 FASI DI SVILUPPO

La fase principale ruota attorno al raggruppo delle idee. Il focus di questo studio, verte sulla simulazione, di dati reali per la comprensione del modello ottimale di trasmissibilità.

Nel dettaglio i dati prima di essere analizzati sono stati, puliti eliminando valori mancanti o errati. Sono stati poi esplorati per comprendere la distribuzione in base alle fasce d'età. Valutando i diversi modelli per identificare quale rappresenta al meglio la trasmissibilità in base alle fasce d'età.

year	month	age_group	Unit	cases	cases_percent	incidence
2014	1	0_4	N/100000	36	13.5	NA
2014	1	5_14	N/100000	68	25.6	NA
2014	1	15_39	N/100000	133	50.0	NA
2014	1	40_64	N/100000	29	10.9	NA
2014	1	65	N/100000	0	0.0	NA
2014	2	0_4	N/100000	60	12.1	NA
2014	2	5_14	N/100000	116	23.5	NA
2014	2	15_39	N/100000	272	55.1	NA
2014	2	40_64	N/100000	46	9.3	NA
2014	2	65	N/100000	0	0.0	NA
2014	3	0_4	N/100000	100	12.6	NA
2014	3	5_14	N/100000	172	21.6	NA
2014	3	15_39	N/100000	440	55.3	NA
2014	3	40_64	N/100000	81	10.2	NA
2014	3	65	N/100000	2	0.3	NA

Tabella : Rappresentazione di tutti i dati a livello nazionale

Sempre in questa fase si decide come dividere e sezionare i dati. Andando a scegliere solo le colonne di reale interesse, per lo studio.

L'interesse è focalizzato sui dati riferiti alla fascia d'età (*age_group*), anno e mese (*year - month*) e casi (*cases*).

Su un foglio excel si è lavorato il file, tramite formule per andare a suddividere i *Casi* in:

- *Casi Nuovi*
- *Casi Totali*

The screenshot displays two tables in an Excel spreadsheet. The top table is titled "Nuovi casi per mese" and has columns "Casi" and "Mese". The bottom table is titled "Casitotali" and also has columns "Casi" and "Mese". Both tables show data for months 1 through 12. The "Casitotali" table includes a row for the total number of cases (499) across all months.

Anno di riferimento		2020
age_group	Tutti	-
Nuovi casi per mese		
Casi	Mese	
0	1	
0	2	
0	3	
0	4	
97	5	
100	6	
0	7	
100	8	
0	9	
100	10	
0	11_12	
102	12	
Casitotali		
Casi	Mese	
0	1	
0	2	
0	3	
0	4	
97	5	
197	6	
197	7	
297	8	
297	9	
397	10	
397	11	
499	12	

Fig.12: Rappresentazione tabellare dei dati

Avendo come riferimento il mese di Gennaio fino al mese di Dicembre, scritti in numerazione da 1 a 12, comprendendo il periodo dal 2014 al 2024. Lo studio è stato svolto su dati nazionali.

Questi rientrano nella parte reale. E fungeranno da controllo per la simulazione finale. Ottenuti, tramite le funzioni excel.

In particolare, due celle sono variabili, una è in *verde*, per rappresentare l'anno di riferimento, e l'altra è rappresentata dalla fascia d'età.

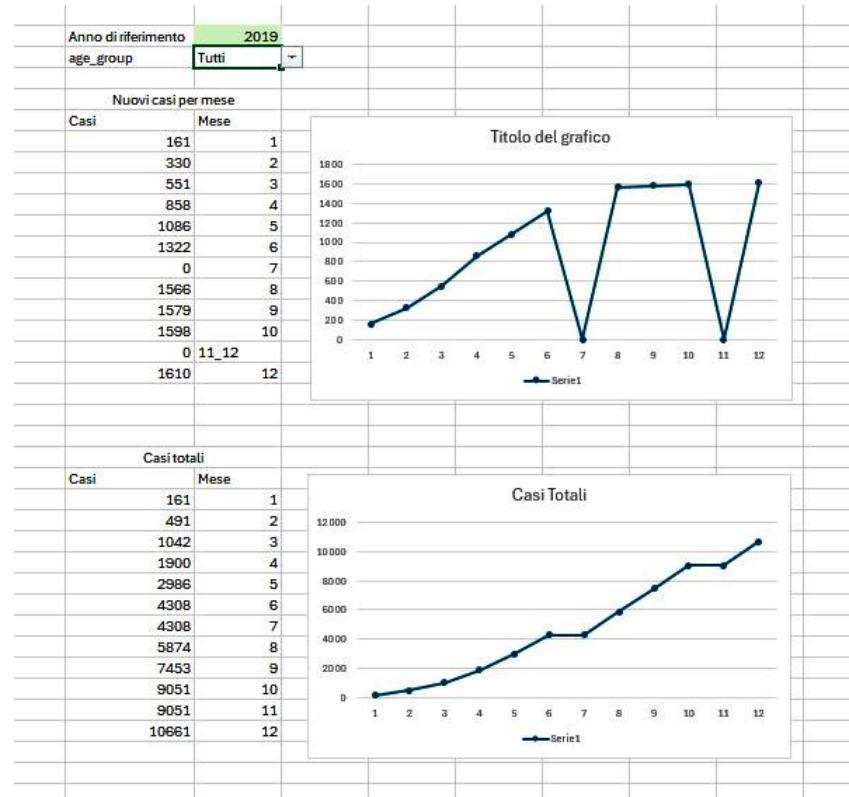


Fig. 13: Esempio riferimento casi reali

Per calcolare i *Casi Nuovi* si ha una logica di restituzione di un valore, se una condizione specificata dà come risultato VERO, e un altro valore dà come risultato FALSO.

Test	<code>U\$5="Tutti"</code>	= VERO
Se_vero	<code>SOMMA.PIÙ.SE(\$E\$2:\$E\$331;\$B\$1)</code>	= 27
Se_falso	<code>SOMMA.PIÙ.SE(\$E\$2:\$E\$331;\$B\$1)</code>	= 0
		= 27

Fig. 14: Esempio formule utilizzate in Excel

Il **test** è un valore o un'espressione che darà il risultato.

Mentre, per ottenere i *Casi Totali* si è utilizzata una funzione base di somma, sommando i numeri presenti in un intervallo di celle.

SOMMA (num1; num2;...)

Quindi vengono calcolati mese per mese, in un anno. Includendo o escludendo in base alle esigenze le fasce d'età.

Le rappresentazioni grafiche, saranno la base di controllo per lo sviluppo di simulazione.

Dove si vede sull'asse delle x i mesi da Gennaio (1) a Dicembre (12), e sulle y il numero degli individui infetti (ovvero i casi reali).

3.3.2 SIMULAZIONI

In fase di sviluppo si è deciso di dividere le simulazioni in due tipologie. Così da avere il controllo del modello.

Il lavoro di programmazione è stato svolto su *Google Colab*, piattaforma basata su cloud che consente di scrivere e eseguire codice Python attraverso il browser.

```
from collections import Counter
from operator import itemgetter

import matplotlib as mpl
import matplotlib.pyplot as plt
import networkx as nx

class StopCondition(StopIteration):
    pass

class Simulation:
    '''Simulate state transitions on a network'''
```

Nello specifico si è utilizzato un codice che ha una classe di “*Simulation*”, utilizzata per simulare le transizioni di stato su una rete. Sviluppato con una libreria “*networkx*”

per la gestione delle reti e infine “ *matplotlib* ” per la visualizzazione.

La classe “ *StopCondition* ” viene usata per interrompere la simulazione quando una condizione di arresto è soddisfatta.

Nell'inizializzazione, vengono specificati i seguenti temi:

- **G** è un'istanza di *networkx.Graph*
- **initial_state** è una funzione che definisce lo stato iniziale dei nodi
- **state_transition** è una funzione che definisce come gli stati dei nodi cambiano a ogni passo.
- **stop_condition** è una funzione opzionale che determina quando interrompere la simulazione.

```
%matplotlib inline
import networkx as nx

# To Do: Load an appropriate graph
G = nx.gnm_random_graph(200, 50)
#nx.draw(G)
```

L'utilizzo della classe *simulation* è combinato da alcuni passaggi fondamentali, affinché il lavoro venga svolto in modo corretto.

Nello specifico :

- A. Creare un grafo networkx;
- B. Definire una funzione per lo stato iniziale dei nodi;
- C. Definire una funzione per la transizione di stato dei nodi;
- D. Definire una funzione per la condizione di arresto (opzionale);
- E. Creare un'istanza della classe “*simulation*” con questi parametri.

Poiché le simulazioni, si basano sulle casistiche di individui affetti da morbillo, vanno introdotti i tassi che influenzano la probabilità con cui un nodo infetto guarisce o un nodo suscettibile si infetta. Questi parametri sono stati definiti per determinare le probabilità di tradizione degli stati.

```
# Parametri del modello SIR
beta = 0.3 # Tasso di infezione
gamma = 0.1 # Tasso di recupero
initial_infected = 1 # Numero iniziale di individui infetti
num_steps = 100 # Numero di passi della simulazione
```

Per implementare la simulazione, si è utilizzata la modellistica a compartimenti (SIR).

Dove si ha uno stato iniziale alla quale, tutti i nodi iniziano come suscettibili (S) tranne uno scelto casualmente che è infetto (I). Questa implementazione del modello SIR è appropriata per studiare la diffusione del morbillo, dato che tiene conto dell'immunità permanente acquisita dopo la guarigione.

Se i parametri utilizzati nella simulazione SIR (tasso di infezione, tasso di guarigione, ecc.) sono scelti accuratamente in base ai dati storici della malattia, la simulazione può riprodurre in modo realistico l'andamento dell'epidemia.

Prima simulazione:

Si analizza la prima simulazione che vede l'*esclusione delle fasce d'età*. Concentrandosi prettamente sui casi e i relativi anni di riferimento. Gli andamenti vengono modellati mese per mese.

```
# Percorso
csv_filename = '/content/drive/MyDrive/TESI_/measles-ita-age-latest.csv' # A

# Carica il grafo dal file CSV
M = load_graph_from_csv(csv_filename, threshold=0.3)

# Disegna il grafo
pos = nx.spring_layout(M) # Posizioni dei nodi per il layout del grafico
cases = nx.get_node_attributes(M, 'cases')
node_sizes = [cases[node] * 10 for node in M.nodes()] # Scala le dimensioni

plt.figure(figsize=(12, 8))
nx.draw(M, pos, with_labels=True, node_size=node_sizes, node_color='skyblue',
plt.title('Grafo basato su "cases"')
plt.show()
```

Vengono analizzati “ 20 andamenti ” , divisi in casi appartenenti a nuovi e totali.

La *curva blu* rappresenta i suscettibili (S), mentre quella *arancione* è la curva degli infetti (I), infine la *verde* è dei guariti (R)

Analizzando dieci anni consecutivi.

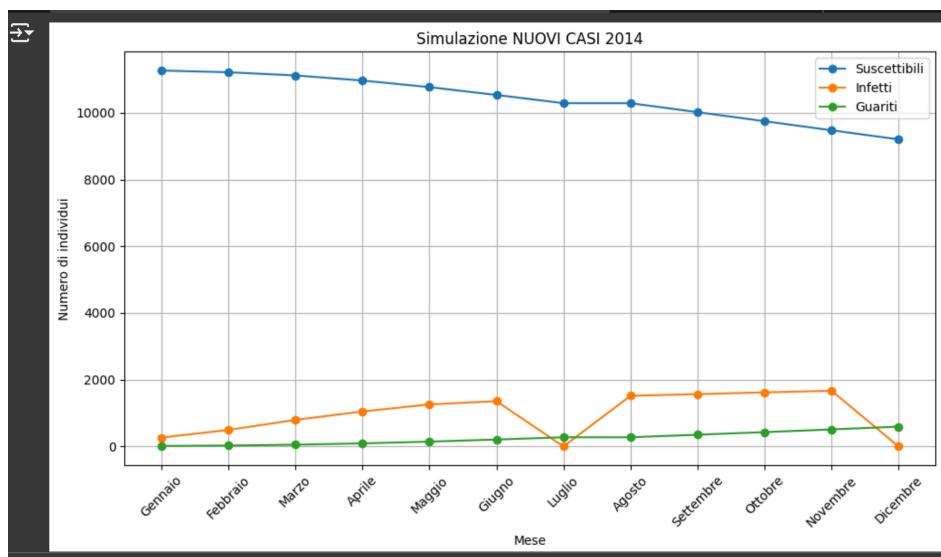


Fig.15: Esempio Simulazione SIR - Casi Nuovi

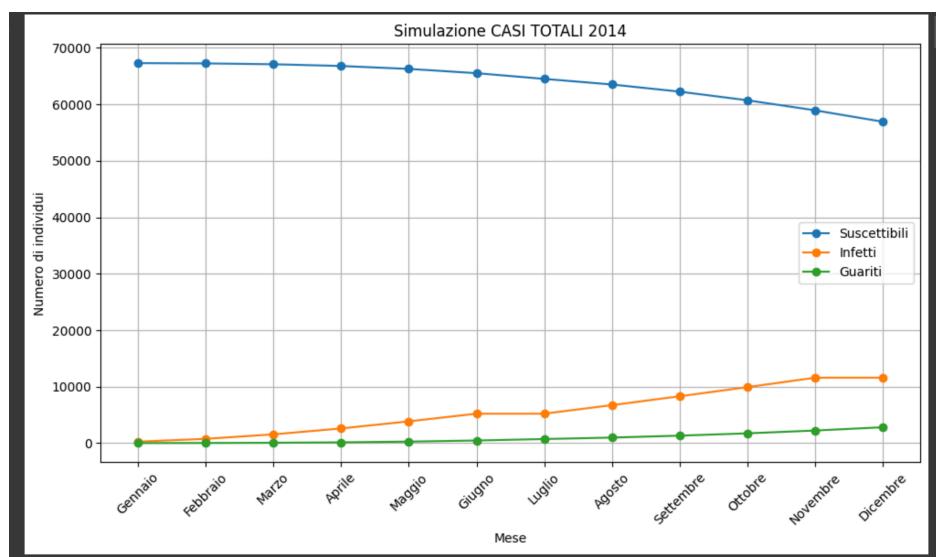


Fig.16: Esempio Simulazione SIR - Casi Totali

Analizzando gli andamenti mostrati come esempio nei grafici, possono essere messi a confronto con i dati reali, ottenuti dai dataset.

I casi nuovi del 2014 (Fig. 15) mostrano una crescita costante nei primi mesi dell'anno, raggiungendo un picco a giugno, per poi essere seguito da un calo drastico a luglio. Questo potrebbe esser dovuto a interventi di controllo o cambiamenti stagionale che influenzano la trasmissione del morbillo. Dopo il calo di luglio c'è una rapida ripresa ad agosto con casi che rimangono elevati per i mesi successivi fino a novembre, per poi notare un calo a dicembre. Questo potrebbe indicare un'onda epidemica stagionale che tende a diminuire con l'inizio della stagione fredda.

Per ottenere questi andamenti aumentando il parametro *beta* β si ha la probabilità di contagio molto alta.

Concludendo questo paragrafo, la simulazione è stata generata con modello random, ma l'esclusione dell'age_group, non può esser validata come considerazione. Valutando così la seconda simulazione.

Seconda simulazione:

In questa simulazione si includono le fasce d'età, andando ad analizzarle fascia per fascia.

Nel dettaglio, ci sono **5 classi d'età** :

- 0_4
- 5_14
- 15_39
- 40_64
- 65

Analizzare l'andamento dei casi di morbillo suddiviso per fasce d'età può fornire ulteriori informazioni dettagliate sulla diffusione della malattia e sull'efficacia delle misure di controllo.

Ad esempio, si può andare ad identificare quali fasce d'età hanno il maggior numero di casi, aiutando così a capire quali gruppi sono più vulnerabili al morbillo. Come i bambini piccoli potrebbero avere tassi di infezione più alti a causa dell'immunità non ancora sviluppata o di un contatto più stretto in ambienti come scuole e/o asili.

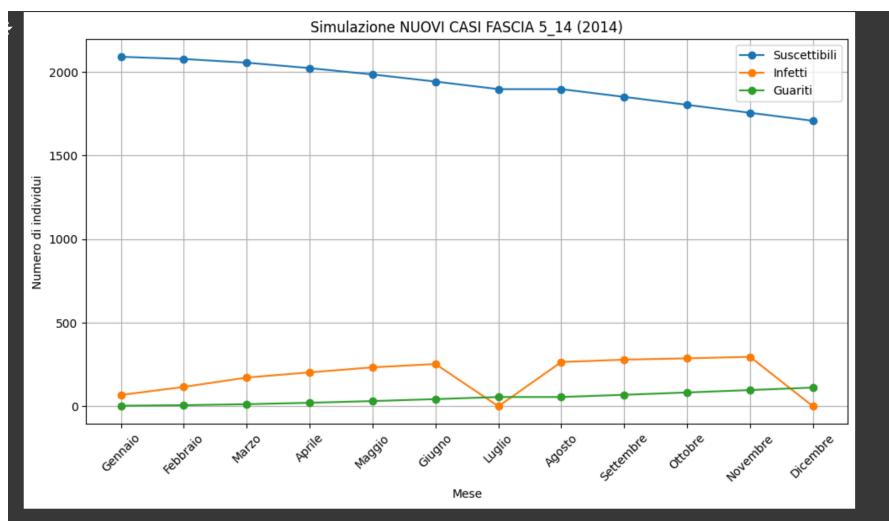
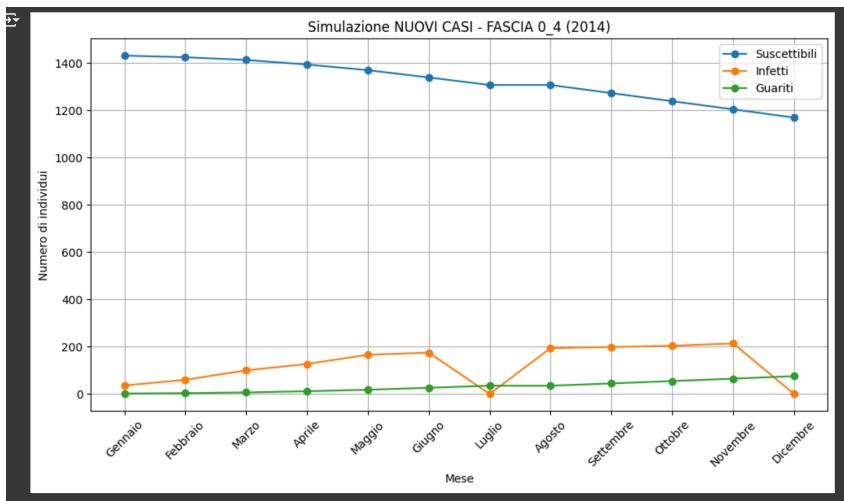
Inoltre analizzando i dati per fasce d'età si può valutare l'immunità di gregge in ciascuna fascia. Se una fascia d'età ha pochi casi, potrebbe indicare una buona copertura vaccinale o una precedente esposizione che ha conferito immunità.

Dunque nei modelli epidemiologici utilizzare dati per fasce d'età può migliorare la precisione delle previsioni.

Analizzando ad esempio tutte le fasce d'età del **2014**.

(*in riferimento ai nuovi casi*)

Otteniamo le seguenti simulazioni.



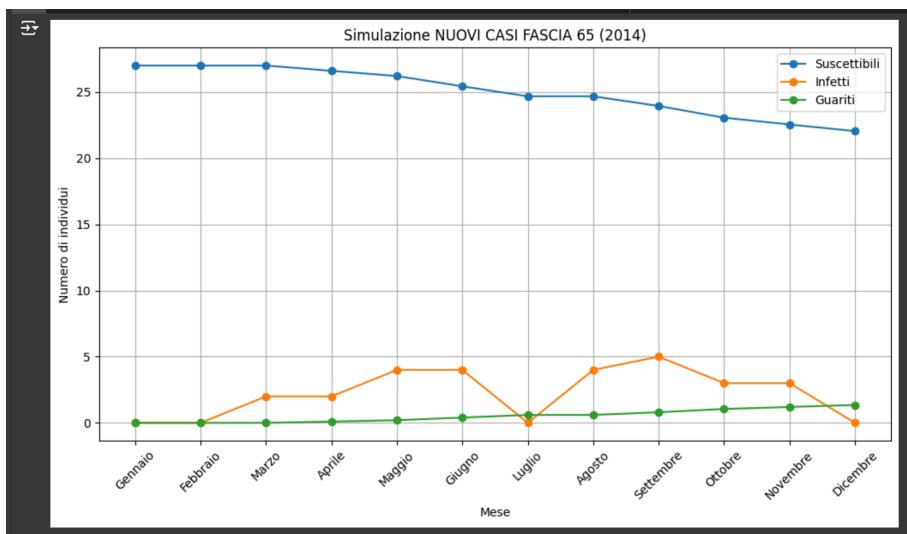
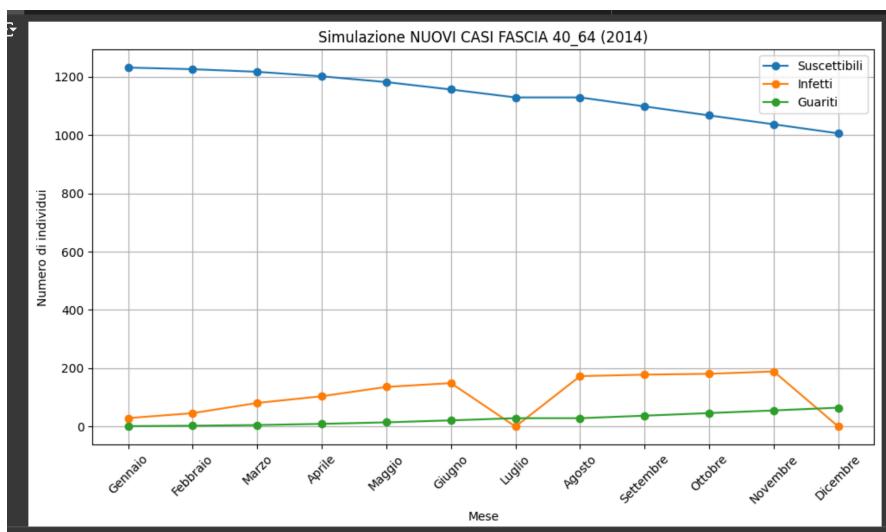
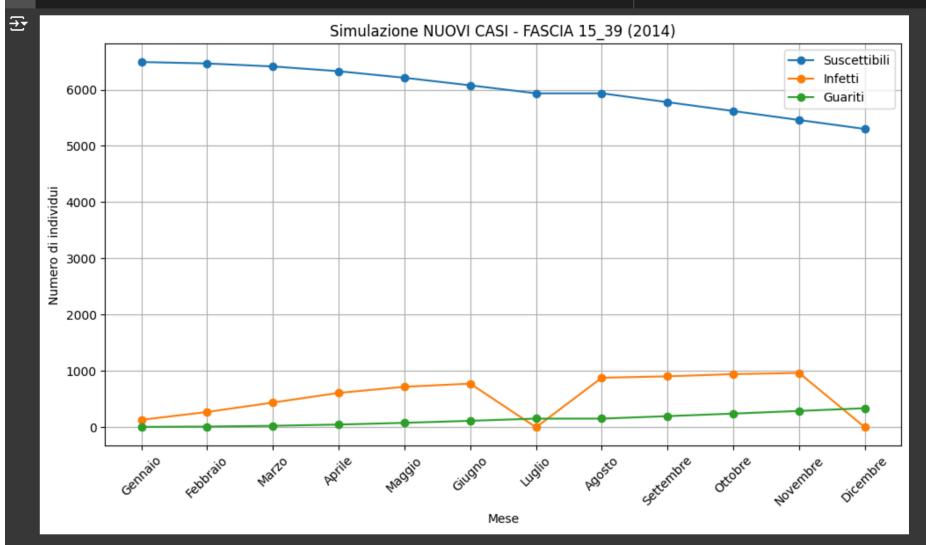


Fig.17: Esempi Simulazione SIR - Casi Nuovi

Volendo spiegare questi andamenti, nella *fascia 0_4*, ad esempio, si vede nei suscettibili (S) una diminuzione costante nel corso dell'anno. Ciò indica che il numero di bambini suscettibili al morbillo diminuisce man mano che alcuni di loro diventano infetti e altri guariti. La curva degli infetti (I) mostra un aumento graduale nei primi mesi, raggiungendo poi un picco a giugno. Successivamente, c'è un calo a luglio, seguito da un nuovo aumento ad agosto. Da settembre in poi, la curva si stabilizza con lievi variazioni fino a dicembre.

L'andamento potrebbe suggerire la possibilità di eventi stagionali o di cambiamenti che influenzano la trasmissione.

I guariti (R) mostrano un aumento graduale nel corso dell'anno. Ciò suggerisce che i bambini infetti stanno guarendo e sviluppando immunità contro il morbillo. Ma questo aumento non è molto rapido, indicando così che la guarigione richiede del tempo o che il numero di nuovi casi infetti non è molto alto rispetto al numero dei guariti.

Infine, facendo un confronto tra i dati reali e i risultati delle simulazioni, la similitudine tra le curve degli infetti della simulazione e quella reale del morbillo mostra che il modello SIR è in grado di rappresentare accuratamente l'andamento della malattia in una popolazione,

confermando la sua utilità e validità come strumento epidemiologico.

3.4 DISCUSSIONI

I risultati verrano illustrati e discussi.

Il *focus* è stato quello di contribuire a definire quale modello potesse rappresentare al meglio la trasmissibilità dei casi di Morbillo.

La misura della trasmissibilità è fondamentale per la valutazione del rischio di epidemie e pandemie, per la pianificazione delle risposte sanitarie e per lo sviluppo di strategie di controllo delle malattie, come vaccinazione, distanziamento e altre misure preventive.

CONCLUSIONI

Le conclusioni verranno analizzate da diverse prospettive e inoltre verranno rappresentate proposte delle riflessioni sui possibili futuri sviluppi correlati a questo studio.

I futuri sviluppi dello studio sarà fare analisi approfondita dei dati. Questo consentirà di capire nel dettaglio quale modello sia più adatto per studiare e comprendere la trasmissibilità del morbillo nella comunità. Portando a possibili miglioramenti e adattamenti su modelli esistenti.

Dal lavoro svolto, l'analisi mostra che il morbillo ha un comportamento stagionale, con una diffusione più intensa in alcuni mesi. Questo potrebbe essere influenzato da fattori ambientali, comportamentali o da interventi di salute pubblica. La capacità del Modello SIR di riflettere questi cambiamenti stagionali ne sottolinea l'utilità per la pianificazione di strategie di controllo e prevenzione.

In conclusione, l'andamento dei casi di morbillo e la simulazione del modello SIR mostrano una buona corrispondenza, indicando che il modello è efficace nel rappresentare la dinamica dell'infezione. Questo tipo di

analisi è fondamentale per comprendere e controllare la diffusione delle malattie infettive.

BIBLIOGRAFIA E SITOGRAFIA

- [1] Castellani G., Verondini E., Giampieri E., Milanesi L., Bersani F., Zironi I., Remondini D.: Large scale modelling of neuronal systems. Il Nuovo Cimento. Vol. 32 C, n. 2, March - April, 2009, p. 9-12.
- [2] Calabria A., Mosca E., Viti F., Merelli I., Milanesi L.: SNPRanker: a tool for identification and scoring of SNPs associated to target genes. J Integr Bioinform. 2010 Mar 25, Vol. 7, n. 3, doi: 10.2390/biecoll-jib-2010-138.PMID: 20375450.
- [3] ITALBIONET-ItalianBioinformaticsNetwork.
- [4] Milanesi, Luciano. "BIOINFORMATICA E CALCOLO AD ALTE PRESTAZIONI."
- [5] BARABÁSI, A.-L., 2004. Link: la scienza delle reti. Einaudi, Torino.
- [6] National Research Council, et al. *Network science*. National Academies Press, 2006.
- [7] Network Science di Albert-Laszlo Barabasi
- [8] Pietro Hiram Guzzi, Francesco Petrizzelli, Tommaso Mazza, Disease spreading modeling and analysis: a survey,

Briefings in Bioinformatics, Volume 23, Issue 4, July 2022,
bbac230, <https://doi.org/10.1093/bib/bbac230>

- [9] Domingos, P. (2005) Mining social networks for viral marketing. *IEEE Intell. Syst.*
- [10] Cohen, R., Havlin, S. & ben-Avraham, D. (2003) Efficient immunization strategies for computer networks and populations. *Pays. Rev. Lett.*, 91, 247901.
- [11] Pastor-Satorras,R.&Vespignani,A.(2002)Immunization ofcomplexnetworks.*Phys.Rev.E*,65,036104.
- [12] Pastor-Satorras, R. & Vespignani, A. (2001) Epidemic spreading in scale-free networks. *Phys. Rev. Lett.*, 86, 3200–3203.
- [13] Pastor-Satorras, R., Castellano, C., Van Mieghem, P. & Vespignani, A. (2015) Epidemic processes in complex networks. *Rev. Mod. Phys.*, 87, 925.
- [14] Nasell, I. (1996) The quasi-stationary distribution of the closed endemic sis model. *Adv. Appl. Probab.*, 28, 895–932.
- [15] Saito,K.,Nakano,R.&Kimura,M.(2008)Prediction of information diffusion probabilities for independent cascade model. *Knowledge-Based and Intelligent Information and Engineering Systems*. pp. 67–75.
- [16] Watts, D. J. (2002) A simple model of global cascades

on random networks. *Proc. Natl. Acad. Sci. U.S.A.*, 99, 5766–5771.

- [17] Schwämmle, V., Gonzälez, M. C., Moreira, A. A., Andrade, J. S. & Herrmann, H. J. (2007) Different topologies for a herding model of opinion. *Phys. Rev. E*, 75, 066108.
- [18] He, J., Ji, S., Beyah, R. & Cai, Z. (2014) Minimum-sized influential node set selection for social networks under the independent cascade model. *Mobile Ad Hoc Networking and Computing, ACM*. pp. 93–102
- [19] Raafat, R. M., Chater, N. & Frith, C. (2009) Herding in humans. *Trends Cognit. Sci.*, 13, 420–428.
- [20] Estrada, E. & Higham, D. J. (2010) Network properties revealed through matrix functions. *SIAM Rev.*, 52, 696–714.
- [21] Mantzaris, A. V., Bassett, D. S., Wymbs, N. F., Estrada, E., Porter, M. A., Mucha, P. J., Grafton, S. T. & Higham, D. J. (2013) Dynamic network centrality summarizes learning in the human brain. *J. Complex Netw.*, 1, 83–92.
- [22] Barabasi, A.-L. & Albert, R. (1999) Emergence of scaling in random networks. *Science*, 286, 5009–5012.
- [23] Hirsch, J. E. (2005) An index to quantify an

individual's scientific research output. *Proc. Natl. Acad. Sci. U.S.A.*, 102, 16569–16572.

[24] Askari Sichani, O. & Jalili, M. (2015) Influence maximization of informed agents in social networks. *Appl. Math. Comput.*, 254, 229–239.

[25] Chen, W., Wang, Y. & Yang, S. (2009) Efficient influence maximization in social networks. *International Conference on Knowledge Discovery and Data Mining*, ACM. pp. 199–208.

[26] Chen, W., Yuan, Y. & Zhang, L. (2010) Scalable influence maximization in social networks under the linear threshold model. *IEEE International Conference on Data Mining*, IEEE. pp. 88–97.

[27] Kempe, D., Kleinberg, J. & Tardos, E. (2003) Maximizing the spread of influence through a social network. *ACM Knowledge Discovery and Data Mining*, ACM. pp. 137–146.

[28] Seidman, S. B. (1983) Network structure and minimum degree. *Soc. Netw.*, 5, 269–287.

[29] Carmi, S., Havlin, S., Kirkpatrick, S., Shavitt, Y. & Shir, E. (2007) A model of Internet topology using k-shell decomposition. *Proc. Natl. Acad. Sci. U.S.A.*, 104, 11150–11154.

- [30] Barabasi, A.-L. & Albert, R. (1999) Emergence of scaling in random networks. *Science*, 286, 5009–5012.
- [31] Liu, Y., Tang, M., Zhou, T. & Do, Y. (2015) Core-like groups result in invalidation of identifying super-spreader by k-shell decomposition. *Sci. Rep.*, 5, 9602.
- [32] Wang, X., Zhang, X., Zhao, C. & Yi, D. (2016) Maximizing the spread of influence via generalized degree discount. *PLoS One*, 11, e0164393.
- [33] Lu, L., Chen, D., Ren, X.-L., Zhang, Q.-M., Zhang, Y.-C. & Zhou, T. (2016) Vital nodes identification in complex networks. *Phys. Rep.*, 650, 1–63.
- [34] de Arruda, G.F., Barbieri, A.L., Rodríguez, P.M., Rodrigues, F.A., Moreno, Y. & da Fontoura Costa, L. (2014) Role of centrality for the identification of influential spreaders in complex networks. *Phys. Rev. E*, 90, 032812.
- [35] Alzaabi, M., Taha, K. & Martin, A. T. (2015) CISRI: a crime investigation system using the relative importance of information spreaders in networks depicting criminals communications. *IEEE Trans. Inf. Foren. Sec.*, 10, 2196–2211.
- [36] Mahdi Jalili, Matjaž Perc, Information cascades in complex networks, *Journal of Complex Networks*, Volume

5, Issue 5, October 2017, Pages 665–693, <https://doi.org/10.1093/comnet/cnx019>

- [37] Stone, L., Olinky, R. & Huppert, A. (2007) Seasonal dynamics of recurrent epidemics. *Nature*, 446, 533–536.
- [38] Pastor-Satorras, R. & Vespignani, A. (2001) Epidemic dynamics and endemic states in complex networks. *Phys. Rev. E*, 63, 066117.
- [39] Khan, K., Arino, J., Hu, W., Raposo, P., Sears, J., Calderon, F., Heidebrecht, C., Liauw, J. & Gardam, M. (2009) Spread of a novel influenza A (H1N1) virus via global airline transportation, *New Engl. J. Med.*, 361, 212–214.
- [40] Cauchemez, S., Bhattarai, A., Marchbanks, T. L., Fagan, R. P., Ostroff, S., Ferguson, N. M. & Swerdlow, D. (2011) Role of social networks in shaping disease transmission during a community outbreak of 2009 H1N1 pandemic influenza. *Proc. Natl. Acad. Sci. U.S.A.*, 108, 2825–2830.
- [41] Brockmann, D. & Helbing, D. (2013) The hidden geometry of complex, network-driven contagion phenomena. *Science*, 342, 1337–1342.

[42] Christley, R. M., Pinchbeck, G. L., Bowers, R. G., Clancy, D., French, N. P., Bennett, R. & Turner, J. (2005) Infection in social networks: using network analysis to identify high-risk individuals. *Am. J. Epidemiol.*, 162, 1024–1031.

[43] Cohen, R., Havlin, S. & ben-Avraham, D. (2003) Efficient immunization strategies for computer networks and populations. *Phys. Rev. Lett.*, 91, 247901.

[44] Madar, N., Kalisky, T., Cohen, R., ben-Avraham, D. & Havlin, S. (2004) Immunization and epidemic dynamics in complex networks. *Eur. Phys. J. B*, 38, 269–276.

[4 5] S a l a t h e

'M.,Kazandjieva,M.,Lee,J.W.,Levis,P. ,Feldman,M.W.&Jones,J.H.(2010) Ahigh-resolution human contact network for infectious disease transmission. *Proc. Natl. Acad. Sci. U.S.A.*, 107, 22020–22025.

[46] Mirzasoleiman, B., Babaei, M. & Jalili, M. (2012) Immunizing complex networks with limited budget. *Europhys. Lett.*, 98, 38004.

[47] Domingos, P. (2005) Mining social networks for viral marketing. *IEEE Intell. Syst.*, 20, 80–82.

- [48] Leskovec, J., Adamic, L. A. & Huberman, B. A. (2006) The dynamics of viral marketing. *ACM Conference on Electronic Commerce*. Ann Arbor, MI: ACM, pp. 228–237.
- [49] Kempe,D.,Kleinberg,J.&Tardos,E .(2003)Maximizing the spreadofinfluencethroughasocialnetwork.
ACM Knowledge Discovery and Data Mining, ACM. pp. 137–146.
- [50] Braunstein, A., Dall'Asta, L., Semerjian, G. & Zdeborová, L. (2010) Network dismantling. *Proc. Natl. Acad. Sci. U.S.A.*, 113, 12368–12373.
- [51] Zdeborová, L., Zhang, P. & Zhou, H.-J. (2016) Fast and simple decycling and dismantling of networks. *Sci. Rep.*, 6, 37954.
- [52] Lorenz, J. (2007) Continuous opinion dynamics under bounded confidence: a survey. *Int. J. Mod. Phys. C*, 18, 1819–1838.
- [53] AskariSichani,O.&Jalili,M.(2015) Influence maximization of informedagentsinsocialnetworks.*Appl. Math. Comput.*, 254, 229–239.
- [54] Shuang,Q.,Zhang,M.&Yuan,Y.(2014) Node vulnerability of water distribution networks under cascading failures. *Reliab. Eng. Syst. Saf.*, 124, 132–141.

- [55] Yan, J., He, H. & Sun, Y. (2014) Integrated security analysis on cascading failure in complex networks. *IEEE Trans. Inf. Foren. Sec.*, 9, 451–463.
- [56] Ghanbari, R., Jalili, M. & Yu, X. (2016) Analysis of cascaded failures in power networks using maximum flow based complex network approach. *Annual Conference of the IEEE Industrial Electronics Society*, IEEE. pp. 4928–4932.
- [57] Ranshous, S., Shen, S., Koutra, D., Harenberg, S., Faloutsos, C. & Samatova, N. F. (2015) Anomaly detection in dynamic networks: a survey. *WIREs Comput. Statist.*, 7, 223–247.
- [58] Vos Fellman, P. & Wright, R. (2004) Modeling terrorist networks, complex systems at the mid-range. *Intelligencer*, 4.
- [59] Magalingam, P., Davis, S. & Rao, A. (2015) Using shortest path to discover criminal community. *Digit. Invest.*, 15, 1–17.
- [60] Yasami, Y. & Safaei, F. (2017) A statistical infinite feature cascade-based approach to anomaly detection for dynamic social networks. *Comput. Commun.*, 100, 52–64.
- [61] Das, K. & Sinha, S. K. (2016) Centrality measure based approach for detection of malicious nodes in twitter

social network. *International Conference on Advanced Material Technologies*.

[62] Kaur, R., Kaur, M. & Singh, S. (2016) A novel graph centrality based approach to analyze anomalous nodes with negative behavior. *Proc. Comput. Sci.*, 78, 556–562.

[63] salute.gov.it

[64] aifa.gov.it

[65] Measles second dose vaccine utilization and associated factors among children aged 24–35 months in Sub-Saharan Africa, a multi-level analysis from recent DHS surveys.

Dagmawi Chilot, Daniel Gashaneh Belay, Kegnie Shitu, Yibeltal Yismaw Gela, Mihret Getnet, Bezawit Mulat, Atalay Goshu Muluneh, Mehari Woldemariam Merid, Desalegn Anmut Bitew & Adugnaw Zeleke Alem

RINGRAZIAMENTI

Mi è doveroso dedicare questo spazio della mia tesi a tutte le persone che mi hanno supportato nel mio percorso di crescita universitaria e professionale.

Questo lavoro non è solo il risultato del mio impegno personale, ma anche del sostegno e della generosità di molte persone straordinarie.

Desidero ringraziare il Professore Pietro Hiram Guzzi, relatore di tesi, che mi ha seguito, passo dopo passo, con disponibilità e pazienza nella stesura di questo elaborato.

Un ringraziamento anche al Professore Pierangelo Veltri , sempre presente. Fornendomi preziosi consigli lungo tutto il percorso accademico.

Barbara e Raffaele grazie per il vostro instancabile supporto.

Non posso non ringraziare le due persone che hanno avuto più influenza nel mio percorso educativo: i miei genitori.

Ed infine, alla mia famiglia, al mio compagno di vita e di avventure, i miei amici e a tutti quelli che hanno incrociato la loro vita con la mia lasciandomi qualcosa di buono.

*“Esiste un solo bene, la conoscenza, ed un solo male,
l’ignoranza”*

Socrate