

Analisi della topologia e dei contenuti delle comunità virtuali per lo studio del fenomeno delle camere d'eco

Relatore: Dr. Marco Viviani

Correlatrice: Prof.ssa Gabriella Pasi

Tesi di Laurea Magistrale

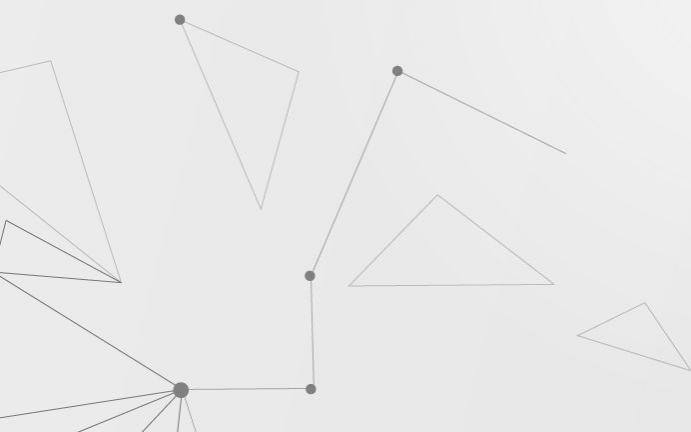
Giacomo Villa 807462

Anno Accademico 2019-2020



Contesto
applicativo e
camere d'eco

Stato dell'arte



01

02

Panoramica

03

La soluzione
proposta

04

Risultati del lavoro

05

Possibili sviluppi
futuri

01

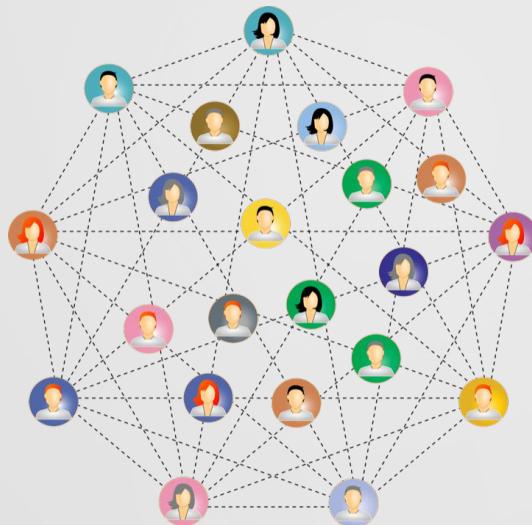
Contesto applicativo e camere d'eco

"Siamo tutti online ma non connessi."
Aran Consentino



Contesto applicativo: i *social media*

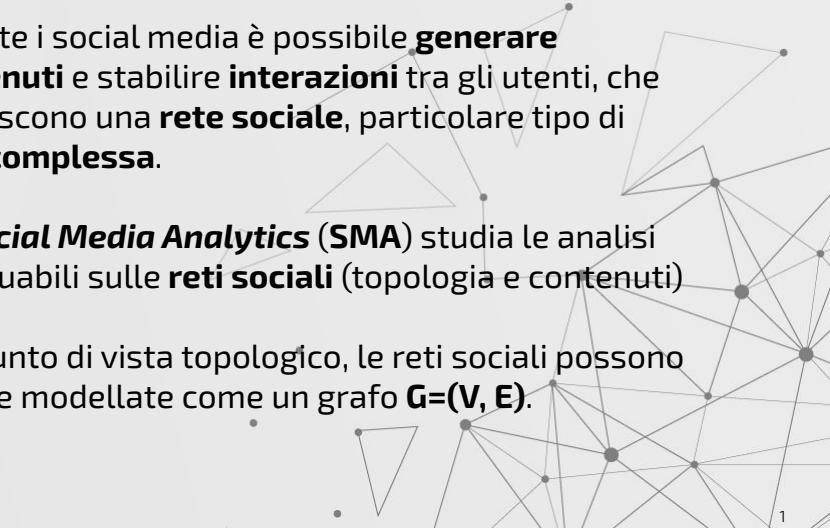
I **social media** sono applicazioni basate sulle tecnologie del Web 2.0. Sono diventati strumenti di comunicazione capillari, estensione nella nostra identità in chiave virtuale.



Tramite i social media è possibile **generare contenuti** e stabilire **interazioni** tra gli utenti, che definiscono una **rete sociale**, particolare tipo di **rete complessa**.

La **Social Media Analytics (SMA)** studia le analisi effettuabili sulle **reti sociali** (topologia e contenuti)

Dal punto di vista topologico, le reti sociali possono essere modellate come un grafo $G=(V, E)$.

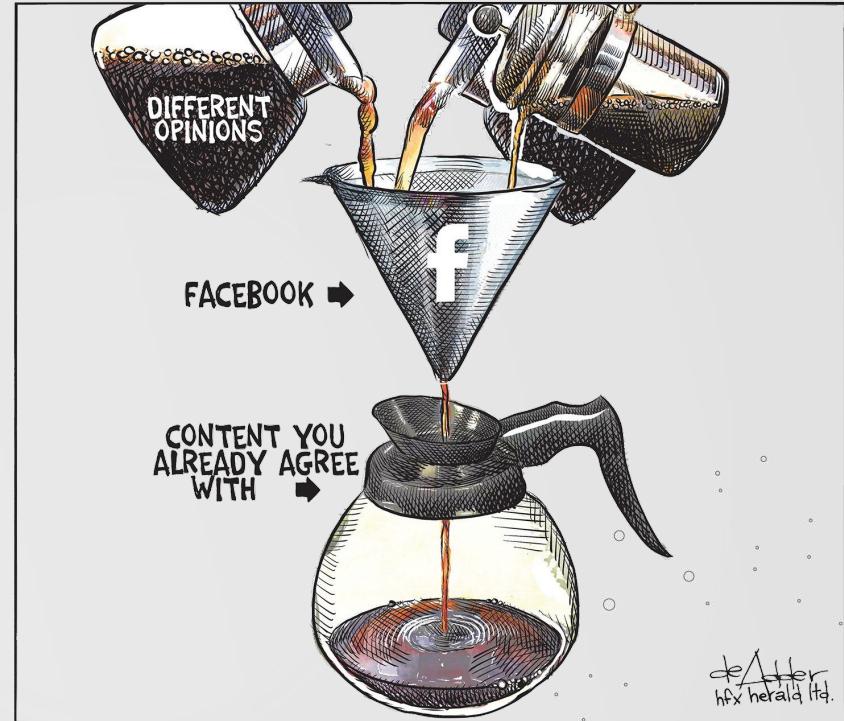


Il problema delle camere d'eco

Realtà in cui certe **informazioni**, **idee o credenze** vengono **amplificate** dalla loro **continua ripetizione** all'interno di un sistema chiuso.

Il **contenuto consumato** dall'utente è **fortemente influenzato** dall'opinione dello stesso e da quella dei suoi "simili".

Caratteristiche tecnologiche e **psicologiche/sociologiche** contribuiscono alla loro formazione.



02

Stato dell'arte

"Non umile dinnanzi alla vita, ma umile dinnanzi all'arte!" Gabriele D'

Annunzio



Le possibili strategie

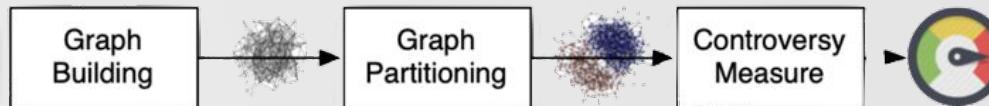
In una comunità virtuale vengono a crearsi solitamente **gruppi polarizzati di utenti** rispetto a determinati *topic* di interesse.

Al fine di superare (o prevenire) tali polarizzazioni, si possono implementare due strategie distinte:

- Agire a monte, prevenendo le **bolle di filtraggio** (*filter bubbles*).
- Agire a valle, operando sulle **camere d'eco** (*echo chambers*).



Come identificare le camere d'eco



Diversi stage proposti negli studi di Garimella et al. (2017)

Al fine di verificare la presenza di **camere d'eco**, viene definita una serie di "step" che insieme vanno a formare una **pipeline di identificazione**.



Costruzione del grafo

Basandosi sulle **interazioni tra utenti** dato un certo *topic*.

Attività di **Community detection tradizionale** con **valutazione della semantica**.



Misurazione controversia

Utilizzo di **metriche o valutazioni automatiche**.



Partizionamento del grafo

Come mitigare il fenomeno delle camere d'eco



Valutazioni a grana fine

Web widget per guidare l'utente.

Prevenzione in funzione del contenuto generato.

Strumenti per i moderatori



Link prediction

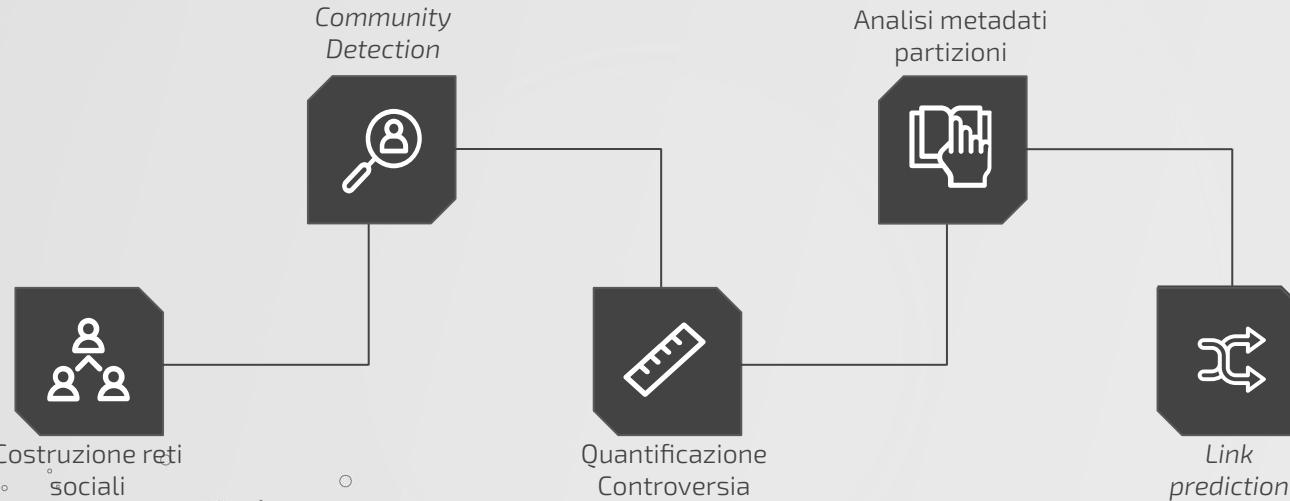
Utilizzo di metriche o valutazioni automatiche.

03

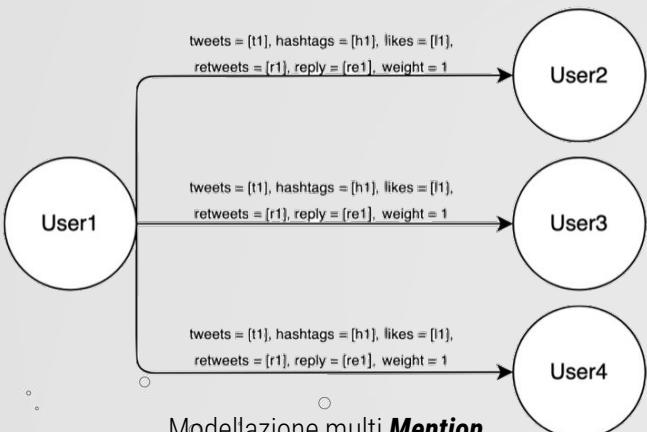
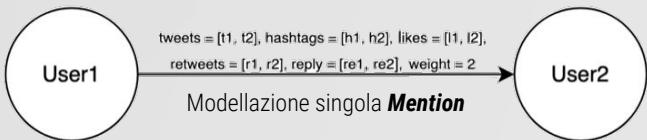
La soluzione proposta

"Laddove gli esseri umani creano problemi, gli stessi esseri umani possono arrivare a delle soluzioni." Carl Sagan

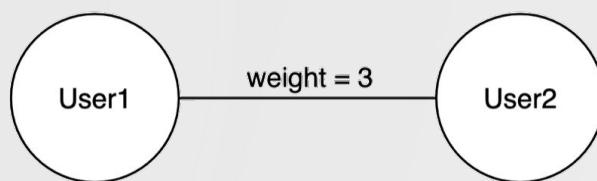
La pipeline implementata



Modellazione delle reti sociali



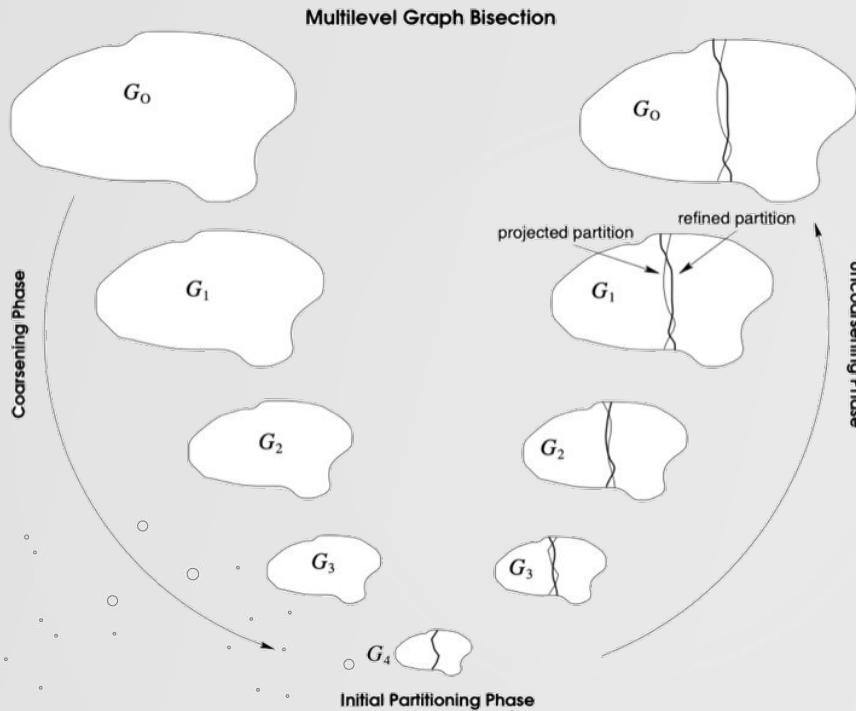
Modellazione dei dati facenti
riferimento a **Covid-19** e
Vaccination



Modellazione dati allo stato dell'arte



Community Detection (Topologia)



L'algoritmo utilizzato per il
partizionamento del
grafo → METIS

Processo diviso in 3 step:

- **Compattazione**
- **Taglio**
- **Scompattazione**

Community Detection (Semantica)

Sentiment (VADER)

- Valutazione *sentiment* medio per ogni utente e riscaling (-30, 30)
- Valutazione $SS(user1, user2) = 60 - |S(user1) - S(user2)|$
- Definizione $SW(user1, user2) = 1 + SS(user1, user2)$

Topic (LDA MALLET)

- Valutazione, mediante *Topic Coherence*, numero ottimale topic
- Assegnazione *topic* per ogni utente
- Valutazione $TS(user1, user2) = \#topic - card((topic(user1) \Delta (topic(user2)))$
- Definizione $TW(user1, user2) = 1 + TS(user1, user2)$

Valutazione informazione semantica

Definizione di nuovi pesi in funzione del *sentiment*,
topic e *ibrido*

Ibrido

- Definizione $HW(user1, user2) = SW(user1, user2) + TW(user1, user2)$

Valutazione controversia



Random Walk Controversy

$$RWC = P_{XX}P_{YY} - P_{XY}P_{YX}$$



Random Walk Top Degree Controversy

$$RWC_{Top} = P_{XX}P_{YY} - P_{XY}P_{YX}$$



Change Side Controversy

$$CSC = \frac{\sum_{\forall v \in N} 1 - (\frac{\kappa(v)}{l_G * 2})}{|N|}$$



GMCK

$$GMCK = \frac{1}{|B|} \sum_{u \in B} \frac{d_i(u)}{d_b(u) + d_i(u)} - 0.5$$

Link Prediction



Coefficiente di Jaccard

$$JC_{(x,y)} = \frac{\Gamma(x) \cap \Gamma(y)}{\Gamma(x) \cup \Gamma(y)}$$



Adamic Adar Index

$$AAI_{(x,y)} = \sum_{z \in \Gamma(x) \cap \Gamma(y)} \frac{1}{\log|\Gamma(z)|}$$



Top Degree Index

$$TDI_{(x,y)} = \text{Degree}(x) + \text{Degree}(y)$$



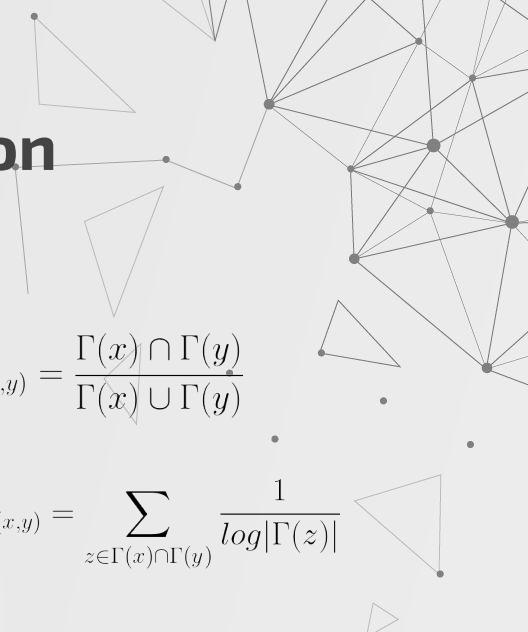
Top to Normal Index

$$TNI_{(x,y)} = \text{Degree}(x)$$



Top Betweenness Index

$$TBI_{(x,y)} = \text{Bet}(x) + \text{Bet}(y)$$



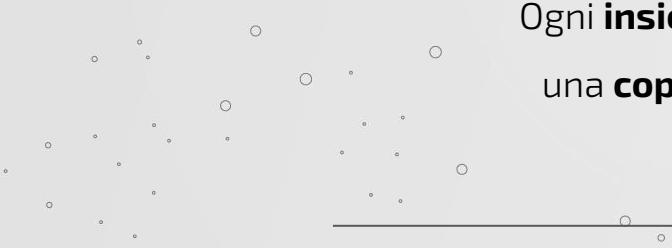
Sentiment Boost



Nel tentativo di modellare una "**probabilità di accettazione**" del collegamento.

$$SB(x, y) = \frac{|E_{SS}(x, y)|}{|E|}$$

Ogni **insieme** definito precedentemente avrà una **copia con valutazione** del **Sentiment Boost**.



04

Risultati del lavoro

*"Potrai non sapere mai i risultati prodotti dalle tue azioni, ma,
se non fai nulla, non ci saranno risultati."*

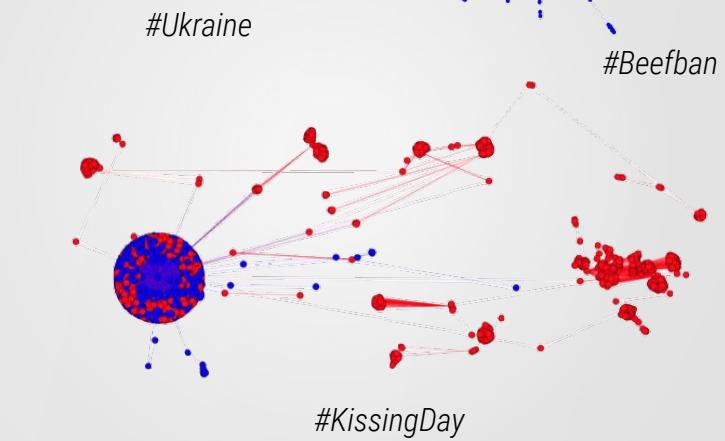
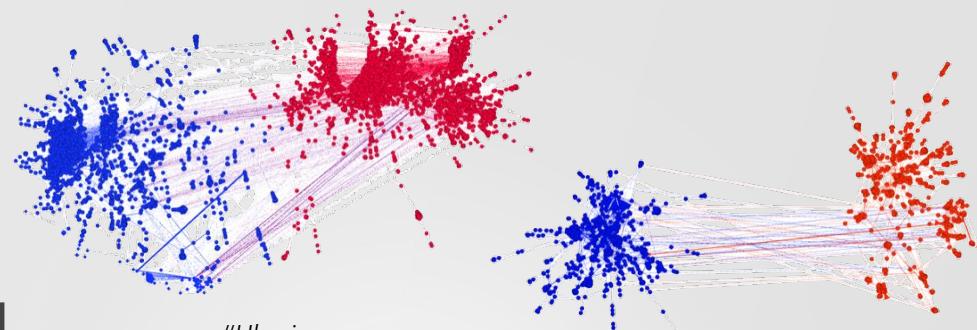
Mahatma Gandhi



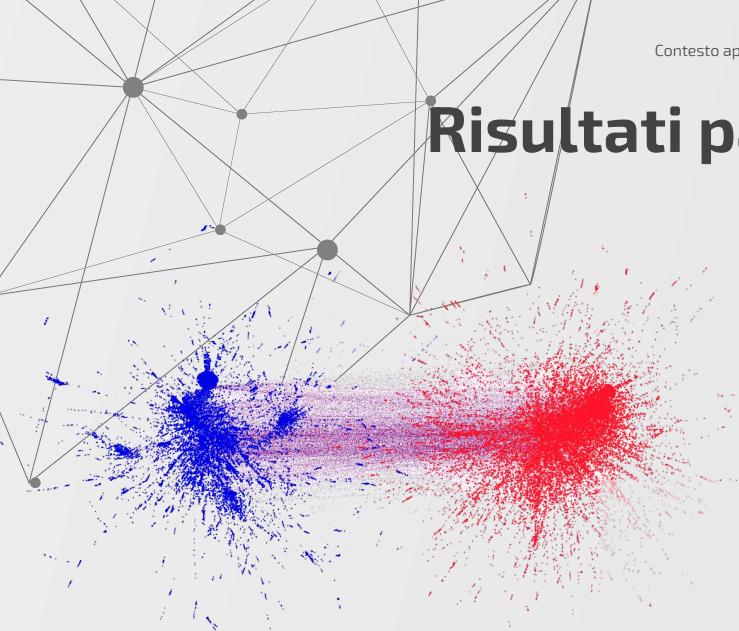
Risultati partizionamento reti (1/2)



	Community 0	Community 1	Modularità	Coverage
#Ukraine	2710	2785	0.4685	0.9697
#BeefBan	824	786	0.4758	0.9798
#KissingDay	2388	2250	0.2608	0.7747



Risultati partizionamento reti (2/2)

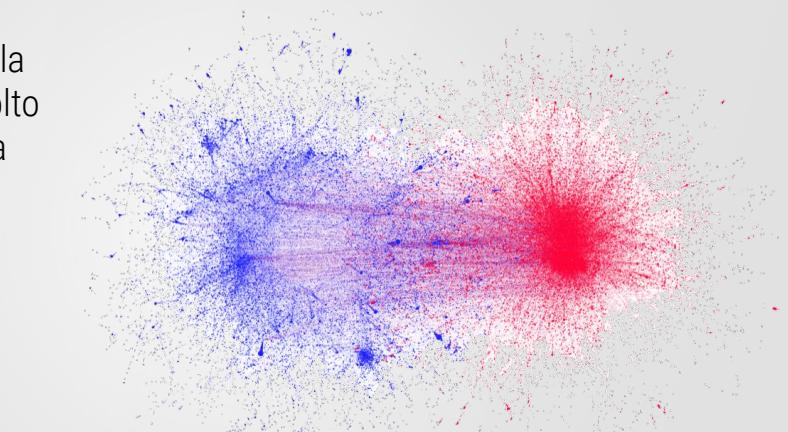


Contrapposizione tra le due **community** identificate con approccio basato sul **sentiment - Covid-19**

Covid-19

	Community 0	Community 1	Modularità	Coverage	Cambi
Strutturale	19096	20240	0.4348	0.9351	0
Sentiment	19163	20173	0.4403	0.9534	6297
Topic	19094	20242	0.4396	0.9403	6920
Ibrido	20240	19096	0.4322	0.9224	0.9224

Contrapposizione tra le due **community** identificate con approccio basato sul **sentiment - Vaccination**

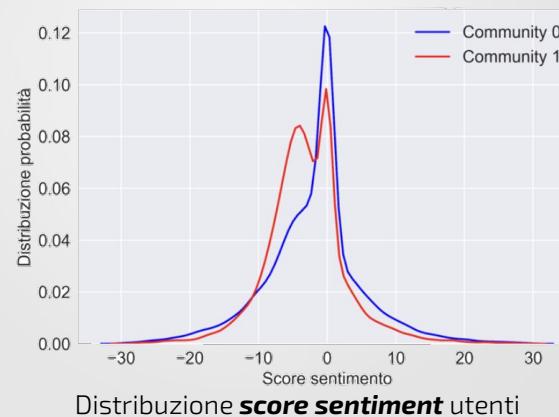
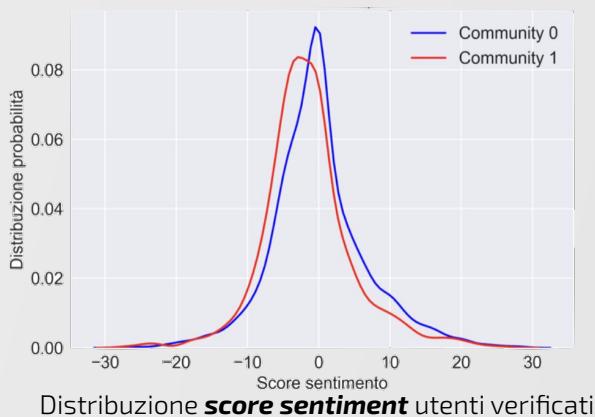


	Vaccination				
	Community 0	Community 1	Modularità	Coverage	Cambi
Strutturale	19808	19796	0.2953	0.9837	0
Sentiment	19525	20079	0.2935	0.9853	35945
Topic	20061	19543	0.2907	0.9853	4169
Ibrido	19970	19634	0.2912	0.985	3776

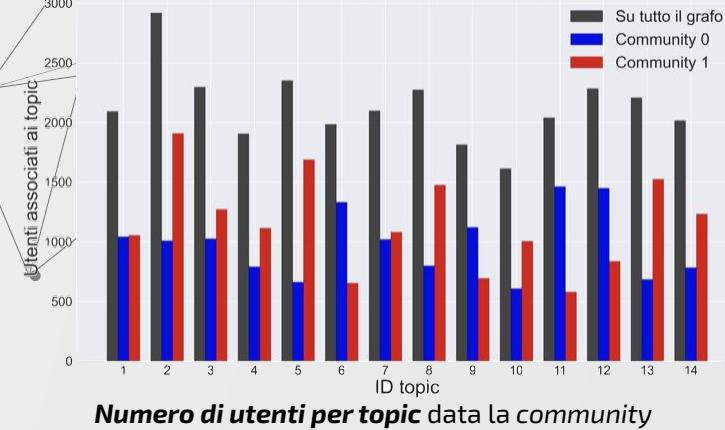
Analisi metadati - Covid 19 (1/2)



Wordcloud date le **descrizione dei profili** degli utenti in funzione della **community**



Analisi metadati - Vaccination (2/2)



Numero di utenti per topic data la community

Vaccineswork, Health, Public Health,
Important...

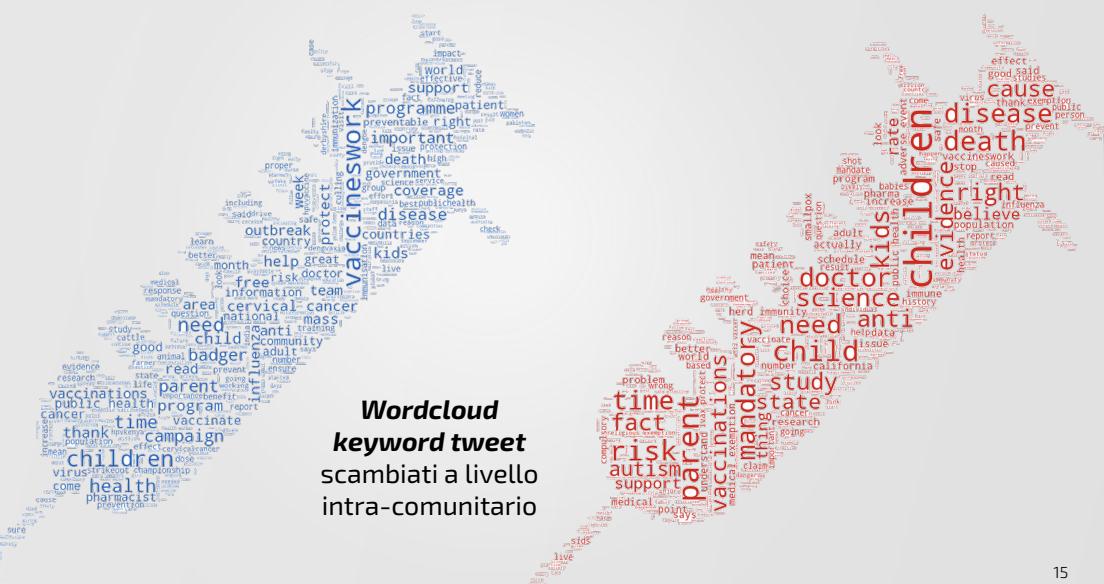
...Risk, Cause, Disease, Autism

Topic 5: *forc, mandatori, stop, peopl, choic, make, pharma, mandat, wrong, american*

Topic 6: *year, polio, prevent, cancer, campaign, vaccineswork, start, million, free, girl*

Topic 8: *caus, kid, child, autism, babi, die, kill, happen, harm, ill*

Topic 12: *health, care, commun, provid, patient, famili, recommend, includ, improv, servic*



Valutazione score controversia

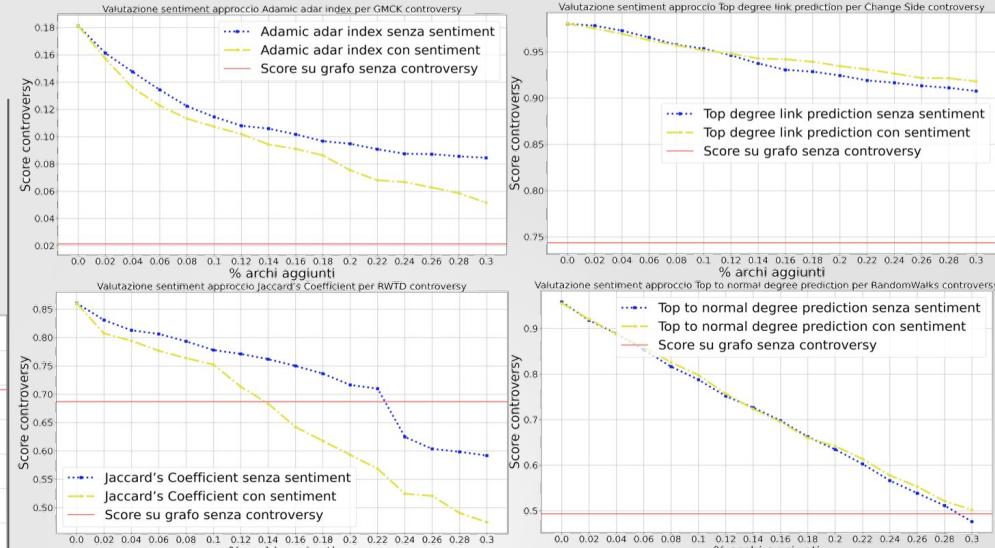
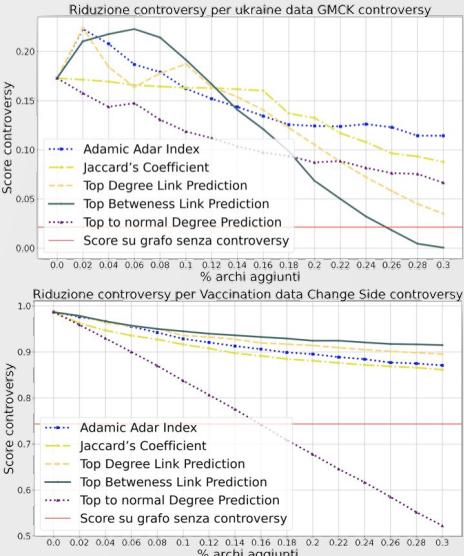
	RWC	GMCK	RWTD	CSC
#Ukraine	0.986	0.2132	0.8583	0.9895
#BeefBan	0.9624	0.1729	0.8889	0.9837
#KissingDay	0.4904	0.0213	0.6875	0.7427
Covid-19	0.9533	0.1813	0.854	0.9807
Vaccination	0.953	0.235	0.8616	0.9871



Link Prediction

Approcci basati sulla **centralità** sembrano registrare **riduzioni più marcate**

In particolare l'approccio **Top to Normal** sembra **eccellere**.



Il **Sentiment Boost** migliora principalmente i risultati nei *linking* basati sulla prossimità.

Analisi post linking (1/3) - #BeefBan & #Ukraine

	Modularità	Coverage	ΔModularità	ΔCoverage
Jaccard's	0.4611	0.9737	-0.0147	-0.0061
Top to Normal	0.3523	0.8551	-0.1235	-0.1247

#BeefBan 1484 nuovi archi aggiunti.

...approcci basati sulla prossimità chiaramente meno performanti!

Metodologie basate sulla centralità indeboliscono in maniera significativa le strutture interne...

#Ukraine 8774 nuovi archi aggiunti.

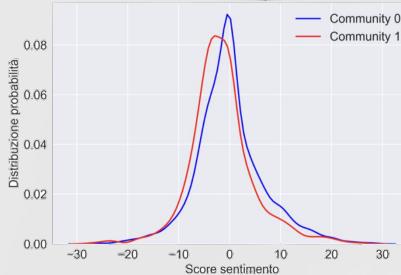
	Modularità	Coverage	ΔModularità	ΔCoverage
Jaccard's	0.4541	0.9742	-0.0144	-0.0045
Top to Normal	0.3604	0.8605	-0.1081	0.1092

Analisi post linking (2/3) - Covid-19

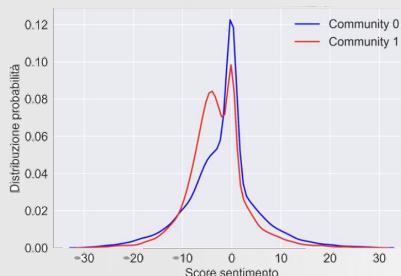


Con Sentiment Boost

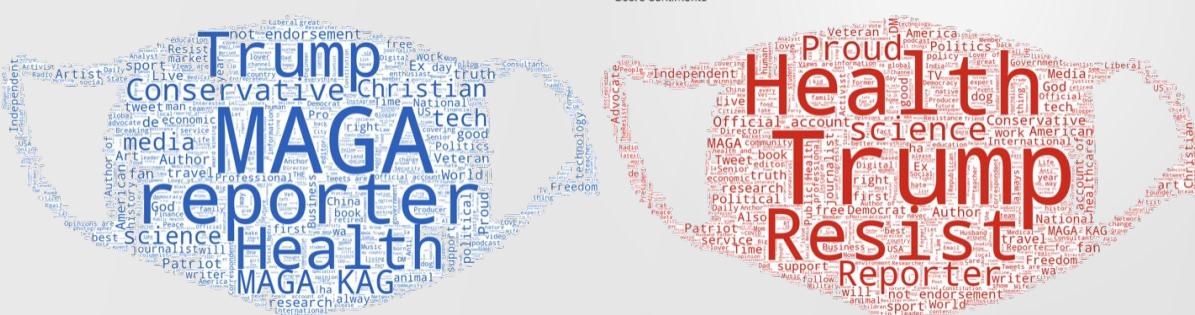
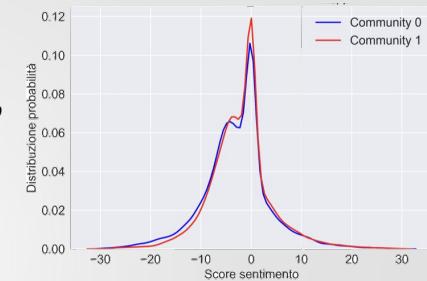
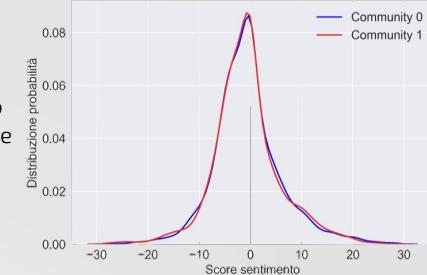
	Modularità	Coverage	Δ Modularità	Δ Coverage
Jaccard's	0.4327	0.932	-0.0076	-0.0084
Top to Normal	0.3023	0.8033	-0.138	-0.1371



Sentiment medio
utenti **verificati** pre
e post linking

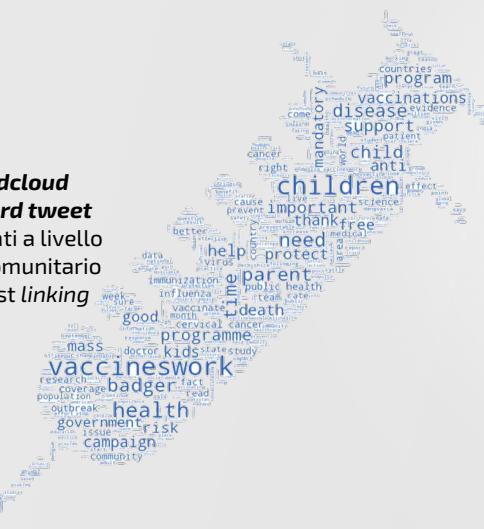
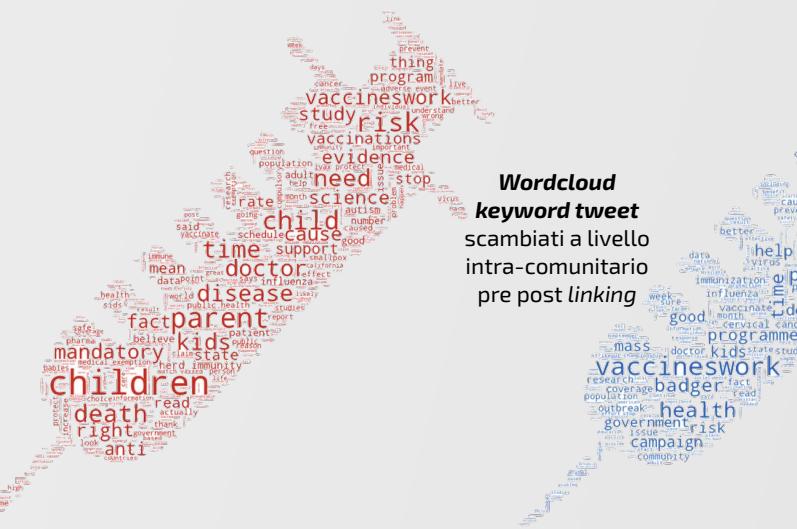
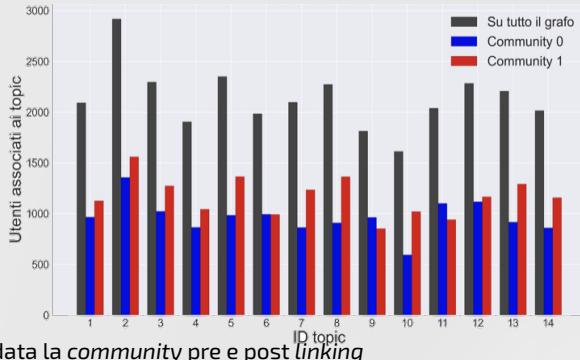
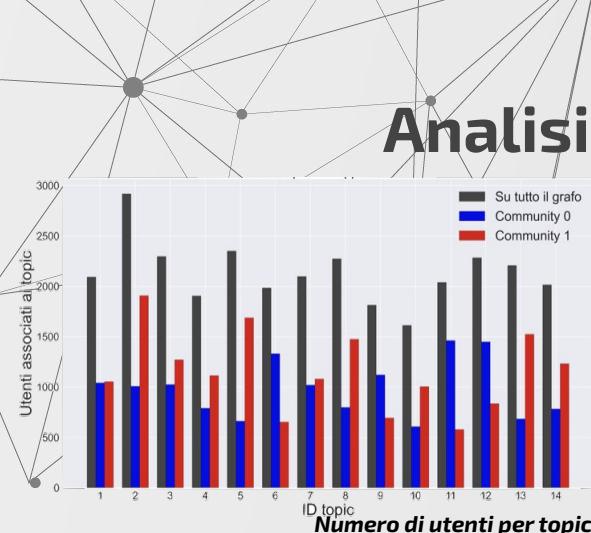


Sentiment medio
utenti pre e post
linking



Wordcloud date le
descrizione dei
profili degli utenti in
funzione della
community post
linking

Analisi post linking (3/3) - Vaccination



	Modularità	Coverage	ΔModularità	ΔCoverage
Jaccard's	0.355	0.9702	0.0615	-0.015
Top to Normal	0.2667	0.8938	-0.0268	-0.0914

	Modularità	Coverage	ΔModularità	ΔCoverage
Jaccard's	0.3466	0.9647	0.0531	-0.0206
Top Bet	0.3588	0.937	0.0452	-0.0482

05

Possibili sviluppi futuri

"Lo sviluppo è un viaggio con molti più naufraghi che navigatori."
Eduardo Galeano

Possibili sviluppi futuri



Superamento bipartizione

Partizionamento più a grana fine per **identificare più punti di vista** sul topic.



Upgrade *Sentiment Boost*

Utilizzo di più **metadati** e informazioni **per** definire il **moltiplicatore**



Tratti della personalità

Valutazione del contenuto semantico condiviso per **definire** la **personalità dominante**.



Hybrid Link prediction

Definizione di **score** che tengano in considerazione **diversi aspetti**.



A large, abstract network graph is positioned in the upper right corner of the slide. It consists of numerous small, dark grey dots representing nodes, connected by a complex web of thin, light grey lines representing edges. The graph is highly interconnected, forming several distinct clusters or communities. In the lower right area, there is a more dense cluster of nodes and edges.

**GRAZIE PER
L'ATTENZIONE**

Ci sono domande?