

STN

Spreading Tweets News Project

Sistemi complessi: Modelli e Simulazione

Report finale di:

Lidia Alecci - 852501

Andrea Guzzo - 761818

Vittorio Maggio - 817034

Manuel Zanaboni - 816105

Anno accademico: 2019-2020

Agenda

Introduzione



1. Introduzione
2. Soluzione proposta
3. Risultati e simulazioni
4. Conclusione e sviluppi futuri

Domande di ricerca

Introduzione



- Studio della diffusione di contenuti all'interno di una rete sociale (Twitter).
- Quante persone visualizzano una notizia all'interno di una rete? È possibile tracciare l'evoluzione?
- Come dei contenuti si diffondono in base alla configurazione della rete e al tempo di esposizione?

Soluzione proposta

Introduzione



- Costruzione di una rete sociale (basandoci sui dati di Twitter)
- Classificazione dei nodi in base agli utenti (Opinion Leader, Bot, Utenti tradizionali)
- Sistema di simulazione multi agente basato su SOIL e modelli epidemiologici

Obiettivo Principale

Studio del migliore posizionamento dei nodi (bot) per
massimizzare la diffusione

Fasi dello studio

Introduzione



1 Download dei dati e creazione del grafo

Grafo orientato rispetto la relazione di "follow" su due gradi di profondità

4 Analisi risultati

Analisi della simulazione e principali statistiche sulla diffusione della notizia

2 Posizionamento dei Bot e simulazione SOIL

Analisi del grafo per scelta del posizionamento dei BOT ed esecuzione simulazione

3 Visualizzazione grafo e statistiche

Visualizzazione, statistiche e informazioni sui risultati



Strumenti e librerie

Soluzione proposta



Per lo sviluppo della soluzione abbiamo utilizzato i seguenti strumenti

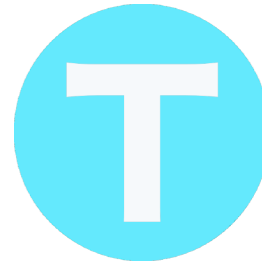
SOIL Simulatore



Streamlit



Twint



Plotly



Python (Pandas, ...)



NetworkX



Repository di progetto: <https://github.com/JeyDi/STN>

Download dei dati, creazione del grafo

Soluzione proposta



- Scraping da Twitter: @GiuseppeConteIT (followers)
- Grafo orientato in base alla relazione di «follow» (NetworkX)
- Diversi grafi considerati
(numero di followers)

Conte followers - STEP 4



Posizionamento dei Bot

Soluzione proposta



Misure di centralità

- **In-Degree:** numero di archi entranti in un nodo
- **Betweenness:** numero di volte che un nodo è un ponte lungo il percorso più breve tra 2 nodi
- **Autovettori:** misura dell'importanza di un nodo

Misure per il posizionamento

- **Random**
- **Top Betweenness:** su nodi che fungono da ponte sui principali percorsi di comunicazione
- **Top Eigenvector:** nodi «importanti»

Agenti

Soluzione proposta



3 Tipi di agenti

- **Opinion Leader:** fonte dell'informazione principale
- **Bot:** diffusori di notizie (automatici e non)
- **Utente:** utente comune

Stati degli utenti

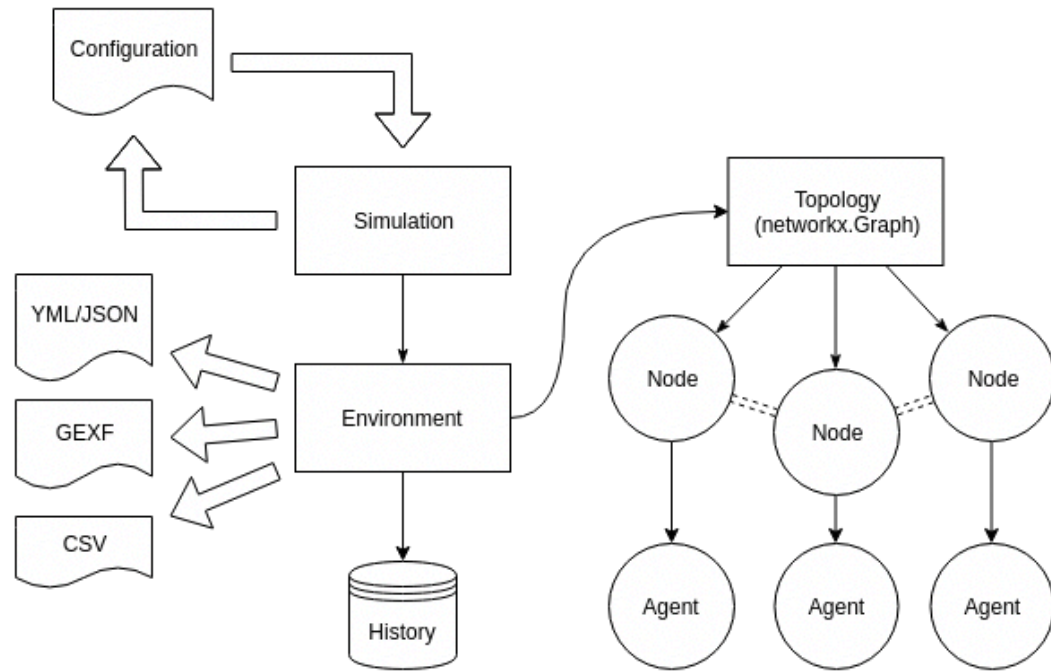
- **Non esposto:** non a contatto con la notizia
- **Esposto:** contatto con la notizia
- **Infetto:** contatto con la notizia e condivisa

Modello

Soluzione proposta



SOIL



- **Agenti**
- **Simulazione**
- **Topologia**
- **SEI Model**
- **Configurazione (probabilità)**
- **GEXF**

Web app

Soluzione proposta



The screenshot shows a 'Configuration Panel' with a close button (X) in the top right corner. It contains the following sections:

- Set parameters for the twitter scraper**
 - Twitter username:
 - ☒ Store the informations?
 - Output result file path:
 - Level of graph distance: with minus and plus buttons.
- Launch the twitter scraper** (button)

STN Project

Simulate news spreading info over a social network

Using data gathered from twitter and simulated

Made with Streamlit

Streamlit app

- Dashboard web configurabile
- Facilità nel creare e lanciare esperimenti
- Interpretazione dei risultati più veloce
- Utilizzabile su un server più potente

Assunzioni e grafi

Risultati



- Opinion leader ha influenza maggiore sui Bot
- I Bot hanno un'influenza più sparsa (ampia)
- Probabilità di infezione tra utente è molto bassa
- Probabilità di azione di un utente (retweet, commento) è bassa

Grafi considerati:

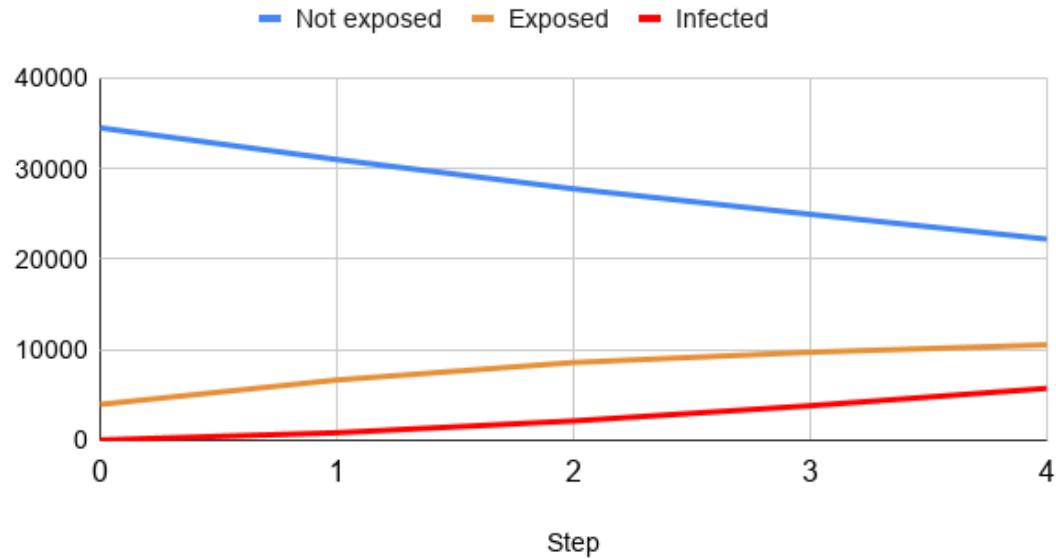
Followers	Profondità	Totale utenti
500	2	38458
1000	2	105774
1500	2	127700
2000	2	215910

Simulazioni: Random (500)

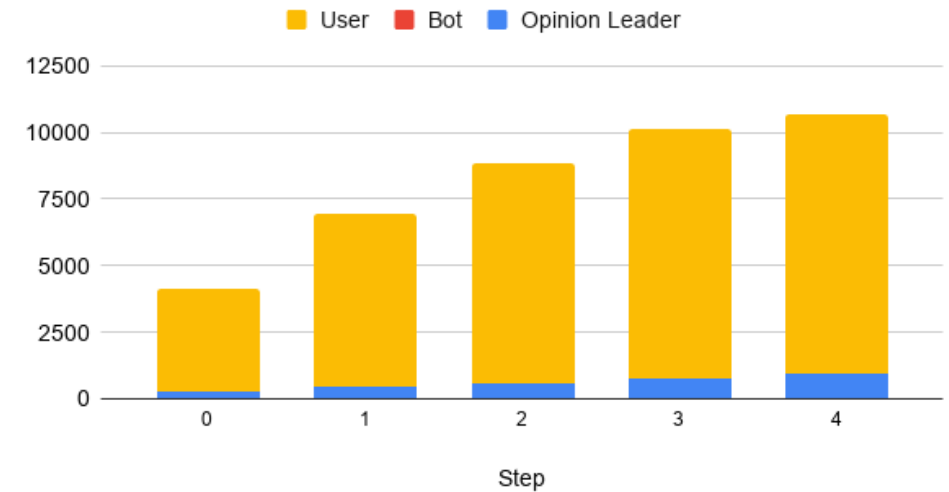
Risultati



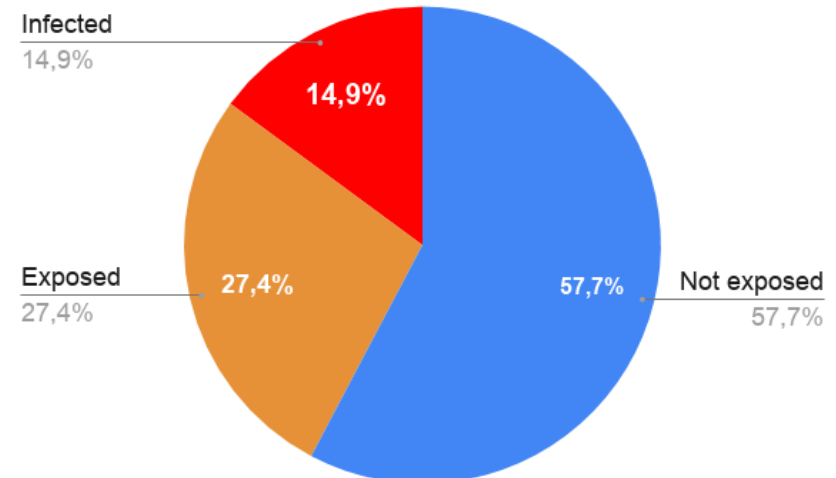
Infection overall: random_500



Type of exposure: random_500



Bot: random
Grafo: 500

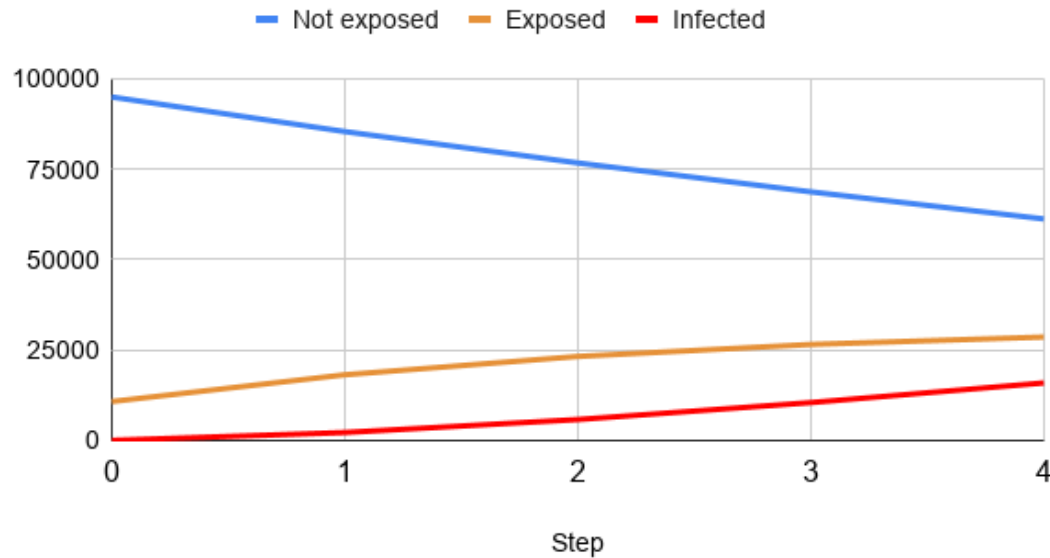


Simulazioni: Random (1000)

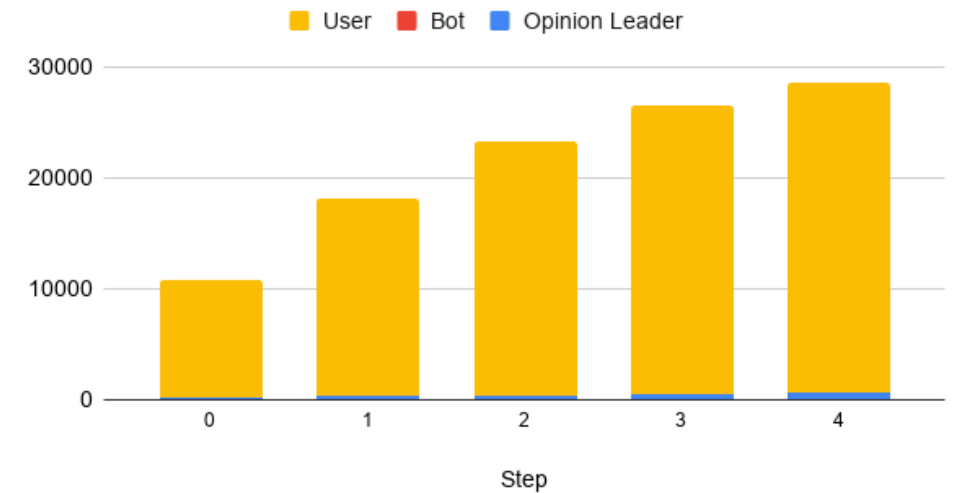
Risultati



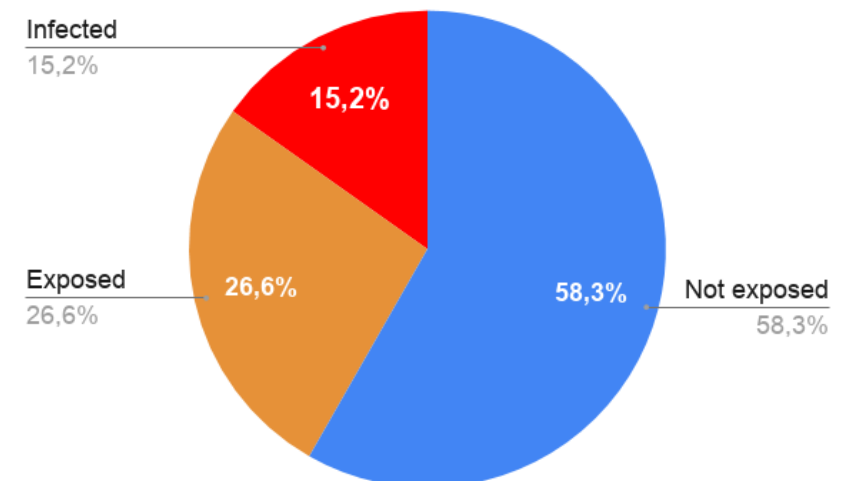
Infection overall: random_1000



Type of exposure: random_1000



Bot: random
Grafo: 1000

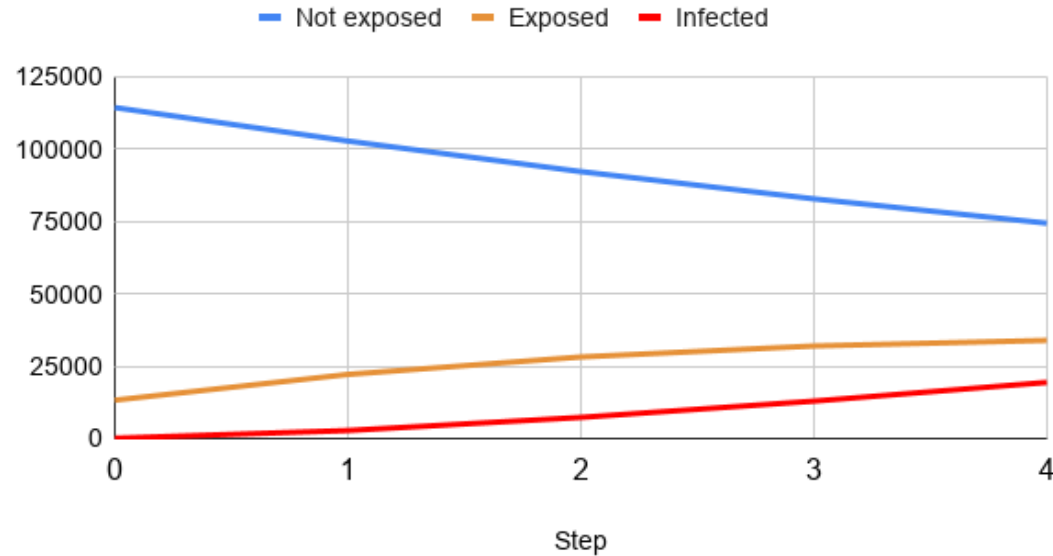


Simulazioni: Random (1500)

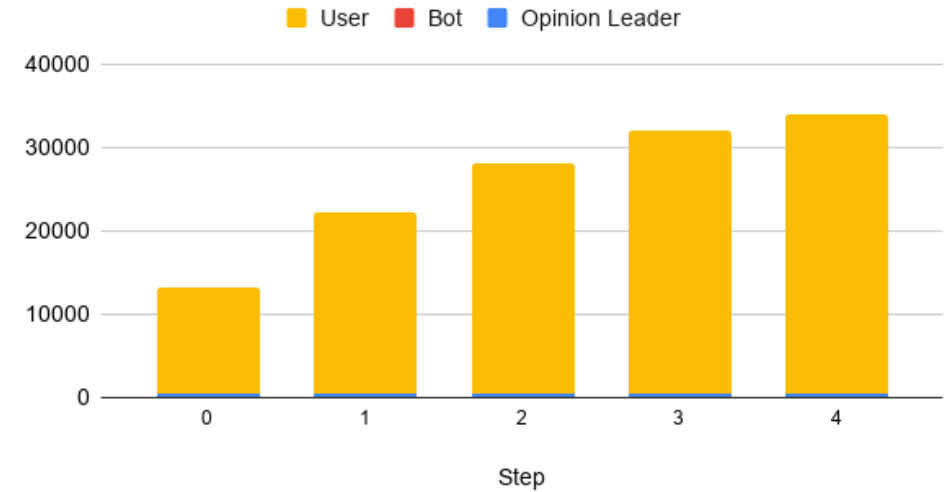
Risultati



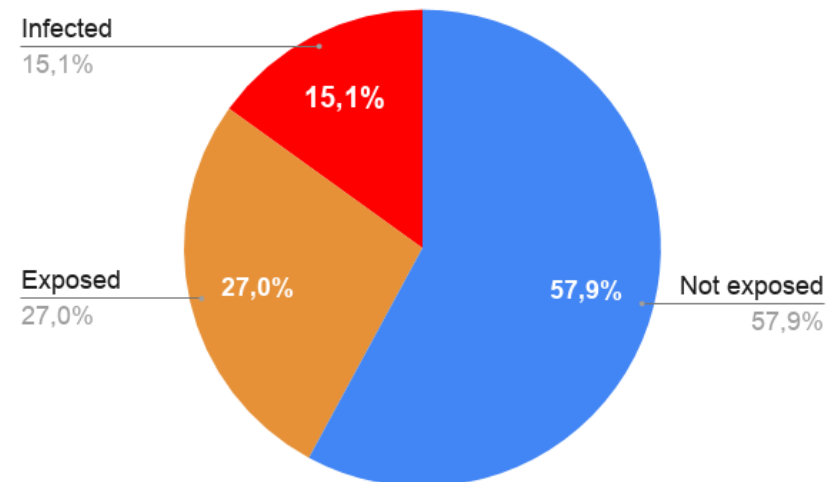
Infection overall: random_1500



Type of exposure: random_1500



Bot: random
Grafo: 1500

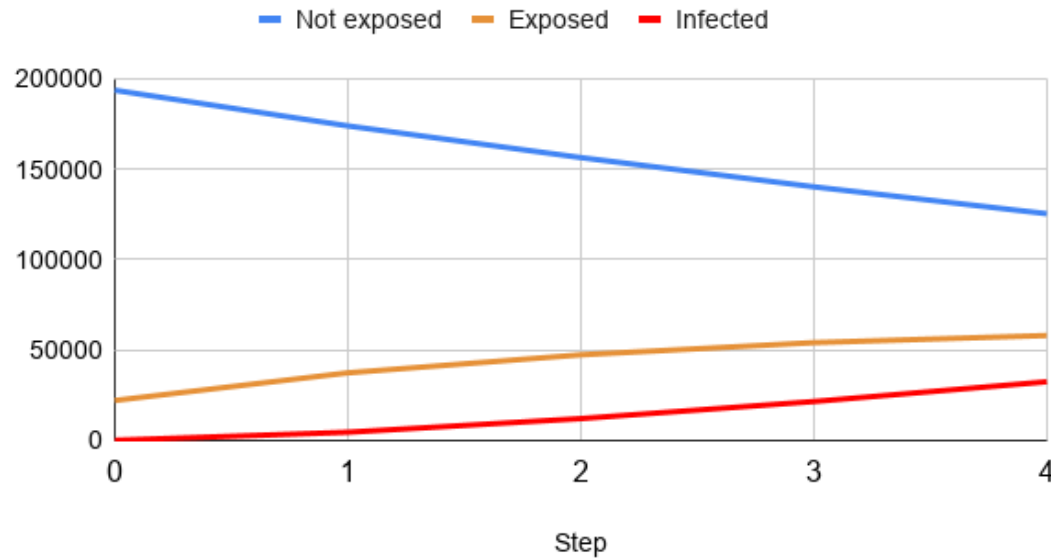


Simulazioni: Random (2000)

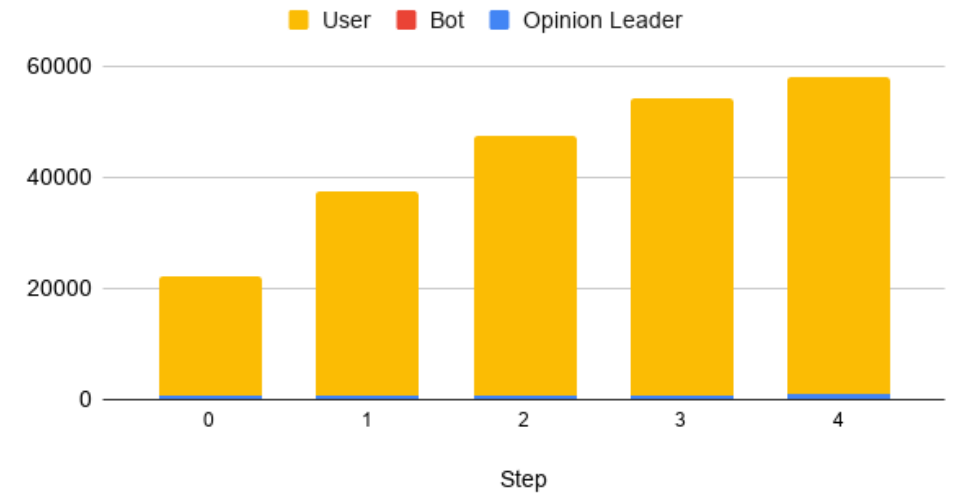
Risultati



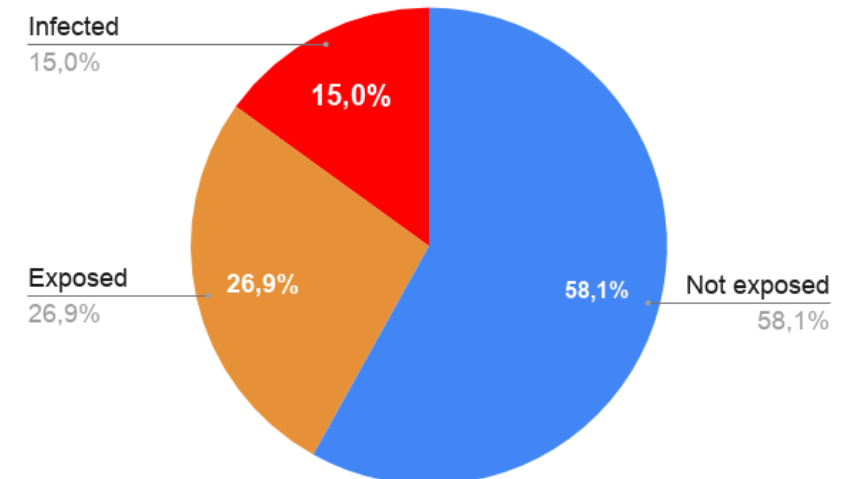
Infection overall: random_2000



Type of exposure: random_2000



Bot: random
Grafo: 2000

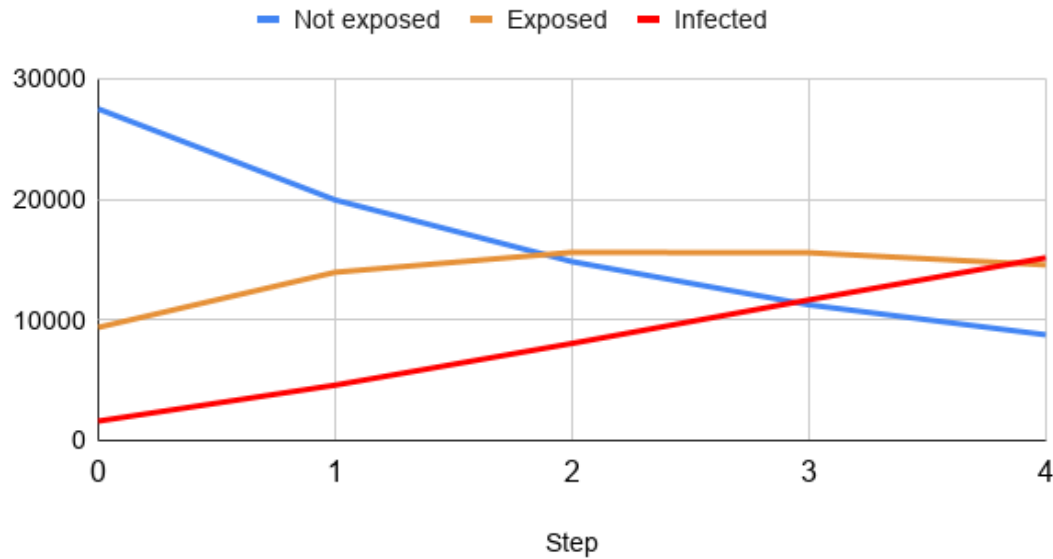


Simulazioni: Betweenness (500)

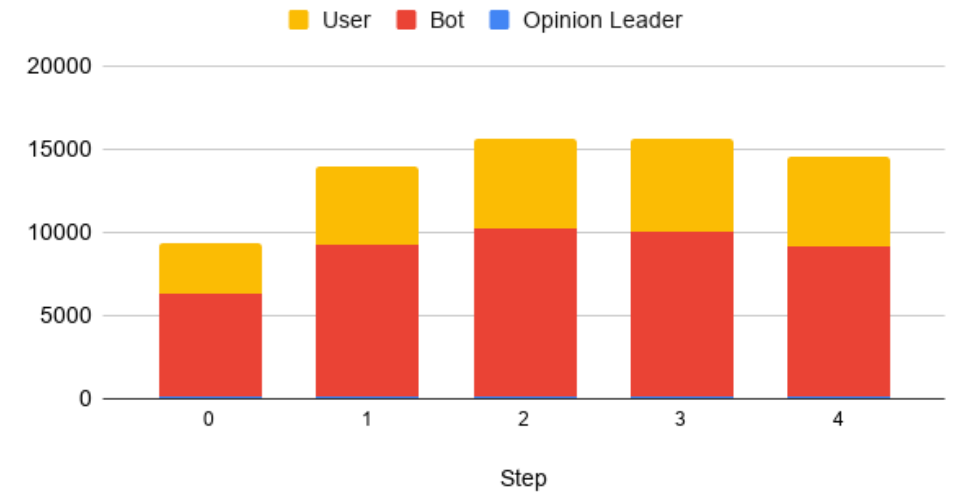
Risultati



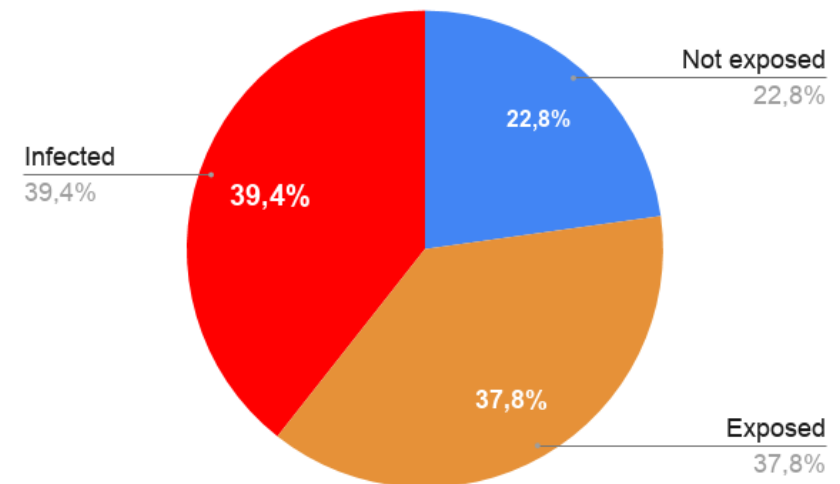
Infection overall: beetweenness_500



Type of exposure: beetweenness_500



Bot: Betweenness
Grafo: 500

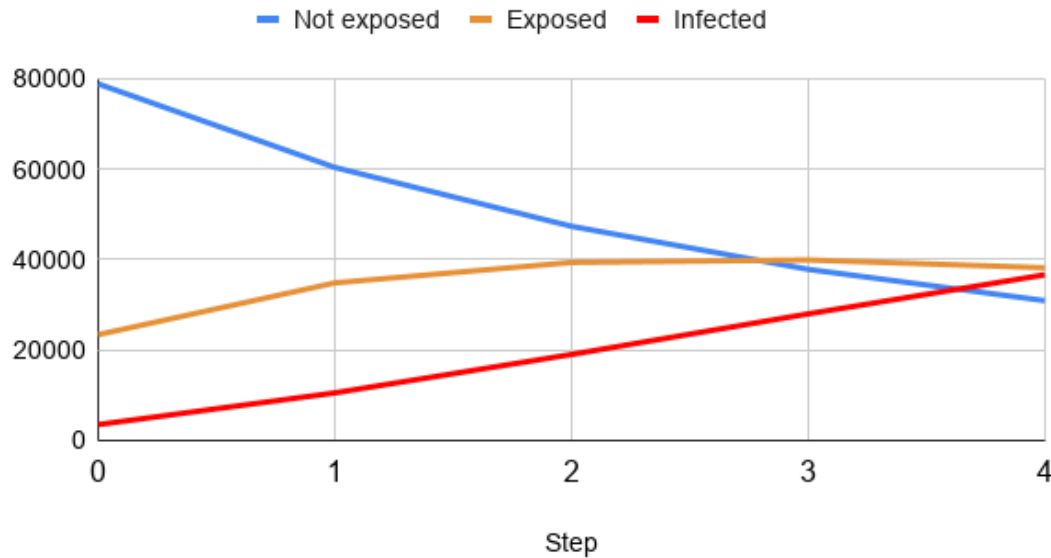


Simulazioni: Betweenness (1000)

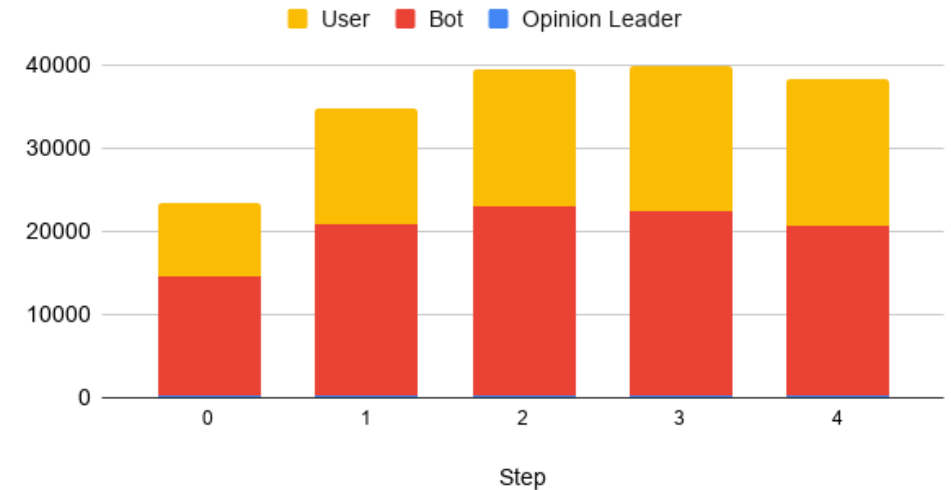
Risultati



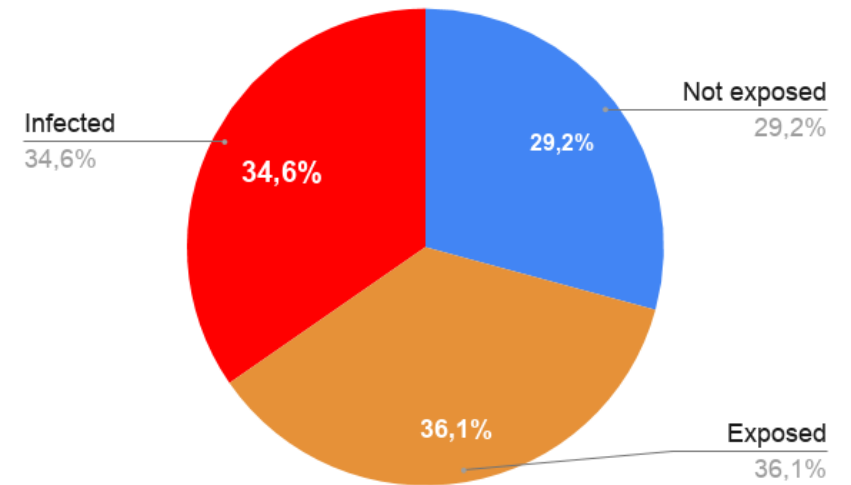
Infection overall: beetweenness_1000



Type of exposure: beetweenness_1000



Bot: Betweenness
Grafo: 1000

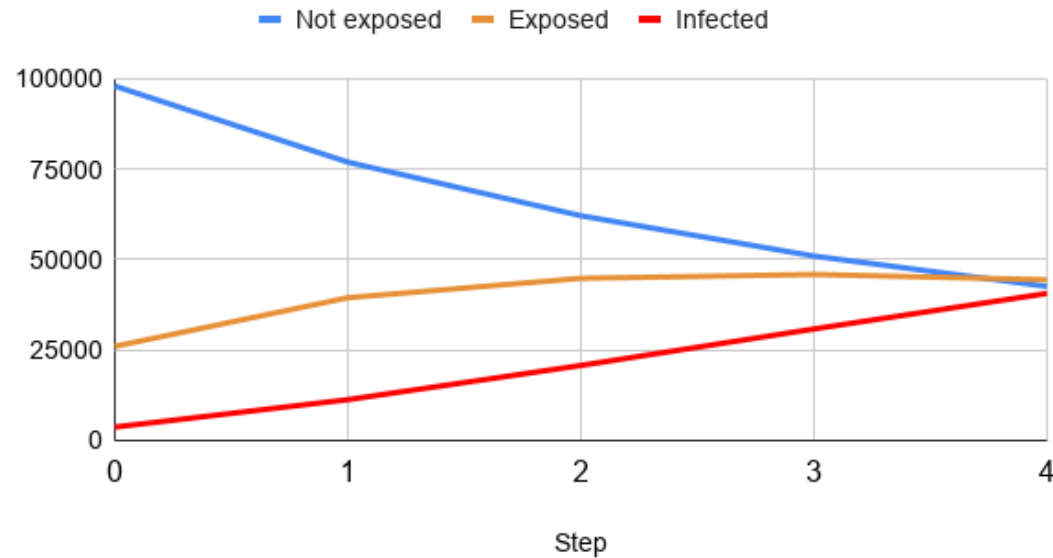


Simulazioni: Betweenness (1500)

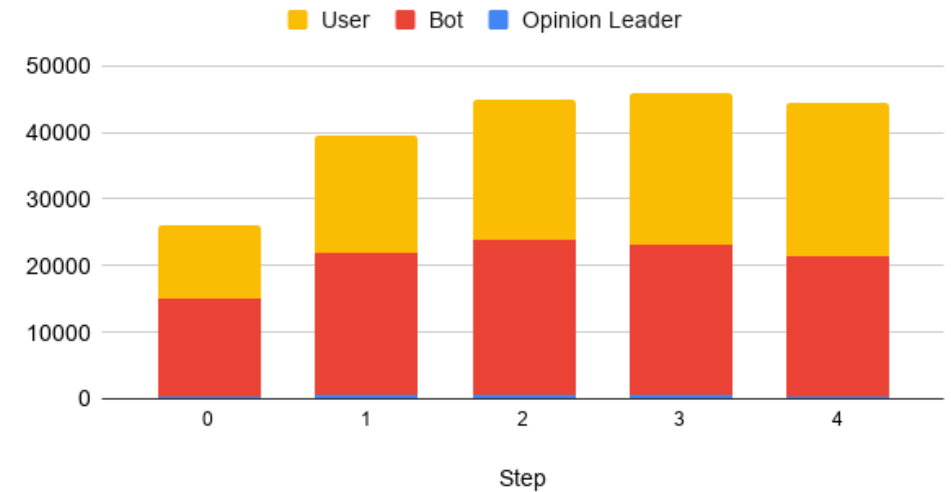
Risultati



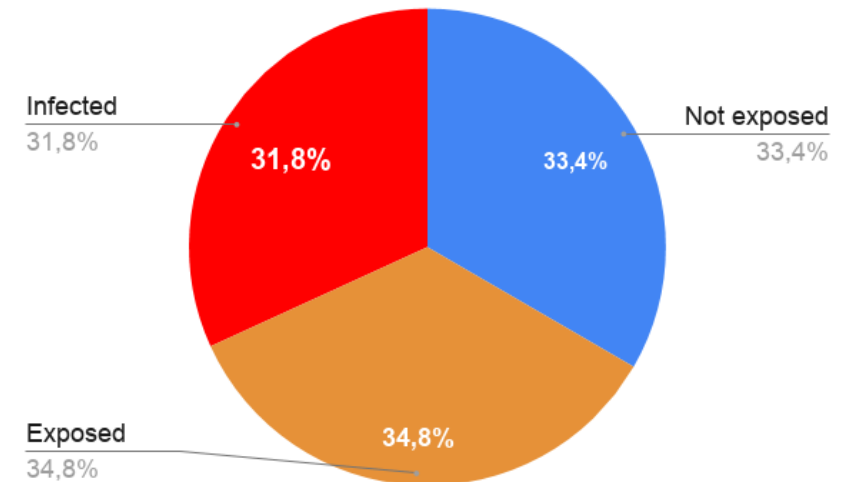
Infection overall: beetweenness_1500



Type of exposure: beetweenness_1500



Bot: Betweenness
Grafo: 1500

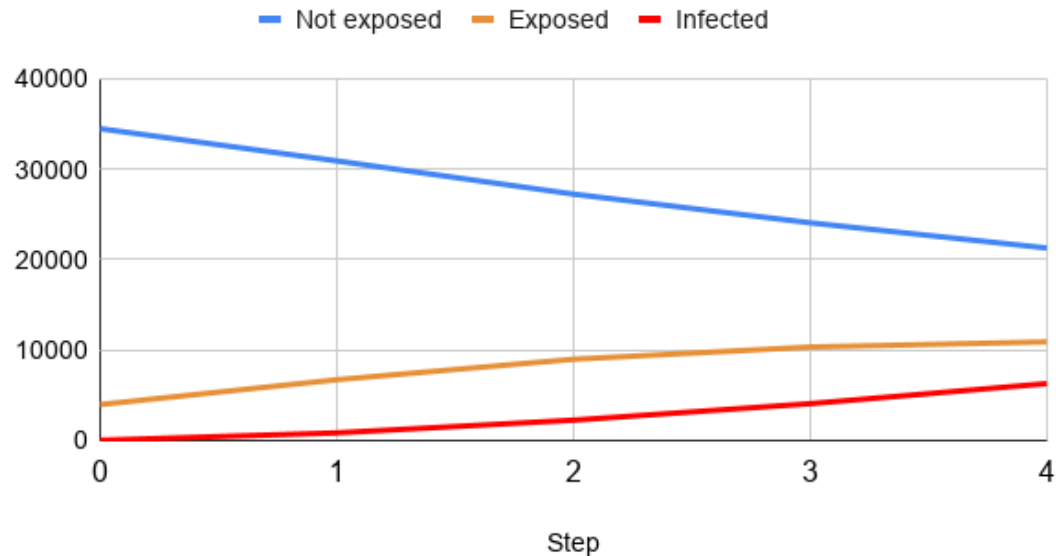


Simulazioni: Eigenvector (500)

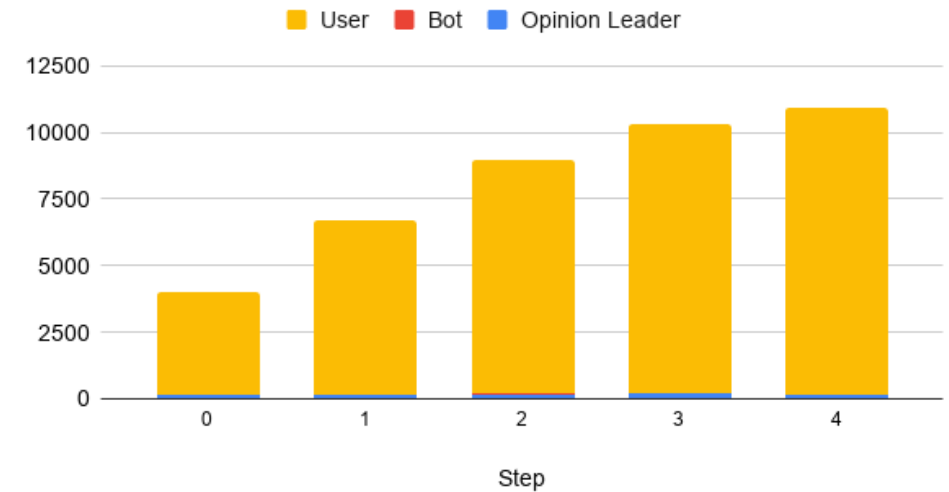
Risultati



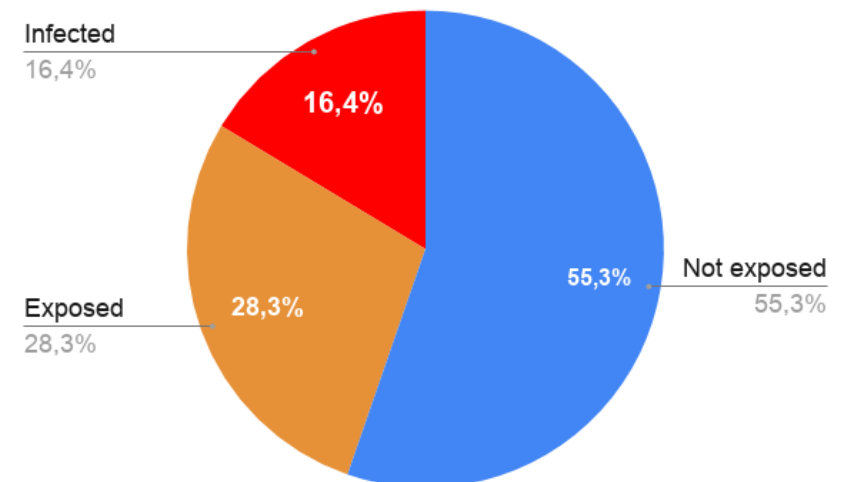
Infection overall: eigenvector_500



Type of exposure: eigenvector_500



Bot: Eigenvector
Grafo: 500

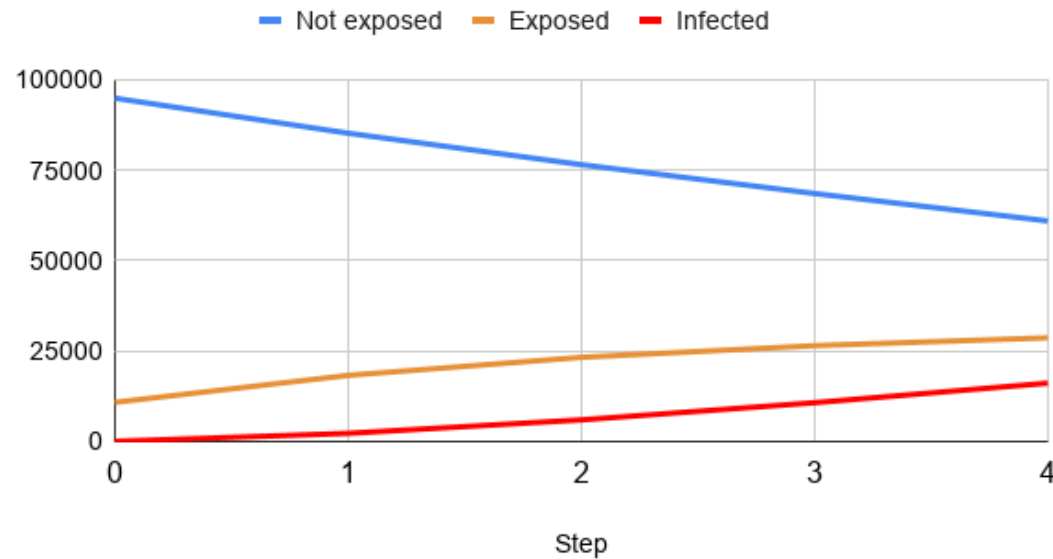


Simulazioni: Eigenvector (1000)

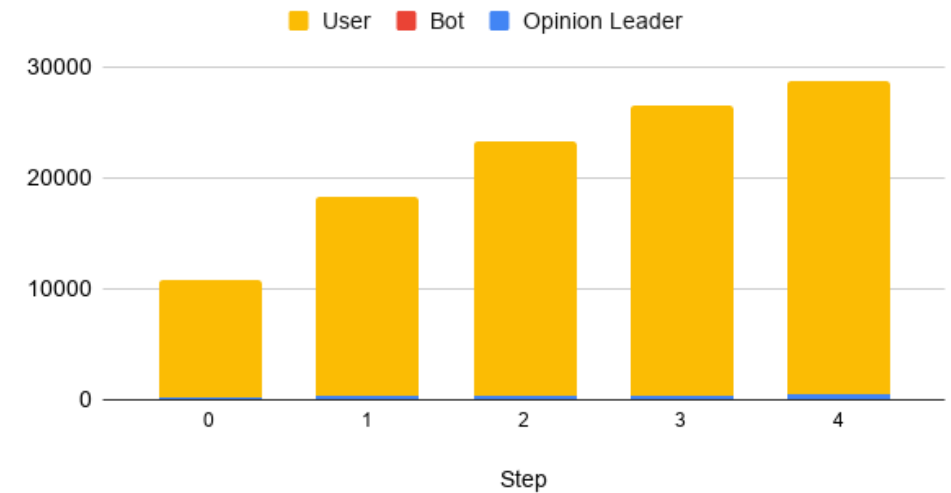
Risultati



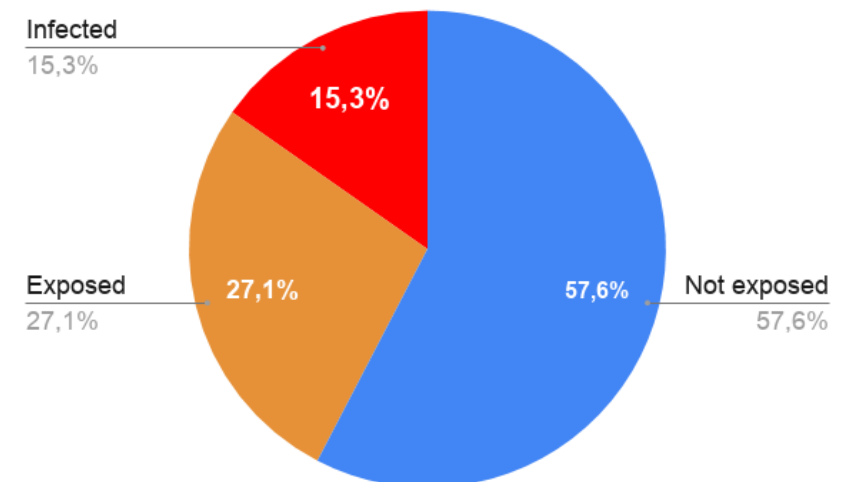
Infection overall: eigenvector_1000



Type of exposure: eigenvector_1000



Bot: Eigenvector
Grafo: 1000

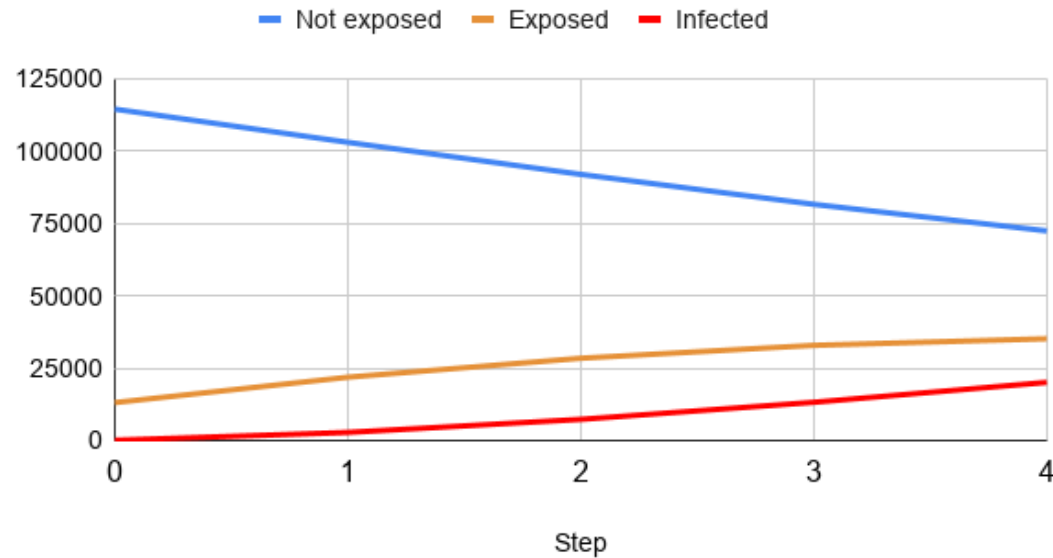


Simulazioni: Eigenvector (1500)

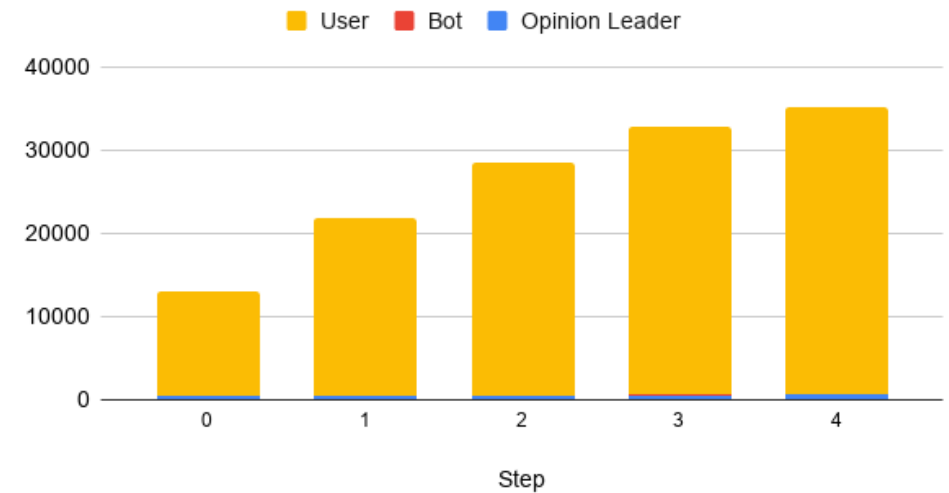
Risultati



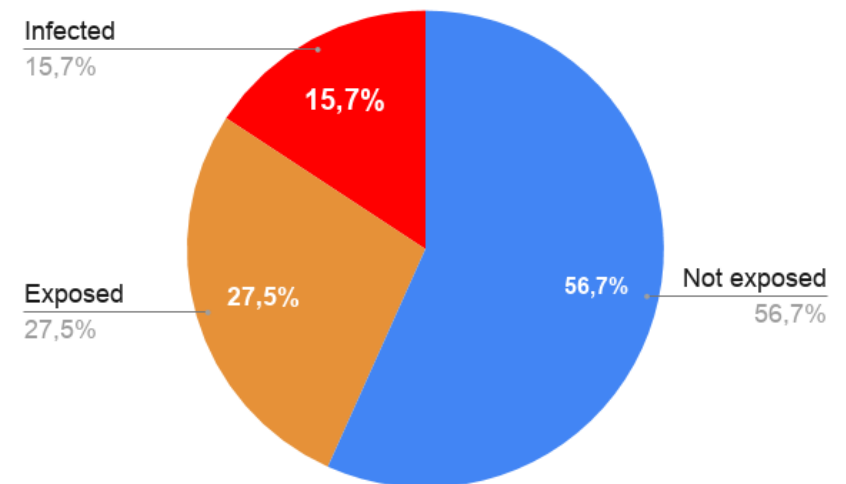
Infection overall: eigenvector_1500



Type of exposure: eigenvector_1500



Bot: Eigenvector
Grafo: 1500

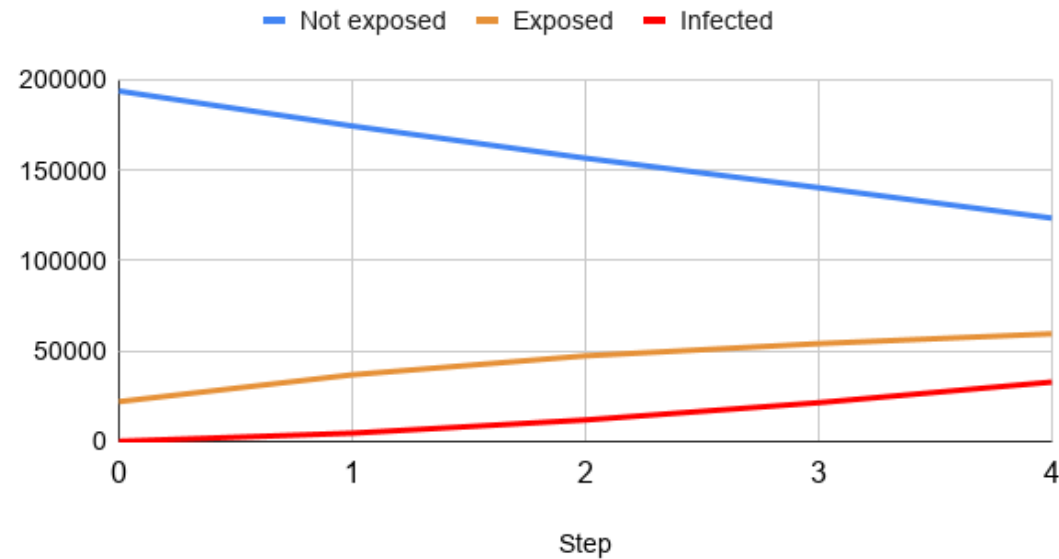


Simulazioni: Eigenvector (2000)

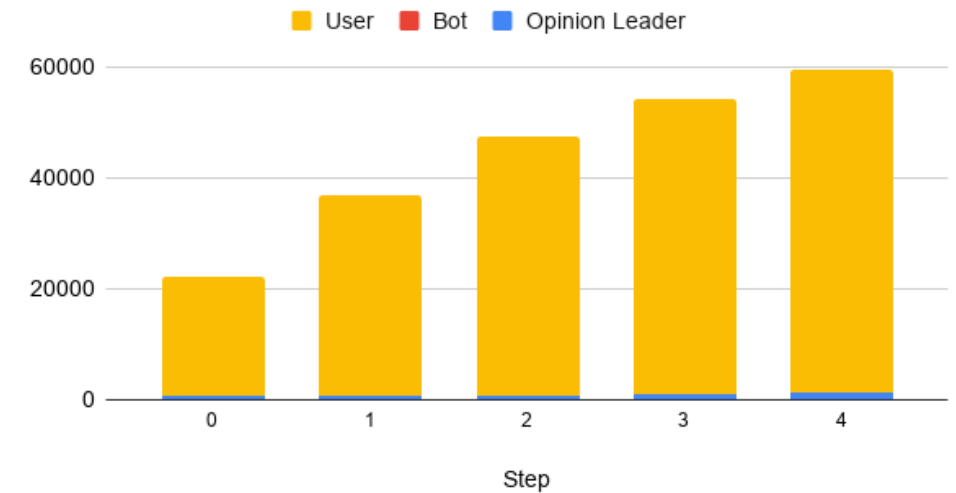
Risultati



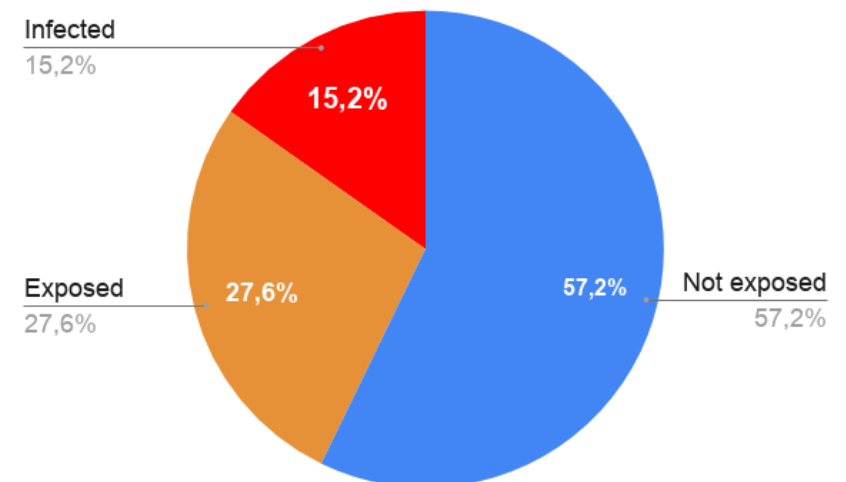
Infection overall: eigenvector_2000



Type of exposure: eigenvector_2000

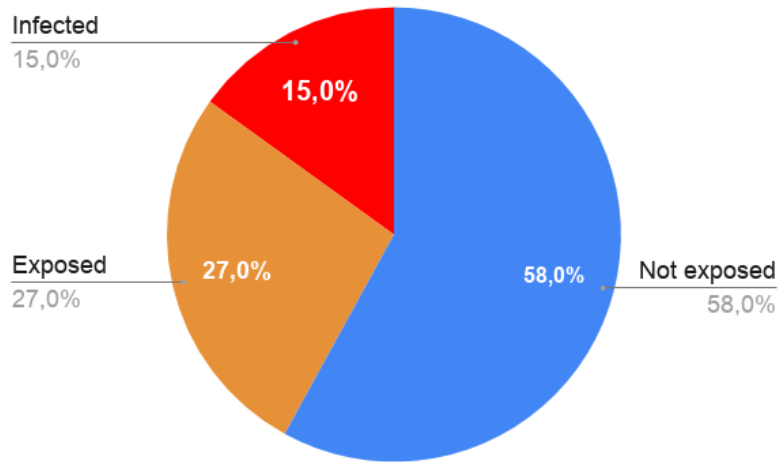


Bot: Eigenvector
Grafo: 2000

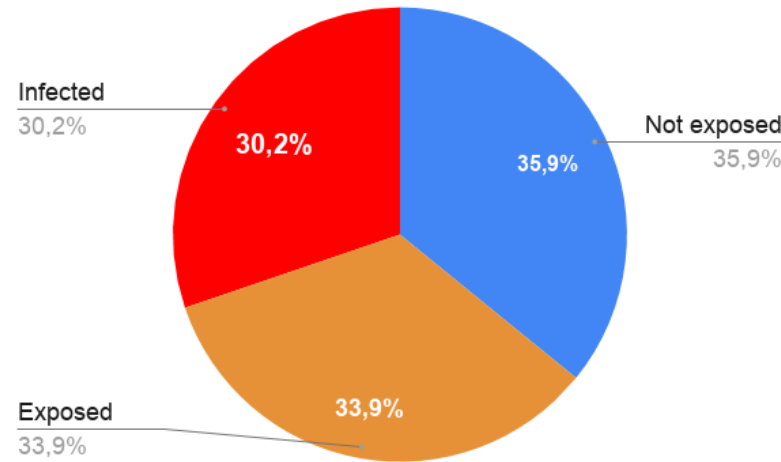


Media simulazioni

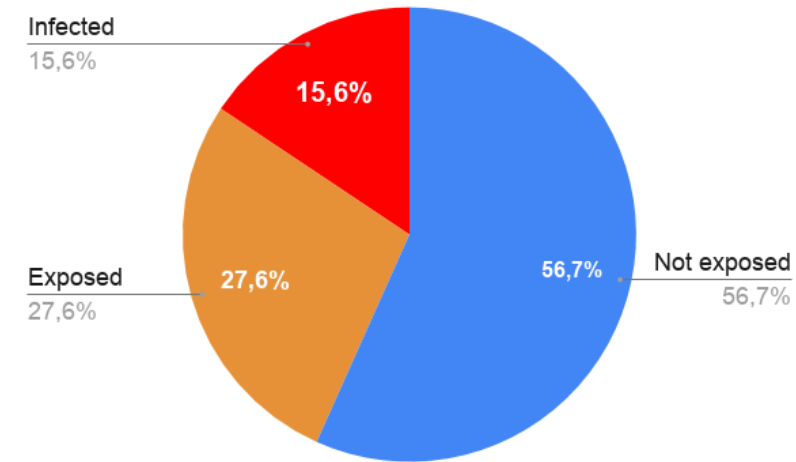
Risultati



Random



Betweenness



Eigenvector

Considerazioni finali

Conclusioni



Risultati

- **Posizionamento secondo Betweenness:**
configurazione migliore (utenti infetti o esposti superano non infetti)
- **Infezione dei bot:** cambia a seconda del loro posizionamento
- **Evoluzione contagio maggiore:** Configurazione con Betweenness
- **Opinion Leader:** Non così influente

Problemi riscontrati

- **Problemi nel reperire dati di valutazione realistici (da twitter)**
- **Problemi computazionali su grandi grafi**
- **Documentazione di SOIL e webapp**

Sviluppi futuri

Conclusioni



Area modellistica

- **Arricchimento modellazione utenti**
(caratteristiche specifiche per utente aggiuntive)
- **Riscontro con dati reali**
- **Caratterizzazione dei contenuti** (tipologie)
- **Test con nuovi modelli**

Area applicativa

- Miglioramento UX
- Incremento performance
- Deploy dell'applicazione
- Aggiornamento codice e repository

Grazie!

