

Thomas Del Moro

29 aprile 2024

Sommario

Questo studio esplora l'applicazione di vari modelli di epidemie nell'ambito della fisica statistica e dei sistemi complessi, in particolare per quanto riguarda la dissuisione di malattie infettive in una popolazione o la percolazione di informazioni in una rete sociale. Attraverso l'applicazione di questi modelli, esaminiamo fattori come la densità di popolazione, i contatti fra gli individui e le strategie di mitigazione basate sulla percezione del rischio. Le simulazioni illustrano come tali modelli possano fornire una prospettiva efficace per comprendere e gestire fenomeni epidemici complessi.

1 Introduzione

La modellazione delle epidemie è un'importante area di ricerca che utilizza concetti della fisica statistica e dei sistemi complessi per comprendere e prevedere la diffusione delle malattie infettive. I modelli epidemiologici forniscono un quadro teorico per analizzare come le malattie si diffondono all'interno di una popolazione e come le strategie di controllo possono influenzare questa diffusione. Tra i modelli più conosciuti vi è il modello SIR (Susceptible-Infectious-Recovered), che suddivide la popolazione in tre compartimenti: suscettibili, infetti e guariti.

Questi modelli, tuttavia, spesso non tengono conto della complessità delle interazioni sociali e della struttura di rete all'interno delle popolazioni umane. È però possibile, utilizzando metodi ispirati alla fisica, come la teoria dei grafi e la dinamica dei sistemi fuori dall'equilibrio, sviluppare modelli più sofisticati che catturino meglio la natura intricata delle interazioni umane.

In questa revisione, esploreremo l'applicazione di vari modelli di epidemie e, attraverso simulazioni computerizzate, esamineremo come questi modelli possano essere utilizzati per comprendere e gestire fenomeni epidemici complessi.

2 Forest Fire

Il modello più semplice di un'epidemia trascura completamente la dinamica dell'agente virale. Se supponiamo inoltre che gli ospiti non si muovano e siano disposti su una griglia regolare con una certa densità. Otteniamo in questo modo il modello di diffusione di un incendio boschivo senza vento (Forest Fire).

Ogni cella della griglia può essere in uno di tre stati: vuota, occupata da un albero sano o occupata da un albero incendiato. Per semplicità abbiamo trascurato il caso di albero già bruciato (immune), che vedremo nella prossima sezione.

Il processo di diffusione segue una curva a S, più o meno ripida e lunga a seconda del parametro di densità. In particolare, se la densità è troppo bassa, l'incendio si estinguerà rapidamente, mentre se questa è alta l'incendio si propagherà a tutto il reticolo.

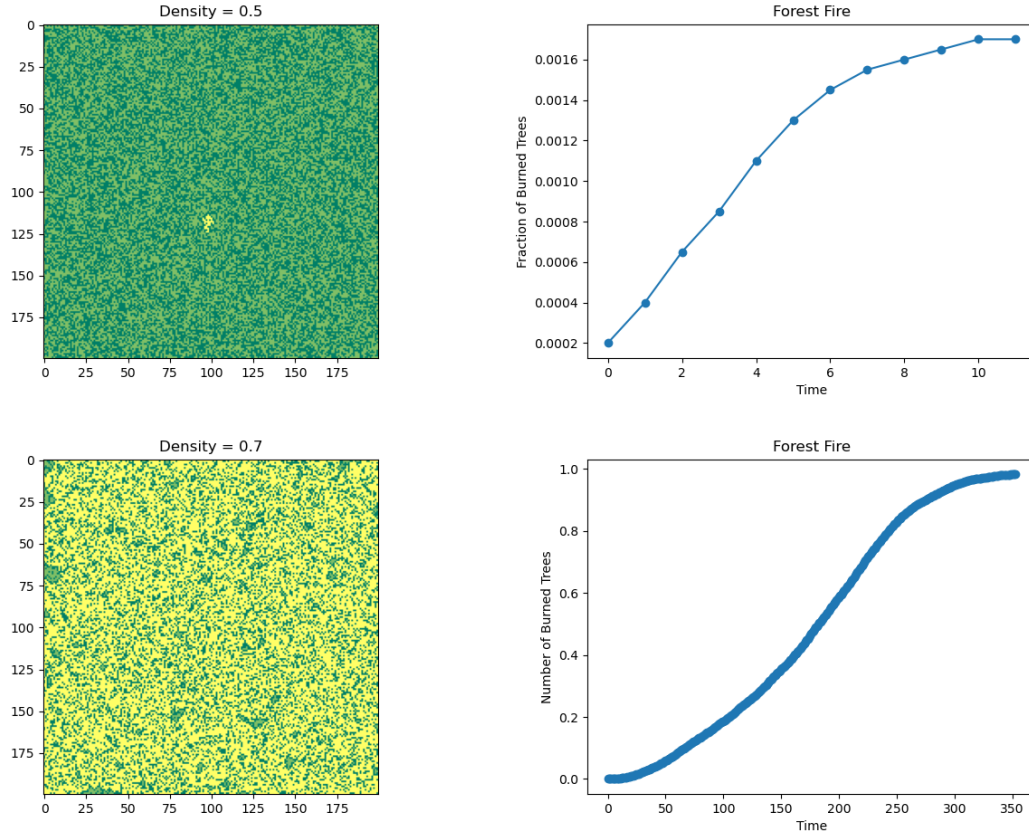


Figura 1: Simulazioni del modello Forest Fire con densità 0.5 (in alto) e 0.7 (in basso). Sulla sinistra si possono osservare le immagini del reticolo, mentre sulla destra la curva di diffusione dell'incendio.

Notiamo facilmente che se la densità è relativamente bassa, è probabile che l'incendio si estingua velocemente. Possiamo quindi osservare il numero finale di alberi incendiati al variare della densità del reticolo. I risultati, mostrati nel grafico in Figura 2, evidenziano di nuovo una curva a S, con un punto di transizione tra un regime in cui l'incendio si estingue rapidamente e uno in cui si propaga a tutto il reticolo.

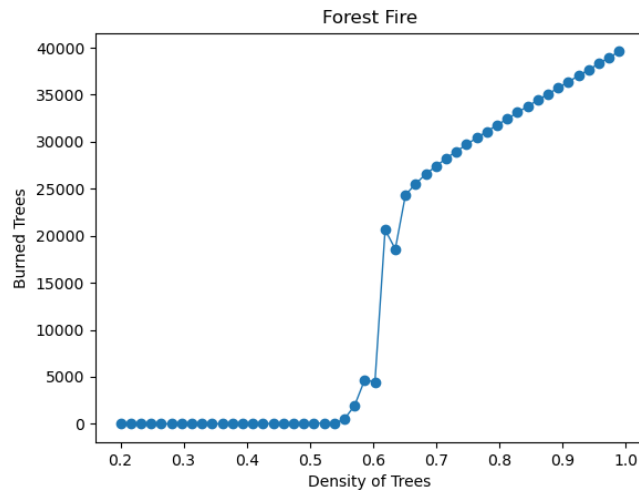


Figura 2: Numero finale di alberi incendiati al variare della densità del reticolo.

3 SIR

Il modello SIR rappresenta uno dei più conosciuti e più semplici modelli epidemiologici. È in parte analogo al modello Forest Fire, ma con una dinamica probabilistica, leggermente più complessa di quella vista sopra.

Supponiamo di avere una popolazione di N individui disposti su una griglia. In questo caso tutte le celle sono occupate da un individuo, ma questo, ad ogni istante di tempo, sarà infettato dagli individui vicini solo con una certa probabilità p . Le celle possono dunque essere suddivise nelle tre categorie:

- **Suscettibili (S)**: individui sani che possono essere infettati.
- **Infetti (I)**: individui infetti che possono infettare gli individui sani.
- **Refrattari (R)**: individui infettati e guariti che non possono più ammalarsi (immuni).

Per eseguire le simulazioni possiamo inizializzare il reticolo con un numero arbitrario di infetti e osservare come la malattia si diffonde a seconda dei parametri di infettività e di guarigione, fino a che non sarà più in grado di infettare nuovi individui.

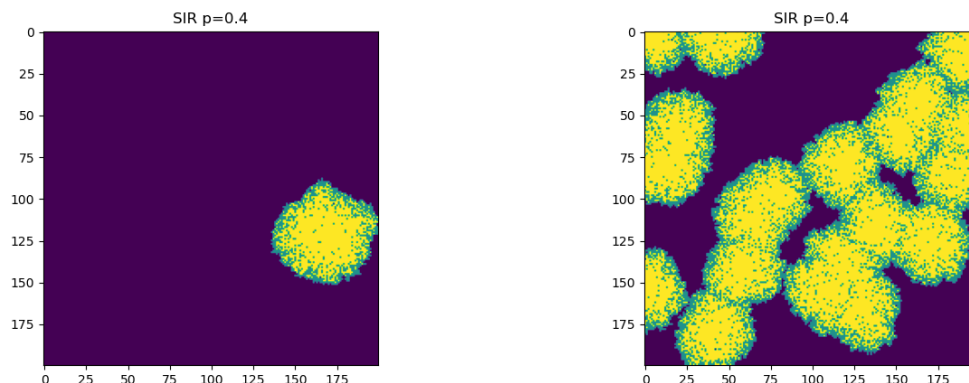
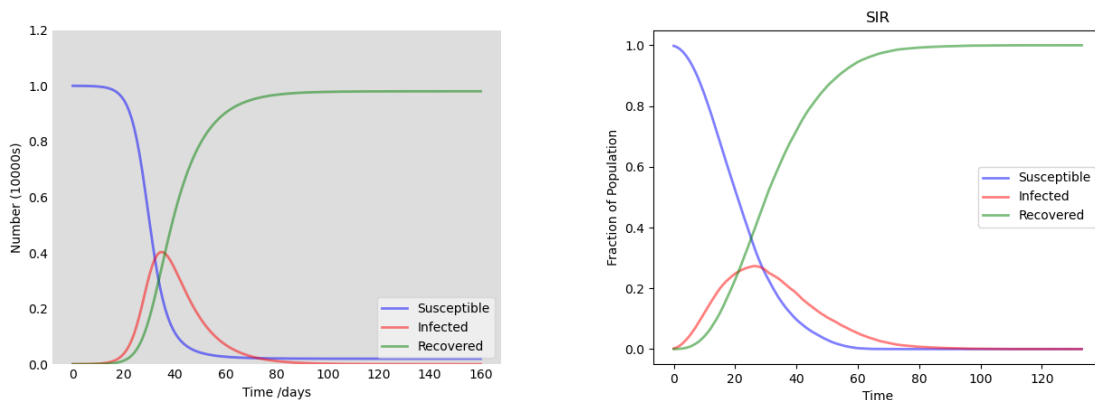


Figura 3: Simulazioni del modello SIR rispettivamente a partire da 1 infetto (sinistra) e da 20 infetti (destra).

Le curve caratteristiche del modello SIR possono essere ottenute applicando le assunzioni di campo medio, ovvero eliminando le correlazioni spaziali tra gli individui. In questa relazione ci limiteremo a confrontare le curve teoriche con quelle empiriche ottenute dalle simulazioni sopra descritte. Tale indagine è mostrata in Figura 4.



(a) Curve teoriche del modello SIR

(b) Curve empiriche di una simulazione del modello SIR

Figura 4: Curve caratteristiche teoriche ed empiriche del modello SIR.

Il fattore più critico per la diffusione della malattia è la probabilità di infezione p . Nel modello molto semplice abbiamo realizzato, l'infezione si diffonde molto probabilmente in tutto il reticolo anche con un livello di infettività relativamente basso: Ciò che possiamo osservare è quindi una diversa velocità di propagazione dell'epidemia a seconda di questo parametro.

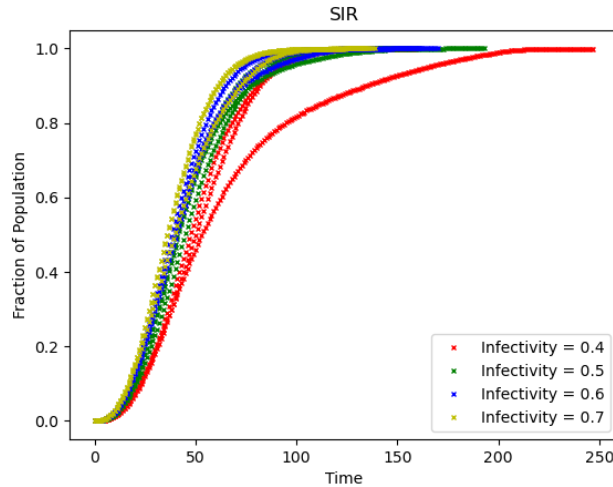


Figura 5: Curve di diffusione di più simulazioni al variare della probabilità di infezione.

Possiamo notare in Figura 5 che all'aumentare della probabilità di infezione, la curva di individui refrattari, ovvero infettati e guariti, si sposta verso sinistra, indicando una maggiore velocità di propagazione dell'epidemia.

Un ulteriore fattore critico oltre ai due già descritti è dato dal numero di vicini scelti per ogni individuo. In una tale griglia è possibile scegliere di considerare solo i primi 4 vicini (alto, basso, destra, sinistra) oppure tutte le 8 celle adiacenti. In questo ultimo caso ci aspettiamo che la soglia di infettività sia più bassa, in quanto la probabilità di trovarsi vicino a un individuo suscettibile è maggiore. Possiamo infatti notare in Figura 6 che la curva relativa agli 8 vicini è più ripida rispetto a quella relativa agli 4 vicini.

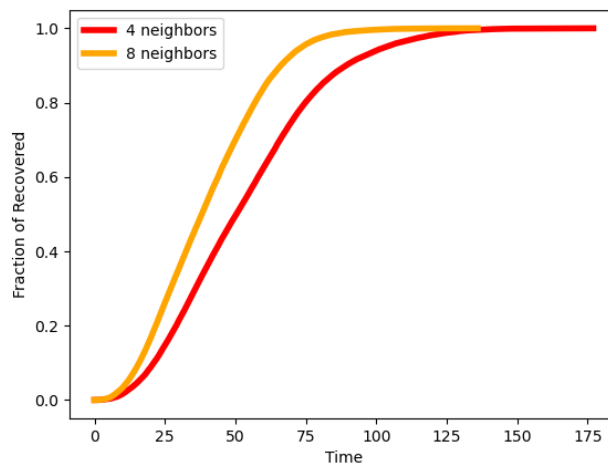


Figura 6: Frazione di individui refrattari relativi a una simulazione con $p = 0.8$

Questo modello molto semplice non tiene ovviamente conto di caratteristiche molto più complesse della malattia e della popolazione, come la mobilità degli individui, la densità di popolazione, la distribuzione dell'età e la presenza di misure di mitigazione come il distanziamento sociale e la quarantena.

Tuttavia, fornisce una base teorica per comprendere la diffusione delle malattie infettive e come le strategie di controllo possono influenzare questa diffusione.

4 SIS

Il modello SIR assume che gli individui guariti acquisiscano immunità permanente alla malattia. Tuttavia, se ciò non valesse, esistono modelli alternativi che considerano una certa probabilità di reinfezione. Il caso all'estremo opposto è infatti il modello SIS (Susceptible-Infectious-Susceptible), in cui gli individui guariti tornano a essere suscettibili in egual maniera rispetto ai sani. Questo modello è particolarmente interessante per le malattie in cui l'immunità non è permanente, come l'influenza stagionale o il raffreddore comune.

Anche per questo modello possiamo eseguire delle simulazioni analoghe alle precedenti, utilizzando lo stesso reticolo ma assumendo che gli individui guariti tornino a essere suscettibili.

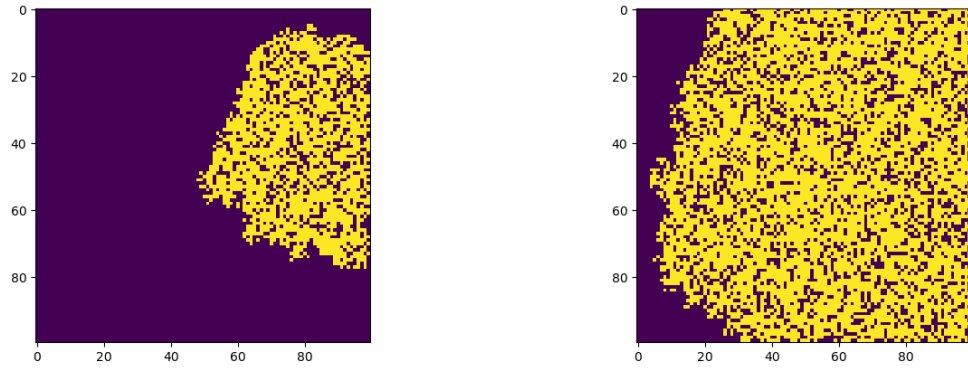


Figura 7: Simulazione del modello SIS con infettività $p = 0.2$ e tasso di guarigione $g = 0.5$.

In questo caso le curve saranno diverse da quelle del modello SIR, poiché non abbiamo una frazione di refrattari. A seconda del grado di infettività e di guarigione è possibile osservare andamenti diversi tra loro, ma che hanno tutti la caratteristica che avere un andamento asintotico nel lungo termine simile a una saturazione. In Figura 8 sono mostrati i risultati ottenuti con diversi valori dei parametri di infettività e di guarigione.

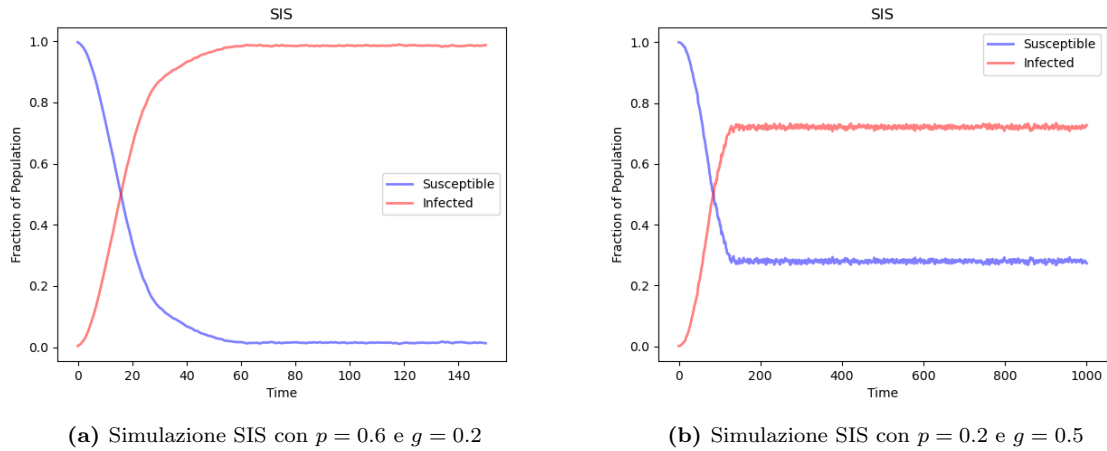


Figura 8: Curve empiriche di due simulazioni del modello SIS.

5 Percolazione

Il modello SIR è analogo alla crescita di una componente di percolazione standard connesso al sito di partenza. La percolazione standard è un modello fondamentale della teoria dei sistemi complessi che studia la formazione di un percorso connesso attraverso una rete in funzione di una certa probabilità di occupazione dei siti nella rete stessa. Il concetto chiave della percolazione standard è la ricerca di un percorso connesso o *cluster* che attraversa la rete da un lato all'altro. Essa studia quindi come la presenza di occupazioni casuali nei siti influenzi la probabilità di avere un percorso connesso che attraversi l'intera rete. La percolazione può essere utilizzata per comprendere fenomeni come la conduttività elettrica dei materiali disordinati, la diffusione di fluidi attraverso mezzi porosi e la propagazione di informazioni o malattie attraverso reti sociali.

In questo senso, dunque, è possibile costruire un modello di percolazione che rappresenti la diffusione dell'epidemia all'interno di una popolazione. In questo modello, i siti della rete rappresentano gli individui della popolazione e i legami tra i siti rappresentano i contatti sociali tra gli individui. La probabilità di occupazione di un sito dipende da dei numeri casuali e corrisponde alla probabilità che l'individuo sia infetto. Ciò che cerchiamo è dunque un percorso connesso di siti infetti che attraversi interamente la rete, indicando una diffusione dell'epidemia in tutta la popolazione.

Vediamo più in dettaglio il modello fin qui descritto, in particolare riferendoci a un sottoinsieme di problemi di percolazione detto *Site Percolation*. Ad ogni sito può essere associato un valore binario x_i , che rappresenta lo stato del sito: $x_i = 1$ se il sito è infetto, $x_i = 0$ altrimenti. La probabilità di infezione di un sito dipende dal parametro di infettività p e da un numero casuale r_i associato al sito. Inseriamo infine una variabile temporale t che ci servirà per descrivere l'evoluzione dell'epidemia. Possiamo definire l'evoluzione dello stato del sito i al tempo t come:

$$x_i(t) = x_i(t-1) \vee [r_i < p] \left[\sum_j a_{ij} x_j(t-1) > 0 \right]$$

dove a_{ij} è la matrice di adiacenza del grafo, ovvero $a_{ij} = 1$ se i siti i e j sono vicini, 0 altrimenti; mentre $[\cdot]$ è la funzione di verità che restituisce 1 se la condizione tra parentesi è vera, 0 altrimenti. Notiamo che i numeri casuali r_i rimangono costanti nel tempo, dunque per un certo valore di p e una data configurazione iniziale i cluster crescono fino a un certo punto e poi si arrestano.

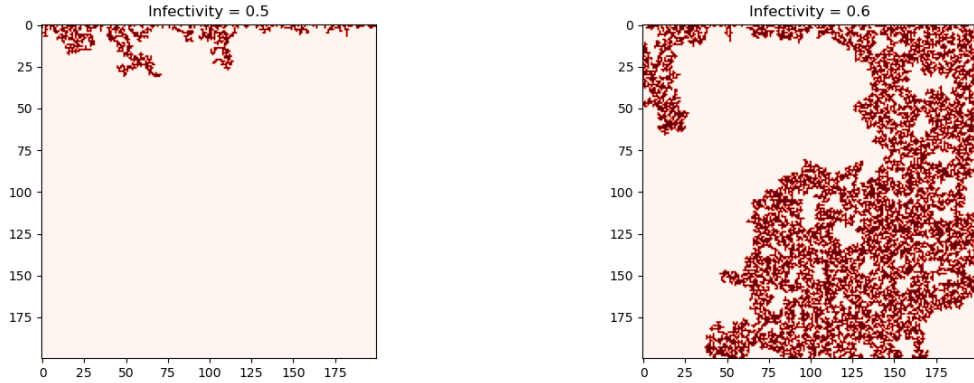


Figura 9: Simulazioni del modello di percolazione con diversi gradi di infettività.

In Figura 9 sono mostrate due simulazioni del modello di percolazione. Nel reticolo sulla sinistra, con una probabilità di infezione $p = 0.5$, non è possibile trovare un percorso connesso di siti infetti che attraversi l'intera rete. Al contrario, nel reticolo sulla destra, con $p = 0.6$, si può osservare un caso di percolazione, ovvero è presente un percorso connesso che attraversa tutta la griglia. Questo indica che l'epidemia si è diffusa in tutta la popolazione.

Può dunque essere interessante cercare la criticità dell'infezione, ovvero il valore di soglia che separa i due regimi di percolazione e non percolazione.

5.1 Criticità

La soglia di criticità, che indichiamo con p_c , può essere stimata attraverso diversi metodi; in questo studio analizzeremo prima di tutto il metodo più semplice, ovvero quello di eseguire molte simulazioni modificando ogni volta il parametro di infettività. Dopodiché utilizzeremo un metodo più sofisticato, basato sulla formulazione SOC (Self-Organized Criticality), che permette di ottenere una stima affidabile eseguendo una sola simulazione.

Vediamo in questa sezione i risultati ottenuti con il metodo più semplice. Costruiamo più volte un reticolo analogo ai precedenti, in cui si inizializzano come infetti tutti i siti sulla prima riga. Ad ogni step, quindi, ogni sito sano sarà infettato da un vicino infetto con una certa probabilità p . Modifichiamo il valore di p a ogni simulazione in modo da ottenere una curva di percolazione.

Figura 10: Percentuale di percolazioni rispetto al numero di simulazioni, al variare del valore di infettività. Valore di soglia teorico $p_c = 0.5927$

Come si può osservare in Figura 10, la curva assume una caratteristica forma a S e rispetta il valore di soglia teorico per un reticolo 2D pari a $p_c = 0.5927$. Nonostante si riesca a trovare una stima affidabile di criticità, questo metodo è computazionalmente molto costoso, può essere vantaggioso utilizzare metodi più efficienti come quello che mostreremo di seguito.

5.2 Percolazione Auto-Organizzata (SOC)

La formulazione SOC (Self-Organized Criticality) è utilizzata in diversi contesti e può essere applicata nel nostro caso per ottenere una stima affidabile del valore di soglia di percolazione. Ciò che è necessario inizialmente è mappare il nostro problema di percolazione in un SOC equivalente e, per farlo, possiamo utilizzare un metodo chiamato *Invasion Percolation*.

Vediamo prima la formulazione matematica del problema, dopodiché passiamo alla progettazione pratica del modello. Supponiamo di avere un reticolo 2D e una certa probabilità di infezione p . Come nel caso precedente, a ogni sito associamo uno stato binario x_i e un numero casuale r_i . Definiamo inoltre una variabile t che rappresenta il i diversi istanti di tempo. Ricordiamo che l'evoluzione dello stato del sito i al tempo t è data da:

$$x_i(t) = x_i(t-1) \vee [r_i < p] \left[\sum_j a_{ij} x_j(t-1) > 0 \right].$$

Riscriviamo per semplicità il secondo termine:

$$x_i(t) = x_i(t-1) \vee \left([r_i < p] \wedge \left(\bigvee_{j:a_{ij}=1} x_j(t) \right) \right).$$

Possiamo sostituire lo stato $x_i(t)$ con $[p > p_i(t)]$ poiché il sito si bagna (si infetta) solo se la probabilità di infezione è maggiore della sua soglia critica $p_i(t)$. Allora l'espressione diventa:

$$[p_i(t+1) < p] = [p_i(t) < p] \vee \left([r_i < p] \left(\bigvee_{j:a_{ij}=1} [p_j(t) < p] \right) \right).$$

In generale, sappiamo che valgono $[a < p] \wedge [b < p] = [\max(a, b) < p]$ e $[a < p] \vee [b < p] = [\min(a, b) < p]$. Allora la nostra espressione diventa:

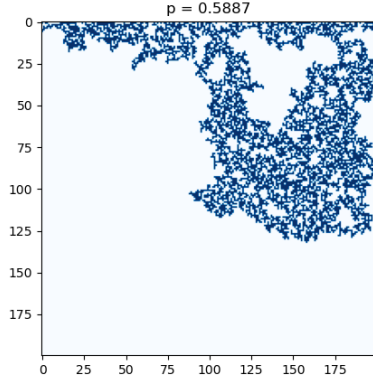
$$[p_i(t+1) < p] = \left[\min \left(p_i(t), \max \left(r_i, \min_{j:a_{ij}=1} (p_j(t)) \right) \right) < p \right].$$

A questo punto diventa semplice estrarre un'equazione per la soglia di criticità p_i :

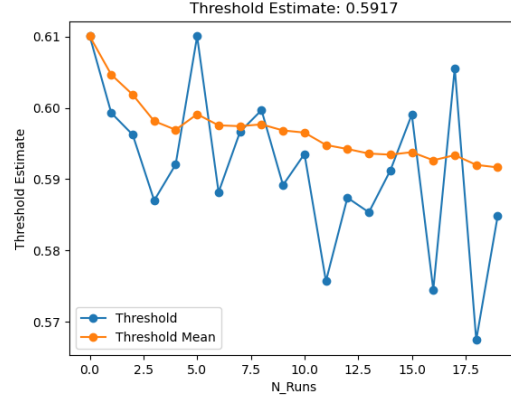
$$p_i(t+1) = \min \left(p_i(t), \max \left(r_i, \min_{j:a_{ij}=1} (p_j(t)) \right) \right).$$

Iterando finché non si raggiunge un fenomeno di percolazione, possiamo ottenere una distribuzione di p_i che tende a p_c .

Vediamo adesso come implementare questo metodo all'interno del nostro reticolo. Si può pensare di inizializzare tutti i siti della prima riga come infetti. La probabilità di infezione p può essere inizializzata a zero e poi aumentata ad ogni step dell'algoritmo. Ad ogni iterazione, si cerca, tra tutti i vicini degli infetti, il sito i con la soglia di criticità p_i minore; si aumenta dunque l'infettività a $p = p_i$ e si infetta tale sito, dopodiché di bagnano anche tutti i nodi adiacenti j che hanno $p_j < p$. Quando non ci sono più nodi infettabili, si esegue una nuova ricerca del minimo e il processo viene ripetuto finché non si raggiunge uno stato di percolazione del reticolo.



(a) Immagine di una simulazione sul reticolo.



(b) Stime della soglia di criticità ottenute con più simulazioni.

Figura 11: Simulazioni del modello di percolazione SOC per la ricerca della soglia di criticità.

Con il metodo descritto è possibile ottenere una stima abbastanza affidabile del valore di soglia di percolazione con un'unica simulazione. In Figura 11 (destra) sono mostrati i risultati ottenuti a partire da una sola simulazione fino a 20. Notiamo che eseguendo il processo più volte si ottiene una stima sempre più precisa, ma già inizialmente il valore ottenuto è molto vicino a quello teorico di $p_c = 0.5927$.

Come si può facilmente intuire, il metodo SOC è molto più efficiente rispetto a quello delle molte simulazioni; anche eseguendolo più volte infatti, il tempo di calcolo è molto inferiore rispetto al metodo precedente.

6 Conclusioni

Riferimenti bibliografici

- [1] Matteo Lapucci, Davide Pucci (2023) *Mixed-integer quadratic programming reformulations of multi-task learning models*
- [2] P. S. Bradley, K. P. Bennett, A. Demiriz (2000) *Constrained K-Means Clustering*
- [3] Mikko I. Malinen, Pasi Fränti (2014) *Balanced K-Means for Clustering*