

EFFETTO REDDIT

ANALISI DELLE DIFFERENZE STRUTTURALI E
COMPORTAMENTALI TRA CINQUE DIVERSI
SUBREDDIT



Teresa Cigna - 813925

Chiara Di Domenico - 815463

Federico De Servi - 812166

Obiettivo

Reddit è un sito Internet di social news, intrattenimento e forum, dove ciascun iscritto (redditor) può creare contenuti testuali o ipertestuali in una serie di subreddit specifici per ciascun argomento.

Vengono considerati **5 subreddit**:

- **r/Politics**
- **r/Coronavirus**
- **r/Science**
- **r/Television**
- **r/Gaming**





- Vi sono **differenze strutturali** tra i subreddit relativi ad argomenti diversi?
- Come circola **l'informazione** all'interno di diversi subreddit?
- Le **interazioni** tra utenti sono principalmente **positive o negative** ? Possiamo notare particolari differenze tra i subreddit?
- Come possono essere spiegate tali differenze?

+

•

Argomenti - NER

Si è deciso di esaminare, per ogni subreddit, i titoli e il corpo dei post analizzati in questo progetto in modo da poter essere più consapevoli nell'interpretazione dei risultati.

Nello specifico si è deciso di utilizzare la **Named Entity Recognition** per poter estrarre le entità nominate.



Argomenti - NER

- **Science:** salute in Germania (1890); cambiamento climatico e inquinamento (1723); risk aversion (467); adolescenza, istruzione e carriera (355); ossigeno su Marte (207); fotoni, quanti (138); felicità (135); remdesivir (antivirale) (48); osteoartrite (18); media (15);
- **Politics:** Donald Trump, Joe Biden, Elezioni, Wisconsin (4331);
- **Television:** Amazon Studios, Pulp Fiction, Lord of the Rings (1165); John Mulaney (1086); Star Wars (445); Criminal Minds (245); Superstore (160); Black Sails (121); Game of Thrones (119); Hawkeye (65); Jason Sudeikis (Ted Lasso) (23); Cinderella (19); Battlebots (13);
- **Coronavirus:** parallelismo morti Coronavirus e 11/9/2001 (4263); Coronavirus (622); vaccino (144); texas; virologia, zoonosi (127);
- **Gaming:** Game of the Year (458); Percezione dei videogiochi tra genitori e figli (377); Final Fantasy VII (368); Minecraft (233); Dark Souls (52); meme (7);



TOPICS

1) RACCOLTA DATI E PROCESSING

2) RETI SOCIALI

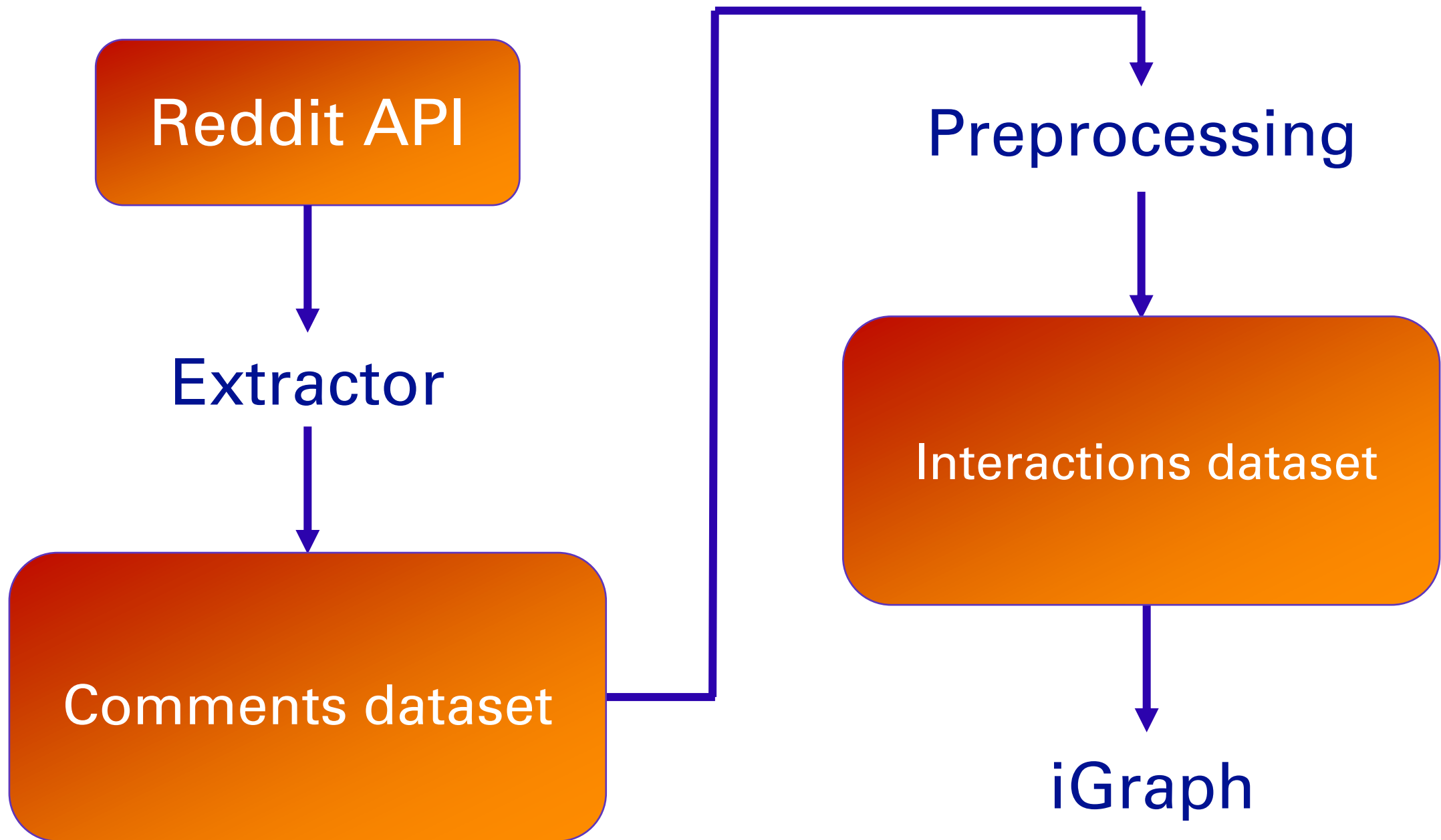
3) COMMUNITY DETECTION

4) SENTIMENT ANALYSIS

5) CONCLUSIONI



RACCOLTA DATI E PREPROCESSING



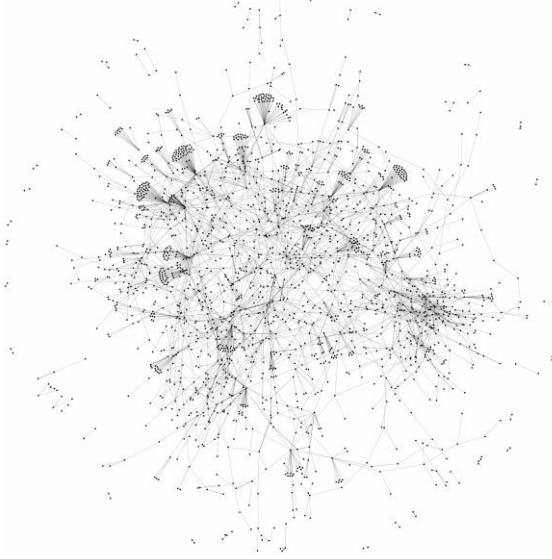
RETI SOCIALI



r/gaming



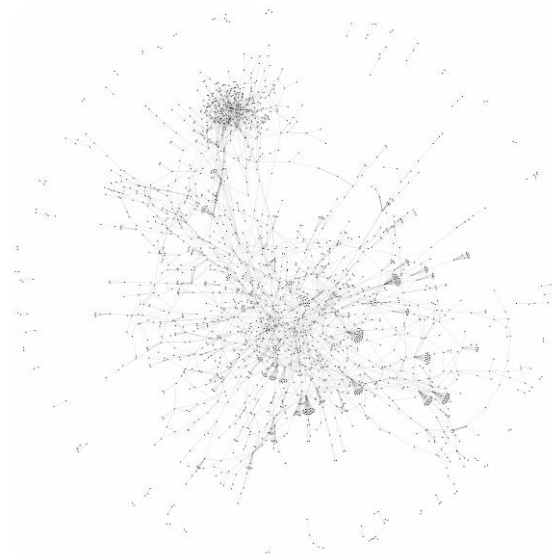
r/coronavirus



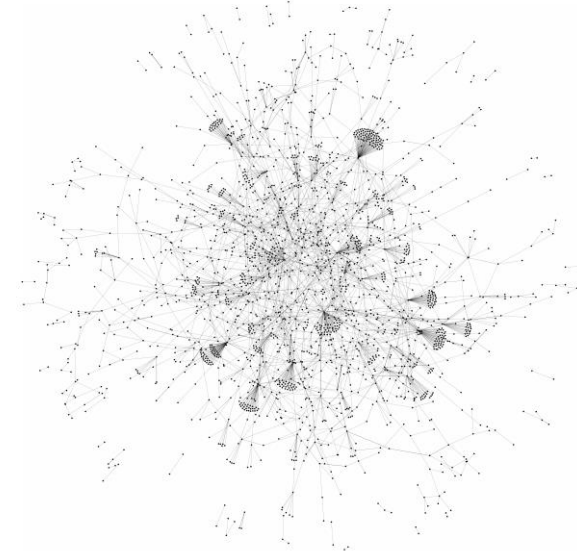
r/science



r/television



r/politics



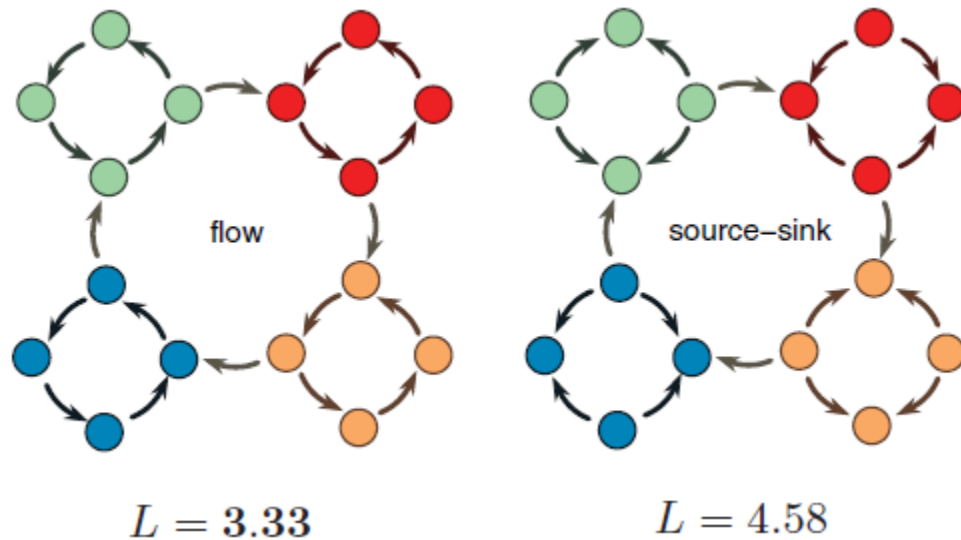
	Gaming	Coronavirus	Science	Television	Politics
Assortativity	-0,16	-0,19	-0,17	-0,14	-0,16
# Authority score > 0.5	1	1	1	1	1
# Hub score > 0.5	1	3	55	1	94
Diameter	23	40	27	41	48
Reciprocity	0,32	0,29	0,33	0,25	0,22

COMMUNITY DETECTION



Il metodo Infomap

Algoritmo di **community detection** che, minimizzando un'equazione chiamata Map Equation, forma communities in cui **l'information flow è massimizzato**.



Risultati



Gaming - 194		Coronavirus - 256		Science - 217		Television - 306		Politics - 286	
# Users	Diameter	# Users	Diameter	# Users	Diameter	# Users	Diameter	# Users	Diameter
81	4	117	7	49	3	68	4	127	3
73	4	65	8	46	8	60	6	97	7
66	7	56	3	43	4	46	5	86	5
44	5	50	6	41	3	45	6	76	5
31	6	45	2	41	5	45	3	73	9

SENTIMENT ANALYSIS

Tecniche

Tecniche di sentiment utilizzate:

- AFINN
- NRCLEX
- Bing-Liu
- TextBlob
- IBM WNLU
- VADER

Preprocessing +

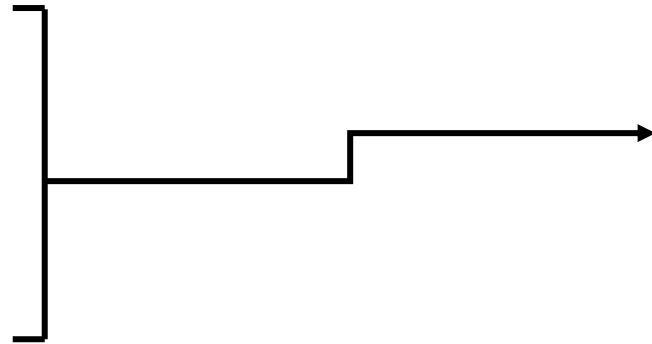
Preprocessing generico:

- Sostituzione termini ingannevoli
- Tokenizzazione
- Eliminazione link e simboli

Tecniche

Tecniche di sentiment utilizzate:

- AFINN
- NRCLEX
- Bing-Liu
- TextBlob
- IBM WNLU
- VADER



Preprocessing +

Preprocessing specifico:

- Lemmatizzazione
- Rimozione stopwords
- Gestione delle negazioni

Valutazione

Dopo aver valutato manualmente 60 commenti di test per ogni subreddit (-1,+1), si sono confrontati i risultati ottenuti dalle varie tecniche di sentiment, calcolando **l'F1 score**. I risultati hanno visto IBM performare meglio, seguito da AFINN.

Le tecniche che sono state quindi utilizzate per il progetto sono state:

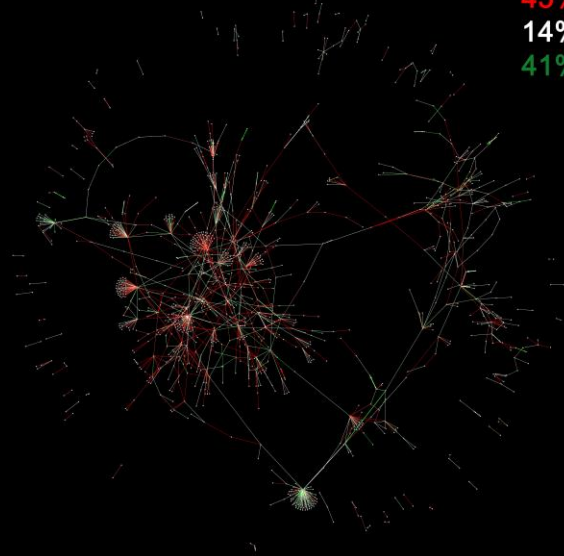
- **IBM** (peso: 0.75)
- **AFINN** (peso: 0.25)

+



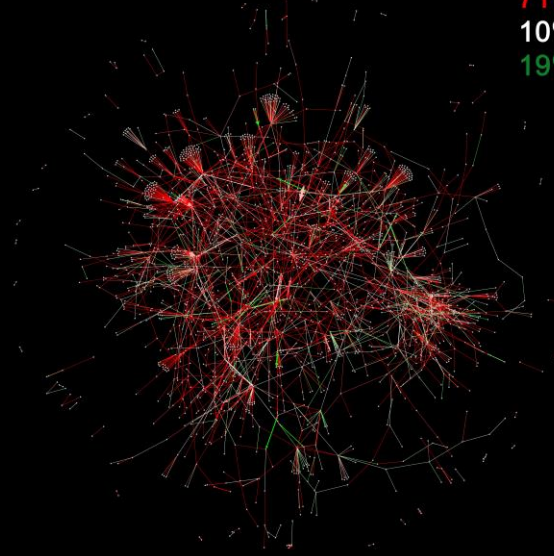
r/gaming

45%
14%
41%



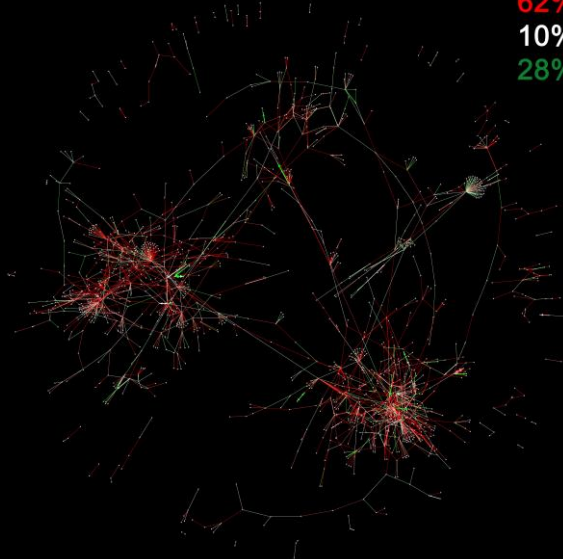
r/coronavirus

71%
10%
19%



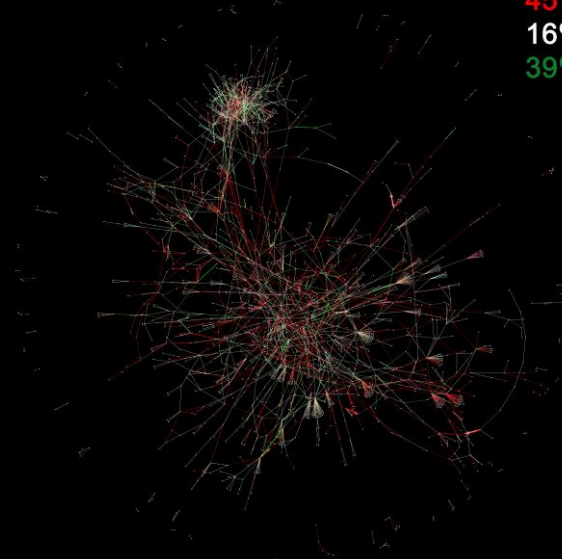
r/science

62%
10%
28%



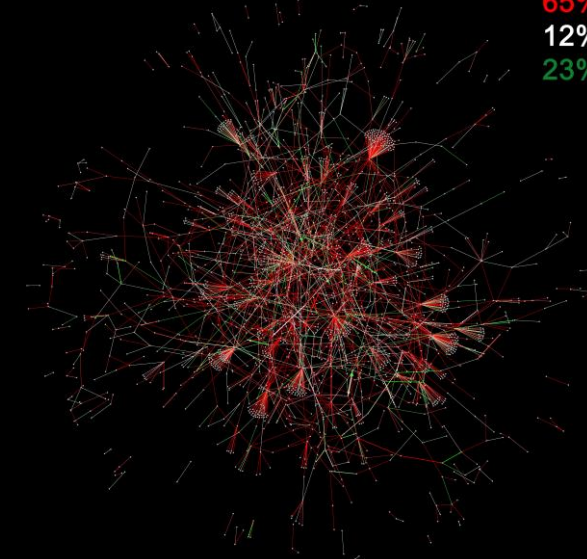
r/television

45%
16%
39%



r/politics

65%
12%
23%



CONCLUSIONI



Differenze strutturali?

Le principali e sostanziali differenze riguardano il **numero di hub**. In particolar modo, si è notato chiaramente come **r/Politics** e **r/Science** risultino avere un numero di hubs particolarmente **elevato** rispetto agli altri subreddit.

Ciò significa che contengono molti più utenti che rispondono ad un alto numero di altri utenti, i quali hanno invece poche interazioni con altri.

Questo può indicare come i subreddit **r/Science** e **r/Politics** siano **particolarmente interconnessi al loro interno**.



2

Information flow? + ●

Si è notato come **r/Politics** e **r/Coronavirus** fossero i network contenenti le **communities più grandi**, le quali presentavano comunque un diametro particolarmente basso. Questo indica ulteriormente che tali communities sono molto interconnesse al loro interno, nonostante l'elevata dimensionalità.

Possiamo quindi affermare che all'interno dei subreddit **r/Politics** e **r/Coronavirus** **l'informazione circoli più facilmente e più efficacemente.**

Interazioni?

r/Politics, **r/Coronavirus** e **r/Science** risultano caratterizzate da una **sentiment** prevalentemente **negativa**

Questo non accade in r/Gaming e r/Television, i quali non risultano avere una sentiment preponderante.



4

Spiegazione?

Pare che quindi, argomenti che possiamo definire più frivoli e leggeri abbiano una sentiment relativamente neutra, mentre argomenti che più suscitano emozioni forti e coinvolgimento psicologico ed emotivo come r/Politics presentino utenti che esprimono **emozioni negative**.

Questo non stupisce particolarmente, e conferma in gran parte l'esperienza quotidiana che tutti notano nelle **discussioni relative a temi caldi come politica e all'ancor più controverso tema del Coronavirus**.

+



EFFETTO REDDIT

+



o



.



GRAZIE

Teresa Cigna - 813925

Chiara Di Domenico - 815463

Federico De Servi - 812166