



Fake News Countermeasures

Raffaele Cerizza
845512



Problema e obiettivi

⊙ Problema:

- La disinformazione è agevolata dai ***social media***.
- I *social media* si basano su un ***social network*** di utenti.
- Un *social network* è un **sistema complesso** di agenti interagenti che esibisce fenomeni emergenti come le *echo chamber*.

⊙ Obiettivi:

- Si vogliono valutare possibili contromisure alla diffusione di *fake news* in *social network*. A questo scopo si è proceduto a:
 - definire **modelli ad agenti** per la creazione di *social network* e per la diffusione di *fake news*;
 - **validare** i modelli definiti;
 - svolgere **simulazioni** di diverse contromisure sui modelli validati.

The background of the slide is a light gray network pattern. It consists of numerous small circles, some of which are solid gray and others are hollow with a gray outline. These circles are interconnected by a web of thin, light gray lines, creating a complex, organic structure that resembles a social network or a molecular structure.

Modelli della rete sociale

Rete sociale - Modello di Barabási-Albert

- ◎ **Growth**: crescita progressiva del numero di nodi.
- ◎ **Preferential attachment**: i nodi tendono a connettersi preferibilmente ai nodi con più connessioni nella rete. La probabilità che un nuovo nodo si connetta al nodo i è:

$$P(k_i) = \frac{k_i}{\sum_j k_j},$$

dove k è il grado di un nodo.

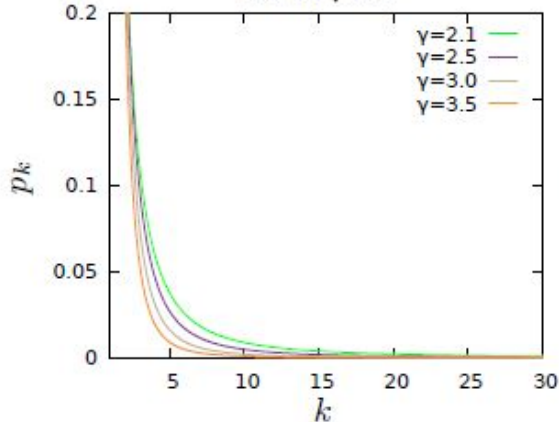


Rete sociale - Modello di Barabási-Albert

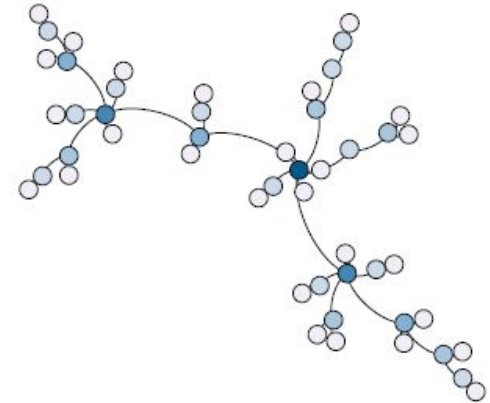
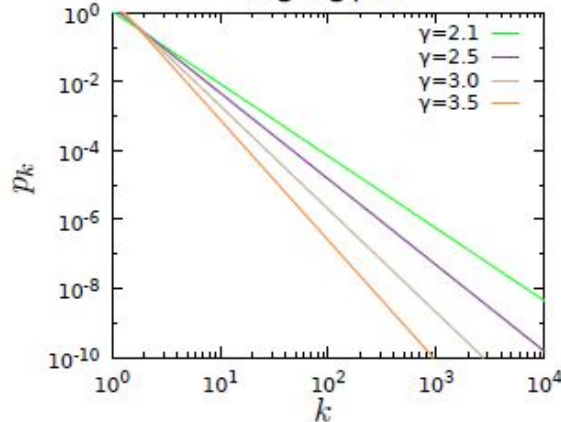
- Viene generata una rete **scale-free** dove i gradi dei nodi sono distribuiti secondo una legge di potenza (**power-law**).
- Il grafo generato è non orientato.

Power Law

Lin-lin plot



Log-log plot



Distribuzione power-law [1]

Rete scale-free [2]

Rete sociale - Modello di Bollobás

- La rete generata è *scale-free: growth e preferential attachment*.
- In questo caso però il grafo è **orientato** e ammette cappi e multi-archi. In particolare:
 - con probabilità α viene aggiunto un nuovo nodo v e un arco da v a un nodo già esistente w . Il nodo w viene scelto con probabilità:

$$P(w) = \frac{(d_{\text{in}}(w) + \delta_{\text{in}})}{(t + \delta_{\text{in}}n(t))}.$$

d_{in} = in-degree

δ_{in} = parametro reale in $[0,1]$

- con probabilità β viene aggiunto un arco da un nodo già esistente v a un altro nodo già esistente w . Il nodo w viene scelto con la probabilità appena specificata. Il nodo v viene scelto con probabilità:

$$P(v) = \frac{(d_{\text{out}}(v) + \delta_{\text{out}})}{(t + \delta_{\text{out}}n(t))}.$$

d_{out} = out-degree

δ_{out} = parametro reale in $[0,1]$

- con probabilità γ viene aggiunto un nuovo nodo w e un arco da un nodo già esistente v al nodo w . Il nodo v viene scelto con l'ultima probabilità specificata;
- la somma di α , β e γ è pari a 1.

Rete sociale - Modello di Anwar

- ⊙ La rete generata è orientata e *scale-free: growth e preferential attachment*.
- ⊙ Anche questa rete ammette cappi e multi-archi.
- ⊙ In questo caso però viene introdotta l'**omofilia**. In particolare:
 - ogni nodo ha un'opinione con valore binario: 0 o 1;
 - l'opinione di un nuovo nodo appartiene alla categoria di maggioranza con probabilità p_M ;
 - viene introdotto un parametro h che regola l'omofilia e che ha valore reale compreso fra 0 e 1;
 - per ogni coppia di nodi viene calcolato il valore $h(v,w)$ come:

$$h(v, w) = \begin{cases} h & \text{se } v \text{ e } w \text{ appartengono alla stessa categoria} \\ 1 - h & \text{altrimenti} \end{cases}$$

Rete sociale - Modello di Anwar

◎ I nuovi archi vengono aggiunti come segue:

- con probabilità α viene aggiunto un nuovo nodo v e un arco da v a un nodo già esistente w . Il nodo w viene scelto con probabilità:

$$P(w) = \frac{(h(v, w)d_{\text{in}}(w) + \delta_{\text{in}})}{(\sum_z h(v, z)d_{\text{in}}(z) + \delta_{\text{in}})}.$$

$d_{\text{in}} = \text{in-degree}$
 $\delta_{\text{in}} = \text{parametro reale in } [0,1]$

- con probabilità β viene aggiunto un arco da un nodo già esistente v a un altro nodo già esistente w . Il nodo w viene scelto con la probabilità appena specificata. Il nodo v viene scelto con probabilità:

$$P(v) = \frac{(d_{\text{out}}(v) + \delta_{\text{out}})}{(\sum_z d_{\text{out}}(z) + \delta_{\text{out}})}.$$

$d_{\text{out}} = \text{out-degree}$
 $\delta_{\text{out}} = \text{parametro reale in } [0,1]$

- con probabilità γ viene aggiunto un nuovo nodo w e un arco da un nodo già esistente v al nodo w . Il nodo v viene scelto con probabilità:

$$P(v) = \frac{(h(v, w)d_{\text{out}}(v) + \delta_{\text{out}})}{(\sum_z h(z, w)d_{\text{out}}(z) + \delta_{\text{out}})}.$$

$d_{\text{out}} = \text{out-degree}$
 $\delta_{\text{out}} = \text{parametro reale in } [0,1]$

Rete sociale - Modello proposto

◎ Il modello proposto si basa sul modello di Anwar con alcune differenze:

- l'opinione dei nodi ha valore reale compreso fra 0 e 1;
- l'opinione dei nodi è campionata da una distribuzione uniforme;
- il fattore $h(v,w)$ viene ridefinito come:

$$h(v, w) = |h - |op_v - op_w||,$$

dove op_v è l'opinione del nodo v ;

- non sono presenti né cappi né multi-archi.

◎ Una volta terminata la creazione della rete, non è più possibile modificarla aggiungendo o rimuovendo nodi o archi.

I nodi rappresentano gli utenti e gli archi rappresentano la relazione di *follow* tipica di Twitter.

Rete sociale - Modello proposto

- I parametri del modello proposto sono riassunti nella seguente tabella:

Parametri	Significato
n	Numero di nodi della rete
α	Probabilità di aggiungere un nuovo arco che ha come sorgente un nodo nuovo
β	Probabilità di aggiungere un nuovo arco tra due nodi già esistenti
γ	Probabilità di aggiungere un nuovo arco che ha come destinazione un nodo nuovo
h	Parametro che regola l'omofilia della rete
δ_{in}	Parametro utilizzato per scegliere un nodo già esistente come destinazione di un arco che parte da un altro nodo
δ_{out}	Parametro utilizzato per scegliere un nodo già esistente come sorgente di un arco che ha come destinazione un altro nodo

The background of the slide is a light blue-grey color with a complex network diagram. It consists of numerous small circles, some of which are solid grey and others are hollow with a grey outline. These circles are interconnected by thin, light grey lines, creating a dense, web-like pattern that covers the entire background.

Modelli della diffusione di *fake news*

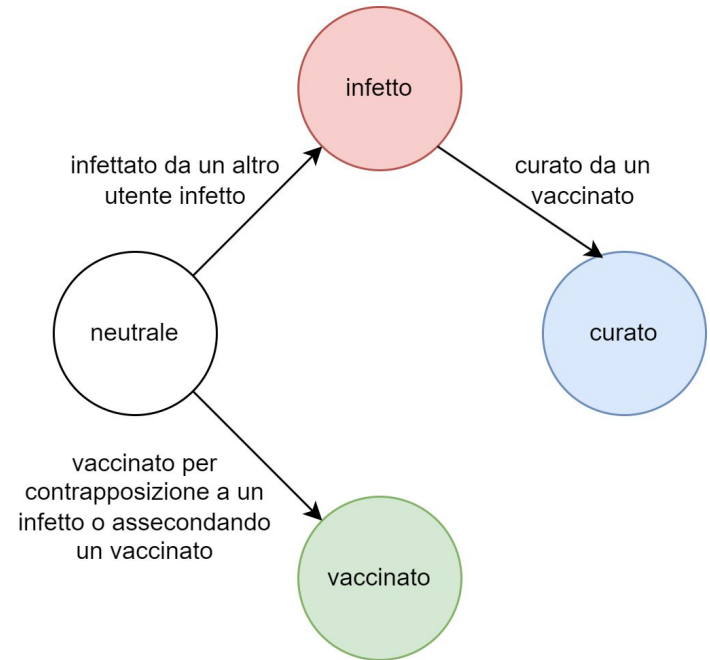
Diffusione di *fake news* - Modello di Serrano

- ⊙ Il modello di Serrano è basato sul modello epidemiologico SIR.
- ⊙ La popolazione è suddivisa in quattro categorie:
 - ***neutral***: individui suscettibili di credere o no alla *fake news*;
 - ***infected***: individui che credono alla *fake news*;
 - ***vaccinated***: individui che non credono alla *fake news* senza mai essere stati infettati;
 - ***cured***: individui che non credono più alla *fake news* dopo essere stati infettati.

Diffusione di *fake news* - Modello di Serrano

◎ Il modello di Serrano presenta i seguenti **parametri**:

- numero iniziale di persone infette;
- probabilità di infettare i vicini;
- probabilità di vaccinarsi a contatto con un infetto;
- probabilità di curare un infetto o vaccinare un neutrale.



Transizioni di stato secondo il modello di Serrano

Diffusione di *fake news* - Modello di Gausen

- Il modello di Gausen si basa sul modello di Serrano. Ne eredita le categorie della popolazione e i relativi parametri.
- Il modello di Gausen introduce alcune novità:
 - introduce la figura dell'**influencer** che ha una maggiore probabilità di infettare o vaccinare;
 - prevede che i vaccinati possano presentare **reclamo** contro gli infetti con una certa probabilità. L'infetto viene **bloccato** dopo 3 reclami. Poi non può più interagire con gli altri utenti;
 - introduce un parametro fisso che riduce la probabilità di infezione:

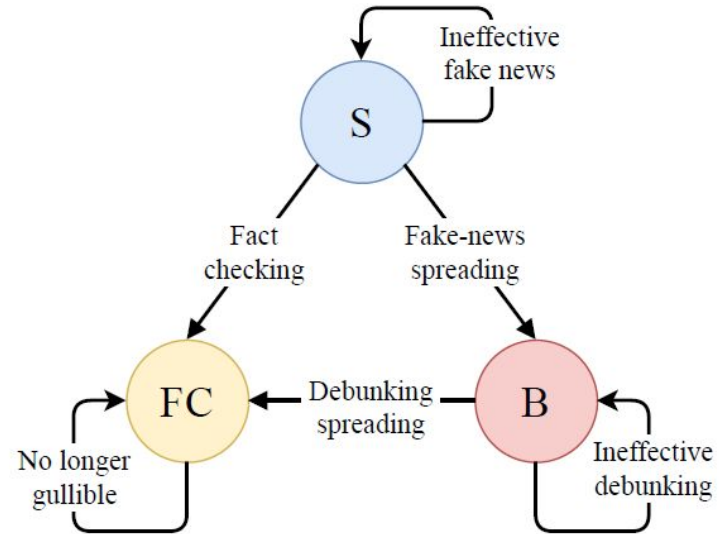
$$P_{\text{inf}} = P_{\text{inf}} - P_{\text{inoc}}.$$

- introduce un parametro fisso che aumenta la probabilità di vaccinazione:

$$P_{\text{vacc}} = P_{\text{vacc}} + P_{\text{acc}}.$$

Diffusione di *fake news* - Modello di Lotito

- Il modello di Lotito prevede due tassonomie ortogonali tra loro.
- Tassonomia epidemiologica:
 - **susceptible**: utenti suscettibili di credere alla *fake news*;
 - **believer**: utenti che credono alla *fake news*;
 - **fact-checker**: utenti che non credono alla *fake news*.



Transizioni di stato secondo il modello di Lotito [3]

Diffusione di *fake news* - Modello di Lotito

◎ Tassonomia dei ruoli:

- ***common***: normali utenti della rete;
- ***influencer***: utenti che hanno una maggiore capacità diffusiva rispetto ai *common*;
- ***bot***: *account* automatizzati che diffondono la *fake news*;
- ***eternal fact-checker***: utenti che combattono continuamente la diffusione di *fake news*.

◎ La sovrapposizione tra le tassonomie opera in questo modo:

- i *common* e gli *influencer* possono essere *susceptible*, *believer* o *fact-checker*;
- i *bot* possono essere solo *believer*;
- gli *eternal fact-checker* possono essere solo *fact-checker*.

◎ Tutti gli utenti possiedono un'opinione e coordinate geografiche che influiscono sulla costruzione della rete sociale. Le probabilità di infezione e vaccinazione sono analoghe a quelle dei modelli precedenti.

Diffusione di *fake news* - Modello di Lotito

- ⊙ Infine il modello di Lotito disciplina la **dinamica temporale**:
 - L'accesso degli utenti al *social network* decresce secondo una distribuzione esponenziale. I *bot* accedono con una frequenza 4 volte maggiore.
 - L'interesse degli utenti a diffondere notizie decresce nel tempo. A questo proposito è previsto un coefficiente di coinvolgimento E :

$$E(t) = E_0 e^{-\lambda t}.$$

In questo caso t è l'istante di tempo e λ è la costante di decrescita.

Diffusione di *fake news* - Modello di Törnberg

- ◎ Il modello di diffusione di Törnberg sfrutta le ***echo chamber*** come segue:
 - viene creata una rete sociale secondo il modello random di Erdős–Rényi, riservando una frazione di nodi per l'*echo chamber*;
 - una frazione di archi aventi un solo estremo nell'*echo chamber* viene sostituita con archi interni all'*echo chamber*;
 - viene attivato (come nodo infetto) un nodo random nell'*echo chamber* e i nodi a esso collegati;
 - gli altri nodi vengono attivati in base a una soglia di attivazione. La soglia di attivazione per i nodi esterni all'*echo chamber* è θ , quella per i nodi interni all'*echo chamber* è $\theta - P_o$, dove θ e P_o sono parametri del modello. La soglia di attivazione per i nodi interni è più bassa rispetto ai nodi esterni;
 - in particolare un nodo viene attivato se i nodi collegati a esso sono attivati per una frazione superiore alla soglia di attivazione.

Diffusione di *fake news* - Modello proposto

- ◎ Il modello proposto ha le seguenti caratteristiche:
 - la popolazione è suddivisa nelle **categorie epidemiologiche** del modello di Serrano;
 - agli utenti viene attribuito un **ruolo** secondo il modello di Lotito;
 - gli utenti vaccinati possono presentare un **reclamo** contro gli utenti infetti con una probabilità $P_{\text{complaint}}$. Con 3 reclami l'utente viene **bloccato** e non può più pubblicare messaggi;
 - l'**attività degli utenti decresce** nel tempo secondo un fattore esponenziale:

$$t_{\text{next}} = t_{\text{current}} + \max\left\{1, \left\lfloor \frac{1}{e^{-\lambda t_{\text{current}}}} \right\rfloor\right\},$$

dove λ è campionato da una distribuzione uniforme tra 0.01 e 0.05;

- i *bot* sono invece attivi in ogni istante di tempo;

Diffusione di *fake news* - Modello proposto

- la probabilità di infezione, vaccinazione e cura è influenzata da un **fattore di differenza delle opinioni** calcolato come:

$$op_{\text{diff}} = 1 - |op_x - op_y|,$$

- viene modellata una **soglia di echo chamber** P_{echo} tale che:
 - un utente neutrale diventa infetto se la frazione di utenti infetti da lui “seguiti” supera questa soglia;
 - un utente neutrale diventa vaccinato se la frazione di utenti vaccinati da lui “seguiti” supera questa soglia;
 - un utente infetto diventa curato se la frazione di utenti vaccinati da lui “seguiti” supera questa soglia.
- il meccanismo della soglia di *echo chamber* si attiva solo se la soglia è superiore a 0.5.

Diffusione di *fake news* - Modello proposto

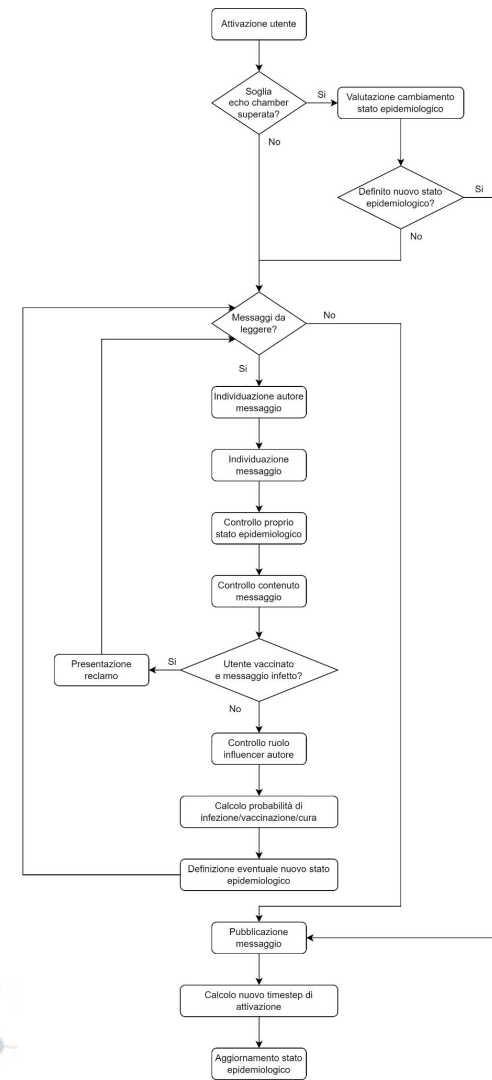
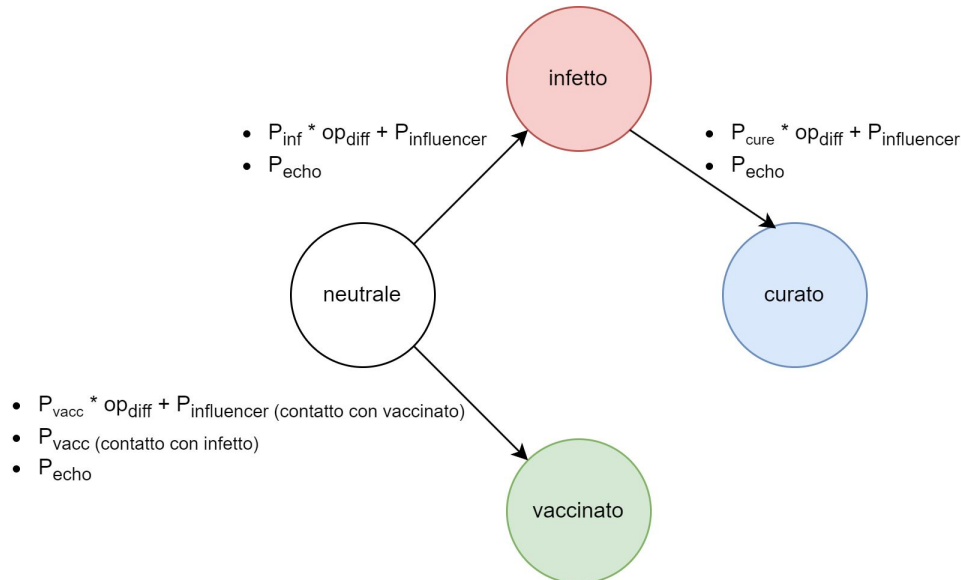
- Vengono qui riportati i parametri principali del modello proposto:

Parametri	Significato
P_{inf}	Probabilità di essere infettato
P_{vacc}	Probabilità di essere vaccinato
P_{cure}	Probabilità di essere curato
$P_{influencer}$	Maggiore probabilità degli <i>influencer</i> di infettare o vaccinare o curare
P_{echo}	Soglia per il meccanismo delle <i>echo chamber</i>
$P_{complaint}$	Probabilità di presentare un reclamo contro un utente infetto

Diffusione di *fake news* - Modello proposto

Workflow delle attività di un utente: 

Schema delle transizioni di stato:

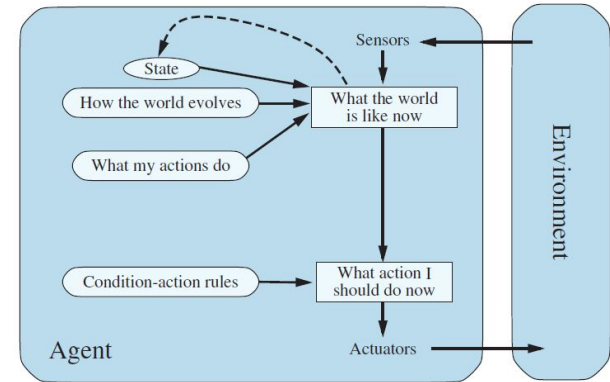


The background of the slide is a light gray network pattern. It consists of numerous small circles, some of which are solid gray and others are hollow with a gray outline. These circles are interconnected by thin, light gray lines, creating a complex, web-like structure that fills the entire background.

Modellazione basata su agenti

Modelli basati su agenti - Agenti

- Nei modelli proposti gli agenti sono gli utenti di un *social network*.
- In particolare si tratta di **agenti isteretici**:
 - **I**: stato epidemiologico, messaggi pubblicati, reclami, blocco;
 - **E**: stato degli altri utenti della rete;
 - **P**: stato degli utenti “seguiti”;
 - **A**: messaggio da pubblicare e reclamo;
 - i_0 : stato iniziale dell’utente;
 - **see**: lettura dei messaggi;
 - **internal**: aggiornamento stato epidemiologico;
 - **action**: pubblicazione messaggio e presentazione reclamo;
 - **do**: blocco degli utenti.



Architettura agenti [4]

Modelli basati su agenti - Ambiente

- ⊙ Nei modelli proposti l'**ambiente** è la rete sociale degli utenti. I nodi sono gli utenti. Gli archi rappresentano la relazione unidirezionale di *follow*.
- ⊙ L'ambiente presenta queste caratteristiche:
 - **parzialmente accessibile**: gli utenti percepiscono solo il sottoinsieme di utenti che “seguono”;
 - **statico**: terminata la creazione della rete non è più possibile aggiungere o rimuovere nodi o archi;
 - **sequenziale**: l'azione corrente e l'aggiornamento dello stato interno influiscono sulle azioni successive;
 - **stocastico**: gli effetti dei messaggi dipendono da probabilità;
 - **discreto**: il numero di azioni e percezioni è finito.

Modelli basati su agenti - Interazione tra gli agenti

◎ L'**interazione** tra gli agenti è:

- **indiretta**: gli utenti pubblicano messaggi che vengono letti dai *follower*. Gli utenti che pubblicano messaggi non hanno bisogno di conoscere i loro *follower* (*name uncoupling*). E non è necessaria la coesistenza nello stesso istante di tempo (*time uncoupling*);
- **fondata spazialmente**: l'informazione che arriva agli utenti è influenzata dalla topologia del grafo. Gli utenti percepiscono solo una porzione del grafo: gli utenti "seguiti".

The background of the slide is a light gray network pattern. It consists of numerous small circles, some of which are solid gray and others are hollow with a gray outline. These circles are interconnected by a web of thin, light gray lines, creating a complex, organic structure that resembles a molecular or neural network.

Validazione

Validazione - Validazione *prima facie*

- Una prima forma di validazione ha riguardato la **natura *scale-free* delle reti** generate.
- Prima validazione quantitativa della proprietà *scale-free*:

- Test di Kolmogorov-Smirnov.
- Ipotesi nulla che i gradi dei nodi siano distribuiti secondo una *power-law*.

$$H_0 : F_X(t) = F(T) \quad \text{per ogni } t \in R.$$

- La distanza tra la distribuzione cercata e quella empirica è:

$$D_n = \sup_{t \in R} |F(t) - \hat{F}_{X,n}(t)|.$$

- La regione critica che porta a rifiutare l'ipotesi nulla è:

$$C = (d_{1-\alpha}, 1],$$

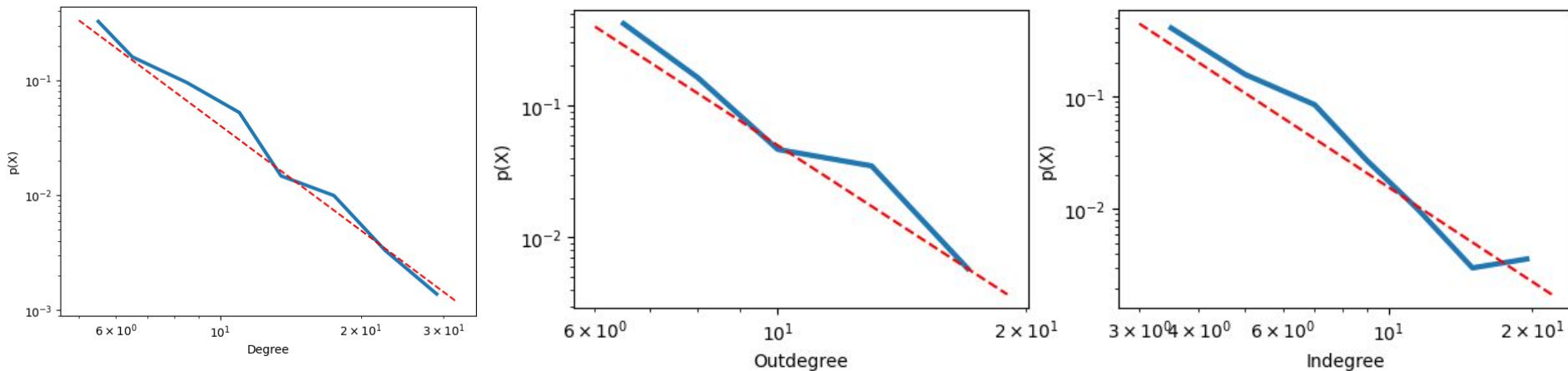
- Risultati ottenuti con 30 reti di 50 nodi ciascuna:

D_n	C	$p\text{-value}$	α	risultato test
0.13 ± 0.01	$(0.19, 1]$	0.72 ± 0.08	0.05	H_0 non rifiutata

Validazione - Validazione *prima facie*

◎ Seconda validazione qualitativa della proprietà *scale-free*:

- Visualizzazione della distribuzione dei gradi di una rete di 1000 nodi.



- La linea tratteggiata rossa rappresenta una distribuzione *power-law*. La linea continua blu rappresenta la distribuzione empirica.

Validazione - Validazione *prima facie*

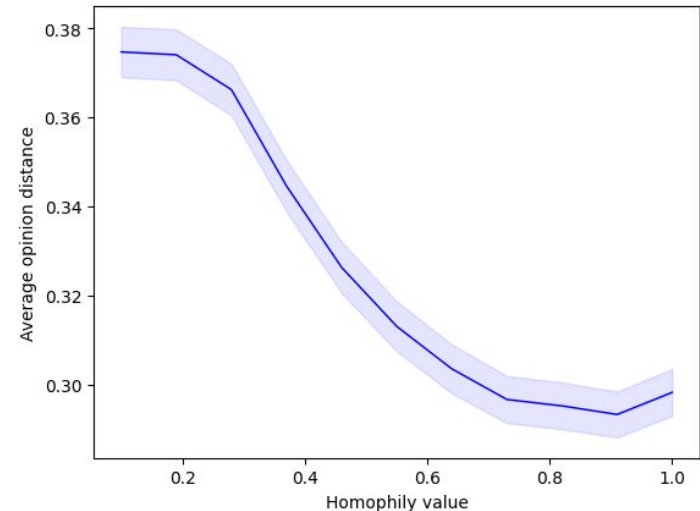
- La seconda validazione *prima facie* ha riguardato l'**omofilia**.
- Si vuole verificare che alla crescita del parametro h si accompagni una maggiore probabilità di creare connessioni tra nodi con opinioni simili:

- sono state create 50 reti di 100 nodi ciascuna per ogni valore di h considerato;
- è stata calcolata la distanza media delle opinioni dei nodi:

$$avg_{op \text{ distance}} = \frac{1}{m} \sum_{e \in E} |op_x - op_y|,$$

dove E è l'insieme degli archi e m è il numero degli archi;

- è stata calcolata la distanza media tra le reti con lo stesso valore di h ;
- al crescere di h diminuisce la distanza fra le opinioni.



Risultato della validazione dell'omofilia

Validazione - *Dataset*

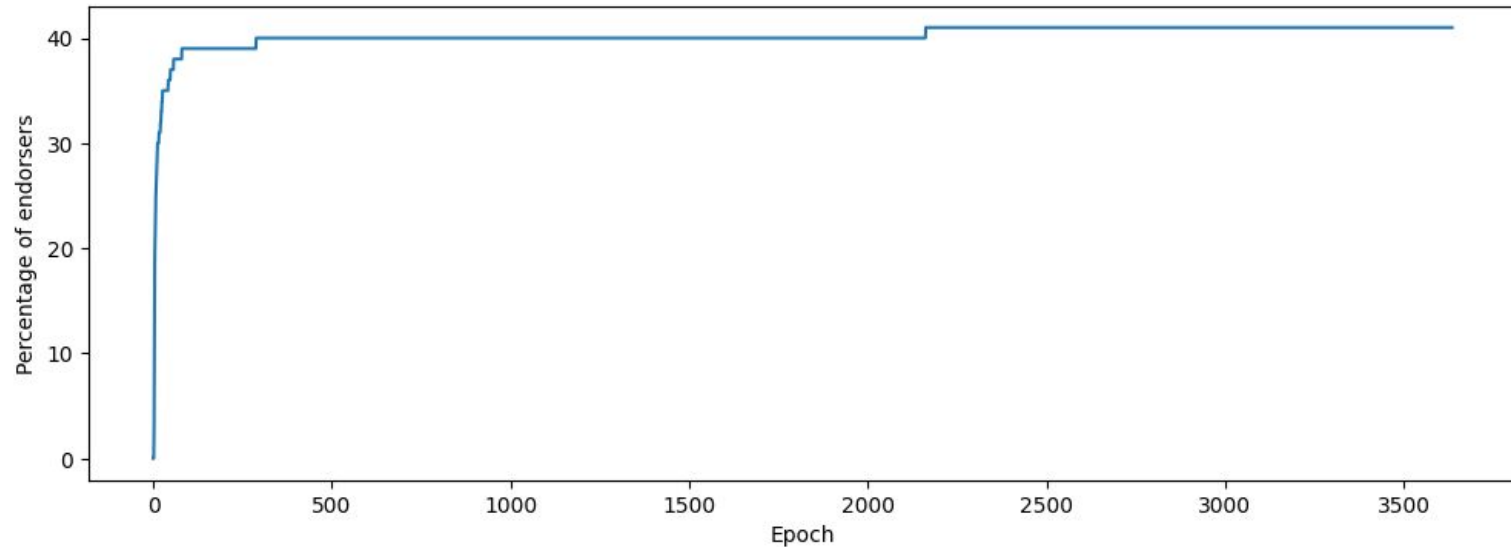
- ◎ Per le ulteriori fasi della validazione è stato necessario individuare dei ***dataset***.
- ◎ Sono stati individuati i *dataset* “palin” e “obama” descritti in [5].
- ◎ Questi dataset contengono alcune informazioni di *tweet* estratti da Twitter:
 - data di pubblicazione;
 - identificativo dell'utente;
 - etichetta che specifica se il *tweet* diffonde la *fake news*, la confuta o è neutrale.
- ◎ Difetti:
 - non contengono tutto il grafo di Twitter;
 - sono circoscritti temporalmente a pochi mesi;
 - non contengono tutti i *tweet* di tutti gli utenti considerati.

Validazione - *Dataset*

◎ *Dataset* “palin”:

<i>Dataset</i>	favorevoli	contrari	neutrali	totali
palin	1709	1895	819	4423

Utenti: 3181

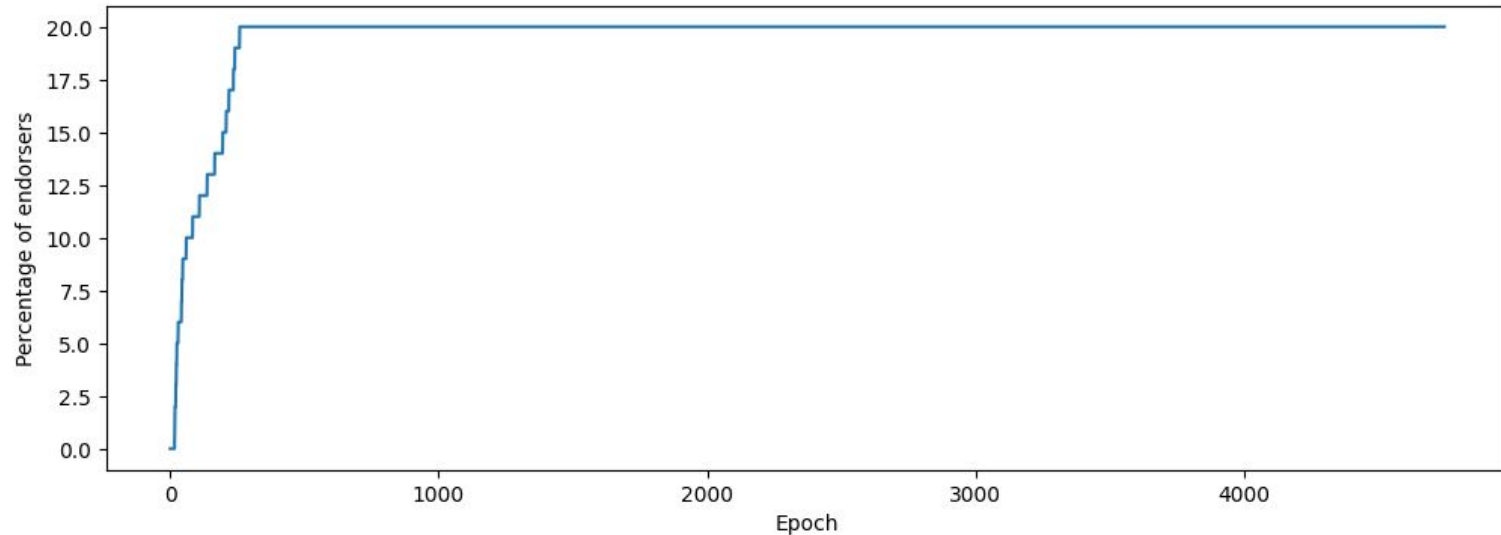


Validazione - *Dataset*

◎ **Dataset “obama”:**

<i>Dataset</i>	favorevoli	contrari	neutrali	totali
obama	854	489	2994	4337

Utenti: 2678



Validazione - Calibrazione dei parametri

- I modelli proposti presentano numerosi **parametri**. Occorre determinarli o calibrarli.
- Per la calibrazione dei parametri è stato usato il *dataset* “palin”.
- **Parametri predeterminati** sulla base del dataset:

<i>Dataset</i>	nodi	δ_{in}	δ_{out}	epoche	<i>infl.</i>	<i>bot</i>	<i>e.f.c.</i>	infetti iniziali
palin	3181	1	1	3634	25	64	64	31

- Il numero di nodi è pari al numero di utenti.
- Gli utenti con più di 50000 *follower* sono stati considerati *influencer*.
- La percentuale di *bot* è stata determinata in base alla letteratura.
- La percentuale di *eternal fact-checker* è stata posta pari alla percentuale di *bot* per limitare *bias*.
- Il numero di infetti iniziali (compresi i *bot*) è stato dedotto dal *dataset*.

Validazione - Calibrazione dei parametri

- ◎ Parametri calibrati con **grid search**:
 - I parametri α , β e γ che disciplinano le modalità di creazione degli archi della rete. Valori considerati per α e β : 0.0, 0.25, 0.33, 0.50, 0.75, 1.0. Il valore di γ è inferito in quanto la somma dei tre parametri deve essere pari a 1.
 - Il parametro h che regola l'omofilia della rete. Valori considerati: 0.25, 0.50, 0.75.
- ◎ Motivazioni dell'approccio *grid search*:
 - combinazioni di valori plausibili note *a priori*.

Validazione - Calibrazione dei parametri

- ⊙ Parametri calibrati con **ottimizzazione bayesiana**:
 - Probabilità di infezione, vaccinazione, cura, *influencer* e soglia *echo chamber*.
- ⊙ Motivazioni dell'ottimizzazione bayesiana:
 - computazionalmente efficiente per spazi di ricerca ampi.
- ⊙ Procedimento:
 - è stata utilizzata una funzione di valutazione per guidare la ricerca dei valori verso regioni più promettenti dello spazio dei valori dei parametri;
 - in particolare è stato utilizzato il **Root Mean Squared Error** (RMSE) tra la percentuale reale di infetti per epoca e la percentuale predetta di infetti per epoca:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2},$$

- l'RMSE per ogni combinazione di parametri è stato calcolato come media su 2 reti per ottenere risultati statisticamente più robusti (anche se di poco);
- l'ottimizzazione dei parametri per ogni combinazione è avvenuta con 8 iterazioni;
- l'RMSE migliore è quello con il valore più basso.

Validazione - Calibrazione dei parametri

- ◎ **Parametri migliori** del procedimento di calibrazione dei parametri per i tre valori di h considerati:

<i>Dataset</i>	α	β	γ	h	P_{inf}	P_{vacc}	P_{cure}	$P_{\text{influencer}}$	P_{echo}	RMSE
palin	0.50	0.33	0.17	0.25	0.44	0.06	0.03	0.65	0.31	1.29
palin	0.50	0.33	0.17	0.50	0.36	0.06	0.22	0.10	0.62	1.41
palin	0.25	0.50	0.25	0.75	0.36	0.02	0.18	0.76	0.12	2.81

Validazione - Validazione statistica

- La **validazione statistica** è stata condotta per verificare se i modelli proposti con i parametri calibrati fossero in grado di simulare correttamente la diffusione di *fake news* anche su un *dataset* diverso da quello usato per la calibrazione.
- Per questa fase è stato usato il *dataset* “obama”.
- I parametri predeterminati per questo *dataset* sono stati i seguenti:

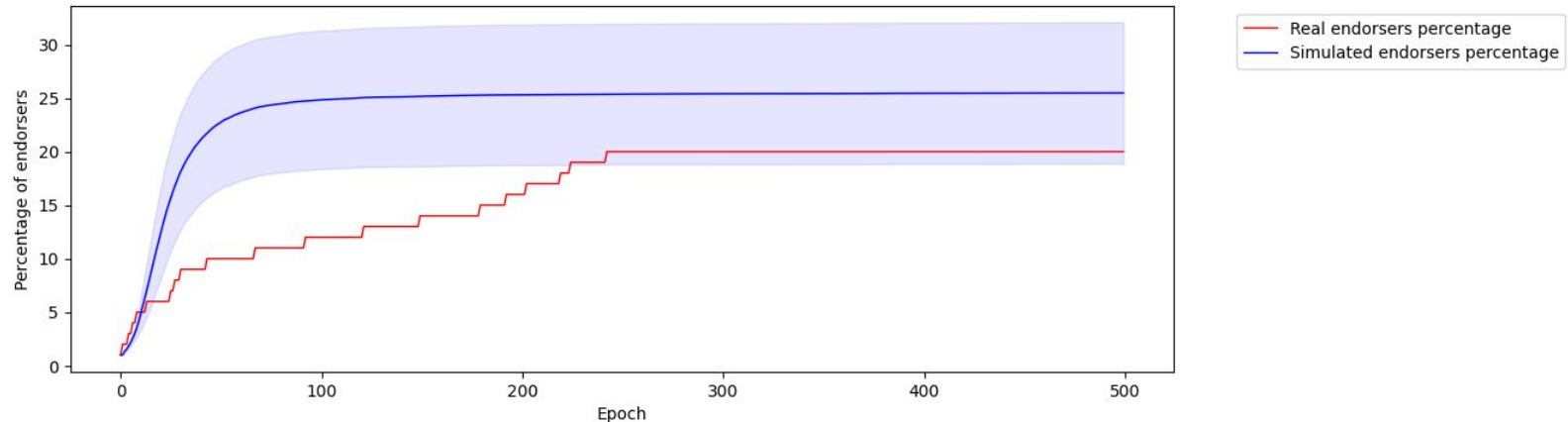
<i>Dataset</i>	nodi	δ_{in}	δ_{out}	epoche	<i>infl.</i>	<i>bot</i>	<i>e.f.c.</i>	infetti iniziali
obama	2678	1	1	4730	29	27	27	27 (con <i>bot</i>)

- I **risultati** sono stati calcolati su 30 reti:

<i>Dataset</i>	<i>h</i>	RMSE
obama	0.25	14.10 \pm 7.74
obama	0.50	7.74 \pm 4.37
obama	0.75	18.18 \pm 5.64

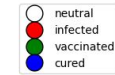
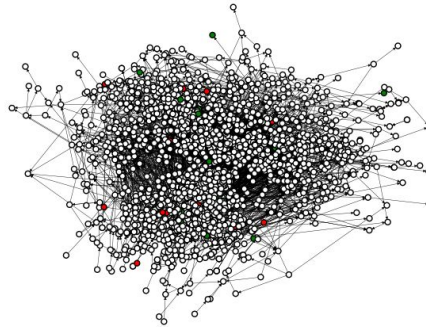
Validazione - Validazione statistica

- I risultati migliori su entrambi i *dataset* sono stati ottenuti con i parametri calibrati con un valore di h pari a 0.50. Questi risultati sono considerati accettabili. Le successive analisi e simulazioni sono state condotte con questi parametri.
- **Percentuali medie di infetti per epoca:**

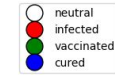
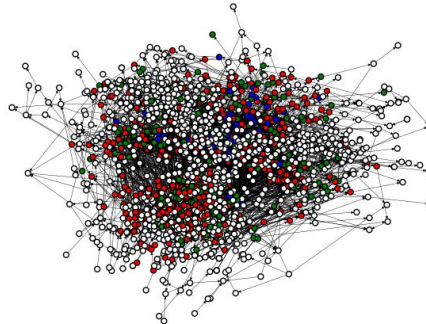


Validazione - Validazione statistica

🕒 Evoluzione del grafo:



Grafo alla prima epoca



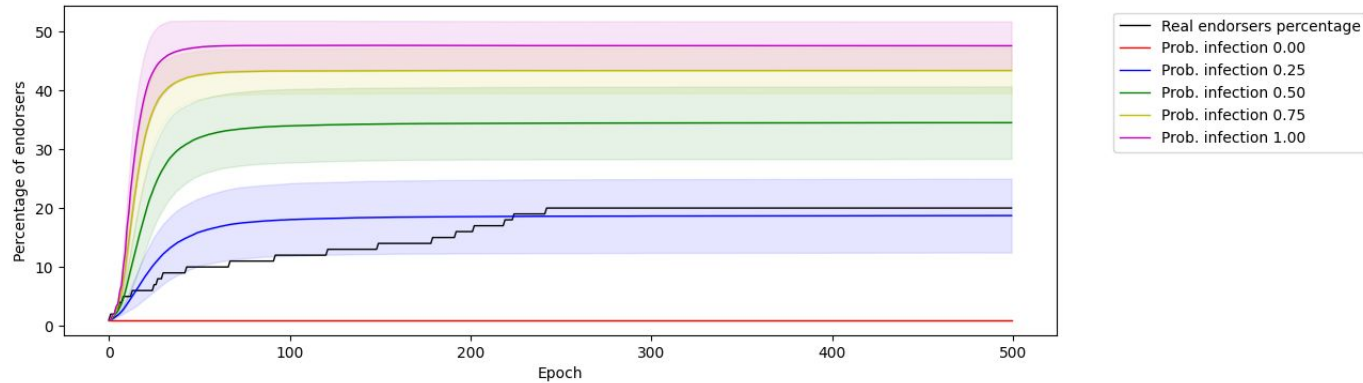
Grafo all'ultima epoca

Validazione - Analisi di sensitività dei parametri

Analisi di sensitività della probabilità di infezione:

P_{inf}	RMSE
0.00	18.70 ± 3.55
0.25	5.21 ± 3.82
0.50	14.89 ± 6.00
0.75	23.64 ± 3.84
1.00	27.86 ± 4.14

RMSE medi



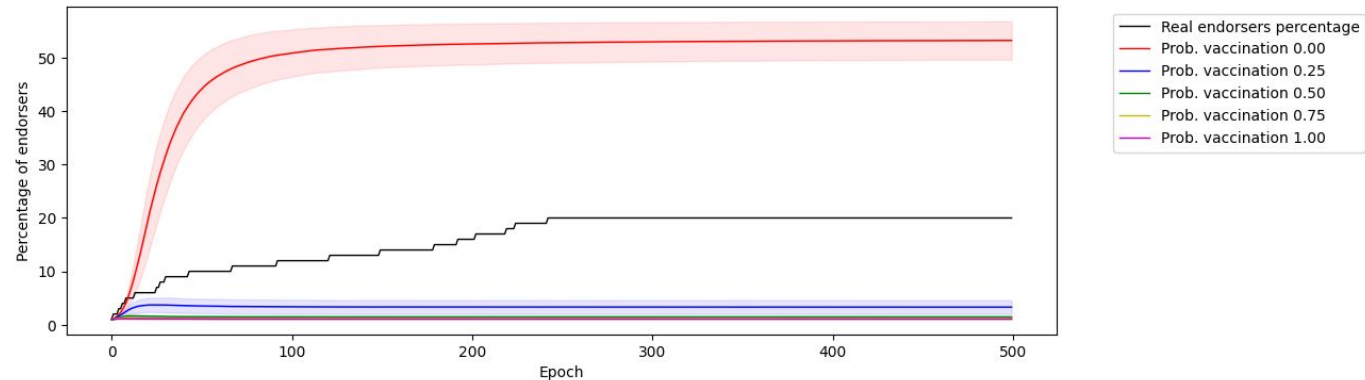
Percentuali medie di infetti per epoca

Validazione - Analisi di sensitività dei parametri

Analisi di sensitività della probabilità di vaccinazione:

P_{vacc}	RMSE
0.00	33.63 ± 3.52
0.25	16.42 ± 1.26
0.50	18.26 ± 0.22
0.75	18.55 ± 0.08
1.00	18.65 ± 0.05

RMSE medi



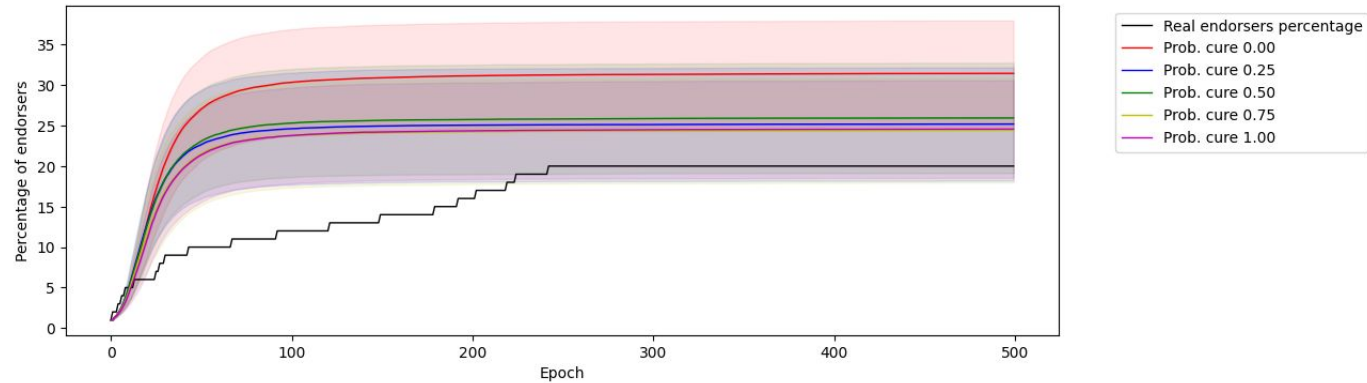
Percentuali medie di infetti per epoca

Validazione - Analisi di sensitività dei parametri

Analisi di sensitività della probabilità di cura:

P_{cure}	RMSE
0.00	12.25 ± 5.70
0.25	8.00 ± 4.16
0.50	8.15 ± 4.55
0.75	7.02 ± 4.04
1.00	6.45 ± 4.50

RMSE medi



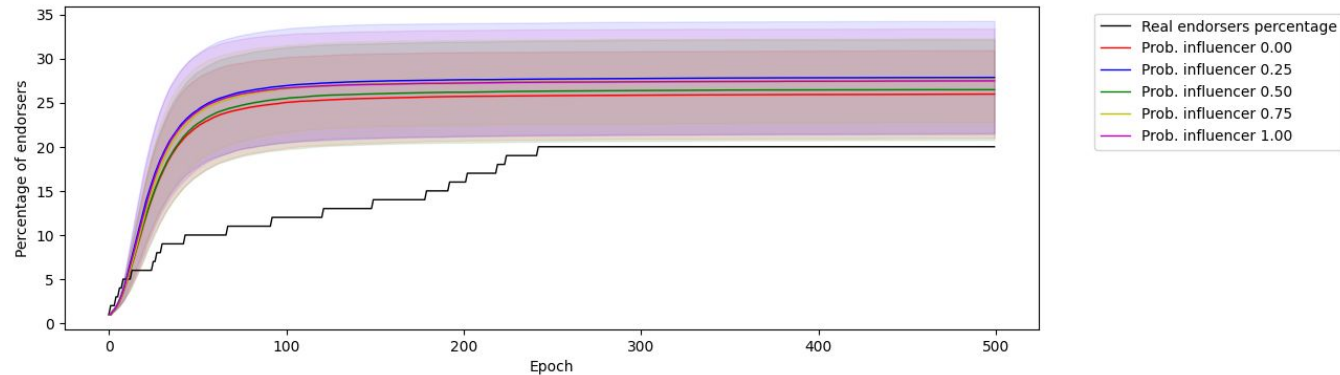
Percentuali medie di infetti per epoca

Validazione - Analisi di sensitività dei parametri

Analisi di sensitività della probabilità degli *influencer*.

$P_{\text{influencer}}$	RMSE
0.00	6.97 ± 4.12
0.25	9.67 ± 3.94
0.50	8.20 ± 3.56
0.75	8.58 ± 3.42
1.00	9.18 ± 3.57

RMSE medi



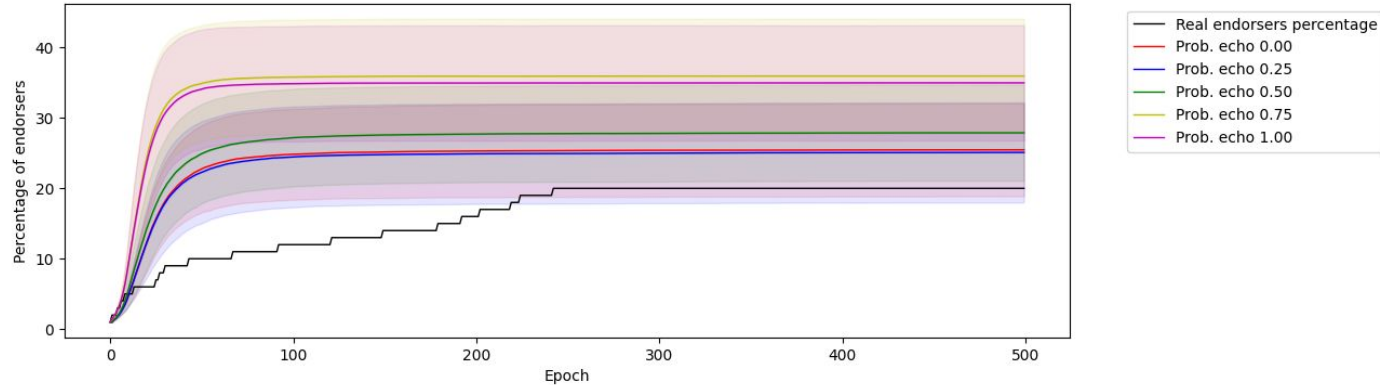
Percentuali medie di infetti per epoca

Validazione - Analisi di sensitività dei parametri

Analisi di sensitività della soglia delle *echo chamber*.

P_{echo}	RMSE
0.00	7.74 ± 4.37
0.25	8.05 ± 4.02
0.50	9.60 ± 4.81
0.75	17.20 ± 5.94
1.00	15.82 ± 7.17

RMSE medi



Percentuali medie di infetti per epoca

The background of the slide is a light gray network pattern. It consists of numerous small circles, some of which are solid gray and others are hollow with a gray outline. These circles are interconnected by a web of thin, light gray lines, creating a complex, organic structure that resembles a molecular or biological network.

Simulazioni

Simulazioni

- ⊙ I modelli calibrati e validati sono stati utilizzati per effettuare **simulazioni** al fine di individuare contromisure efficaci alla diffusione di *fake news* in *social network*.
- ⊙ Per le simulazioni sono stati utilizzati i parametri calibrati con il valore di h pari e 0.50 e i seguenti parametri predeterminati:

nodì	δ_{in}	δ_{out}	epoche	<i>infl.</i>	<i>bot</i>	<i>e.f.c.</i>	infetti iniziali
2000	1	1	1000	20	40	40	100 (con <i>bot</i>)

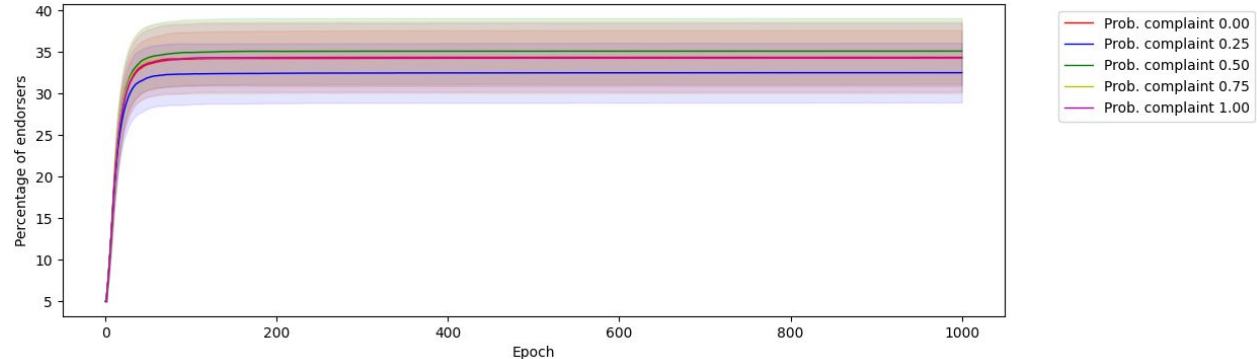
- ⊙ I risultati delle simulazioni sono stati calcolati in termini di percentuali medie di infetti per epoca. La media è stata calcolata su 30 reti.

Simulazioni - Blocco degli utenti

- Sono state svolte simulazioni attivando il **blocco degli utenti** a seguito di **reclami**. Il blocco avviene dopo 3 reclami. Sono state considerate diverse probabilità di presentare reclamo.
- **Risultati:**

$P_{\text{complaint}}$	utenti bloccati
0.00	0.00 \pm 0.00
0.25	0.67 \pm 0.75
0.50	1.63 \pm 1.22
0.75	1.53 \pm 1.36
1.00	2.57 \pm 2.36

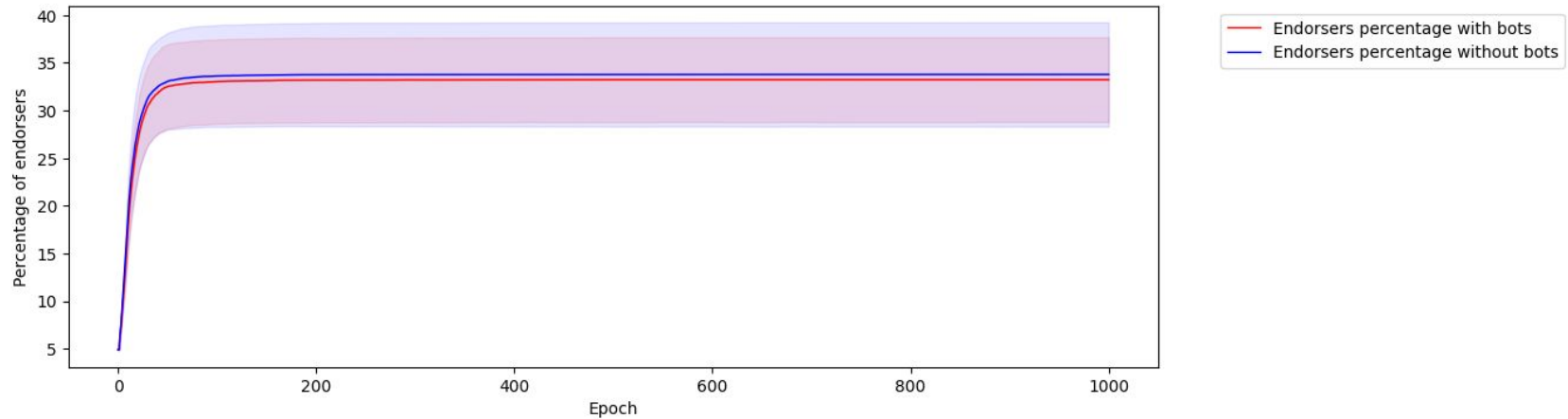
Utenti bloccati medi



Percentuali medie di infetti per epoca

Simulazioni - Rimozione dei *bot*

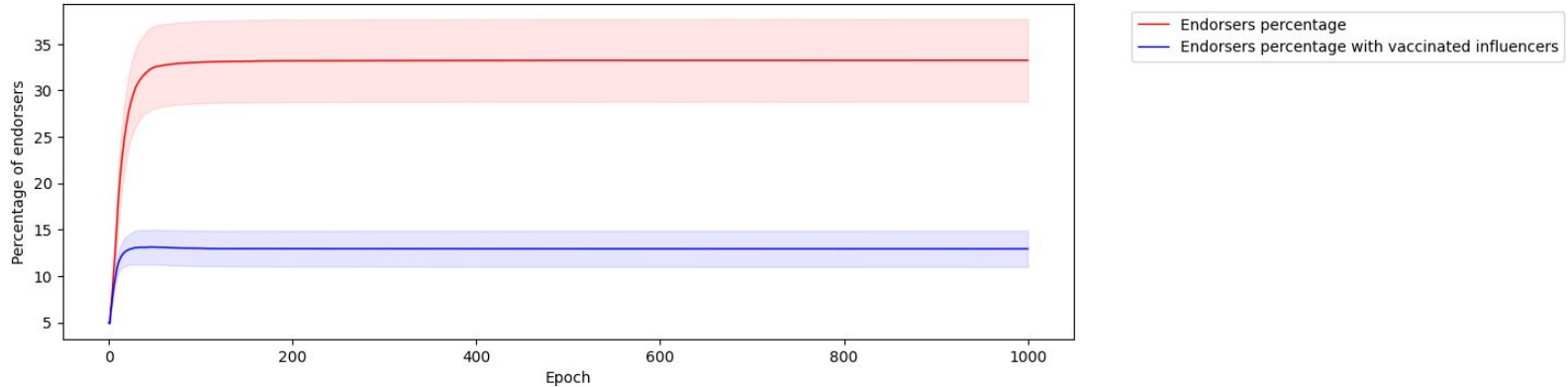
- Sono state svolte simulazioni **rimuovendo i *bot*** e sostituendoli con normali utenti infetti.
- **Risultati:**



Percentuali medie di infetti per epoca

Simulazioni - Vaccinazione degli *influencer*

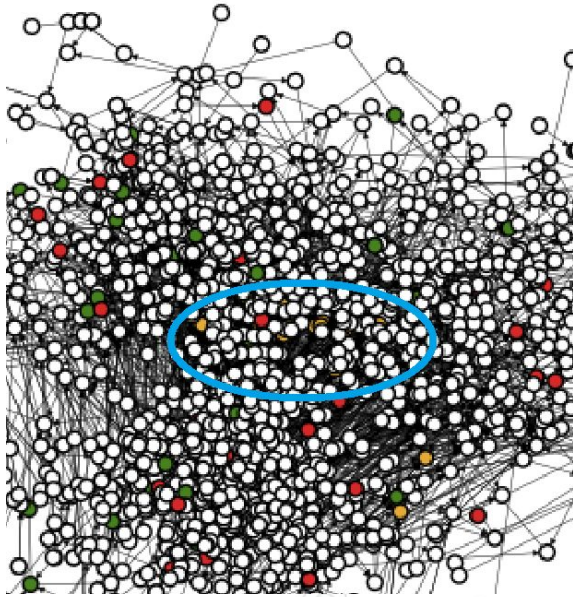
- Sono state svolte simulazioni **vaccinando gli *influencer***.
- **Risultati:**



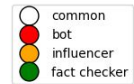
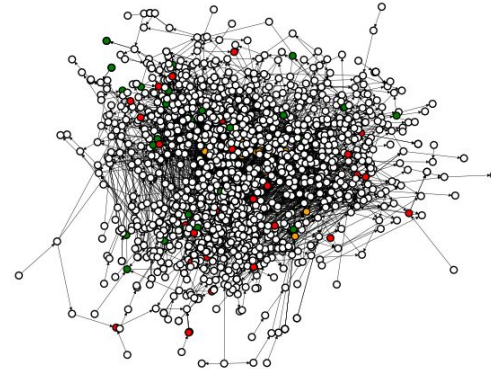
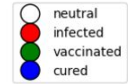
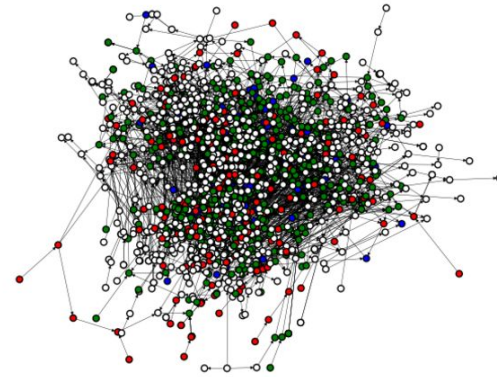
Percentuali medie di infetti per epoca

Simulazioni - Vaccinazione degli *influencer*

◎ Grafo all'ultima epoca:

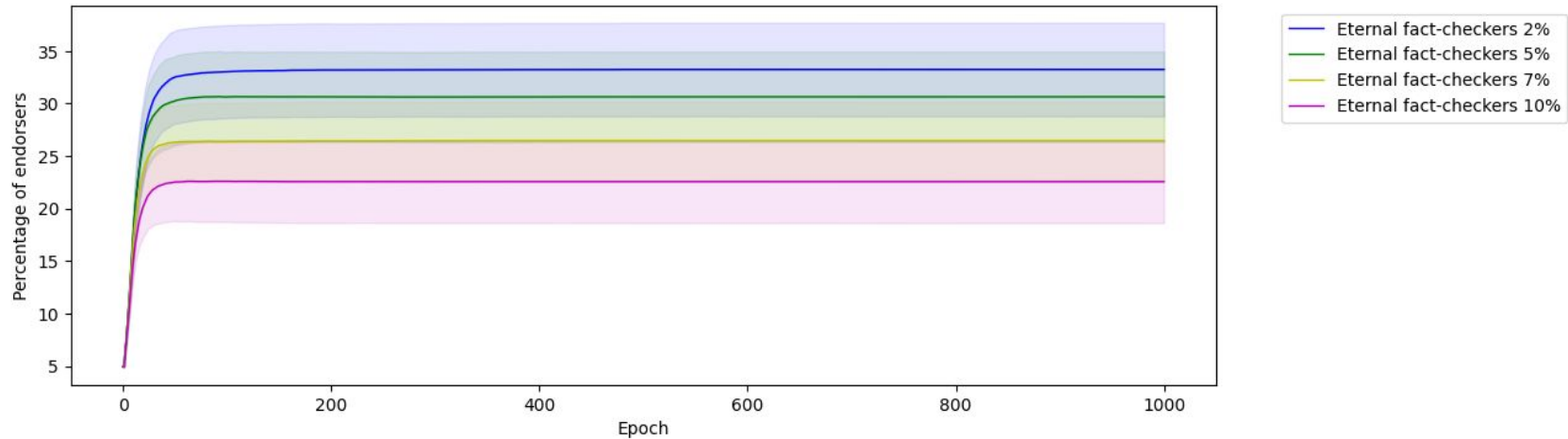


Dettaglio del grafo generato
all'ultima epoca



Simulazioni - Incremento degli *eternal fact-checker*

- Sono state svolte simulazioni **incrementando il numero di *eternal fact-checker***.
- **Risultati:**



Percentuali medie di infetti per epoca

The background of the slide is a light gray network pattern. It consists of numerous small circles, some of which are solid gray and others are hollow with a gray outline. These circles are interconnected by a web of thin, light gray lines, creating a complex, organic structure that resembles a molecular or digital network.

Conclusioni e sviluppi futuri

Conclusioni e sviluppi futuri

◎ Conclusioni sulle contromisure:

- Il blocco degli utenti e la rimozione dei *bot* sono inefficaci.
- L'incremento degli *eternal fact-checker* è efficace ma costoso.
- La vaccinazione degli *influencer* ottiene (i) le maggiori riduzioni di infetti; (ii) con una varianza piccola; (iii) rapidamente; e (iv) con un basso impatto del relativo parametro di probabilità.

◎ Sviluppi futuri:

- Evoluzione dinamica della rete sociale.
- Considerazione di altre relazioni come il *retweet*.
- Utilizzo di *dataset* migliori e migliore calibrazione dei parametri.
- Confronto con la diffusione di *fake news* attraverso i media tradizionali.



Grazie per l'attenzione

Riferimenti bibliografici

- [1] Barabási, A.-L. *Network science*. Cambridge University Press, Cambridge, 2016.
- [2] Serrano, E., and Iglesias, C. A. Validating viral marketing strategies in twitter via agent-based social simulation. *Expert Systems with Applications* 50 (2016), 140–150.
- [3] Lotito, Q. F., Zanella, D., and Casari, P. Realistic aspects of simulation models for fake news epidemics over social networks. *Future Internet* 13 (2021).
- [4] Russell, S. J., and Norvig, P. *Artificial Intelligence: A Modern Approach (4th Edition)*. Pearson, 2020.
- [5] Qazvinian, V., Rosengren, E., Radev, D. R., and Mei, Q. Rumor has it: Identifying misinformation in microblogs. In *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing* (Edinburgh, Scotland, United Kingdom, 2011), Association for Computational Linguistics, pp. 1589–1599.