

Disinformation and network analysis

Marcelo Mendoza

Departamento de Ciencia de la Computación, Pontificia Universidad Católica de Chile

Instituto Milenio Fundamentos de los Datos
Centro Nacional de Inteligencia Artificial

Motivation

- Our approach: To develop AI techniques and models that enhance the understanding of various issues on social networks, grounded in data-based evidence.
- Areas of study: Phenomena on online social networks (such as bots, polarization dynamics, controversies, and fake news).
- Research based on graph representations, modeled either using representation learning techniques or classical models combined with natural language processing.
- Strong interdisciplinary interaction between computer science and other fields such as communication studies and sociology.
- Teamwork, coordinated with interdisciplinary research at IMFD.

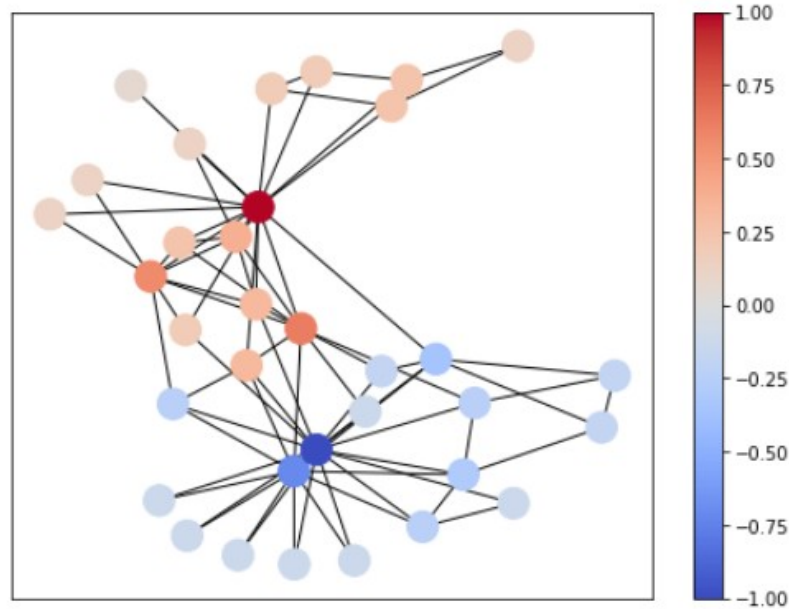
Agenda

- Definitions and basic concepts.
- Bot detection.
- Disinformation dynamics.
- Perspectives in the era of ChatGPT.

- Definitions and basic concepts -

Definitions and basic concepts

Social networks:

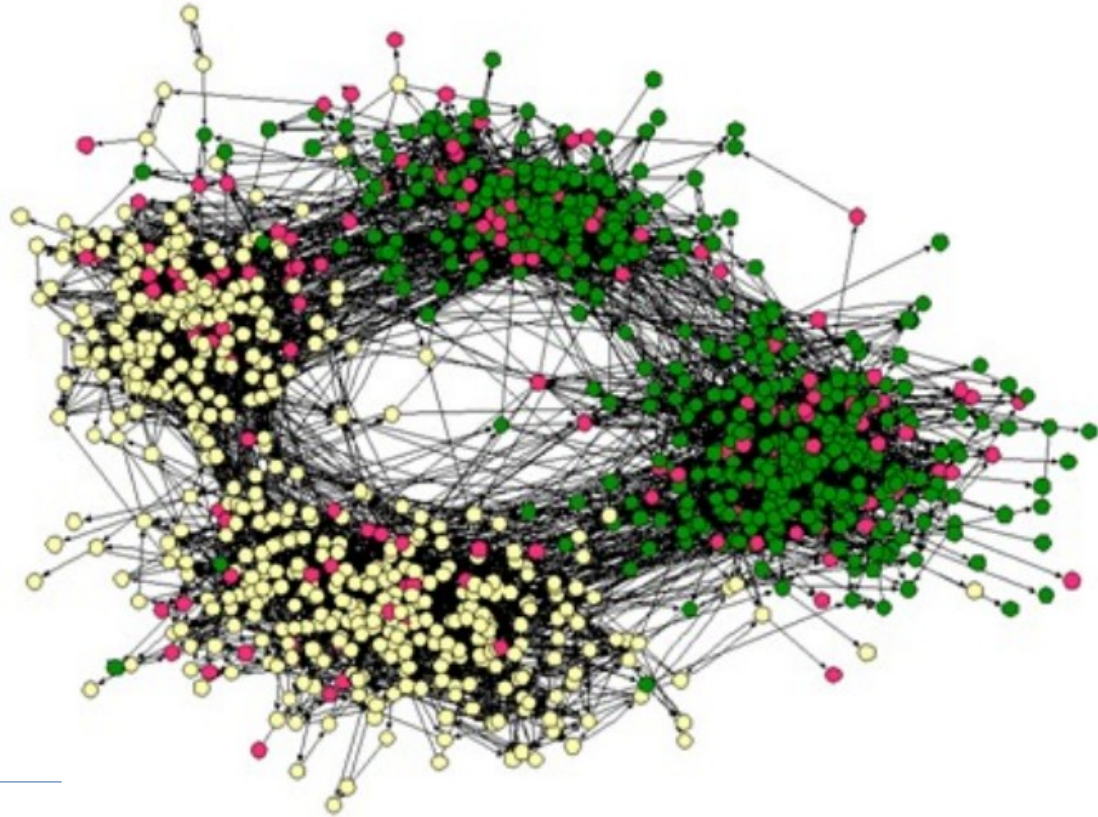


Zachary, W. (1977). An information flow model for conflict and fission in small groups. *Journal of Anthropological Research*, 33, 452–473.

Definitions and basic concepts

Homophily:

Social interactions
Who works with whom?



Moody, J. Race, school integration, and friendship segregation in America. American Journal of Sociology, 2001

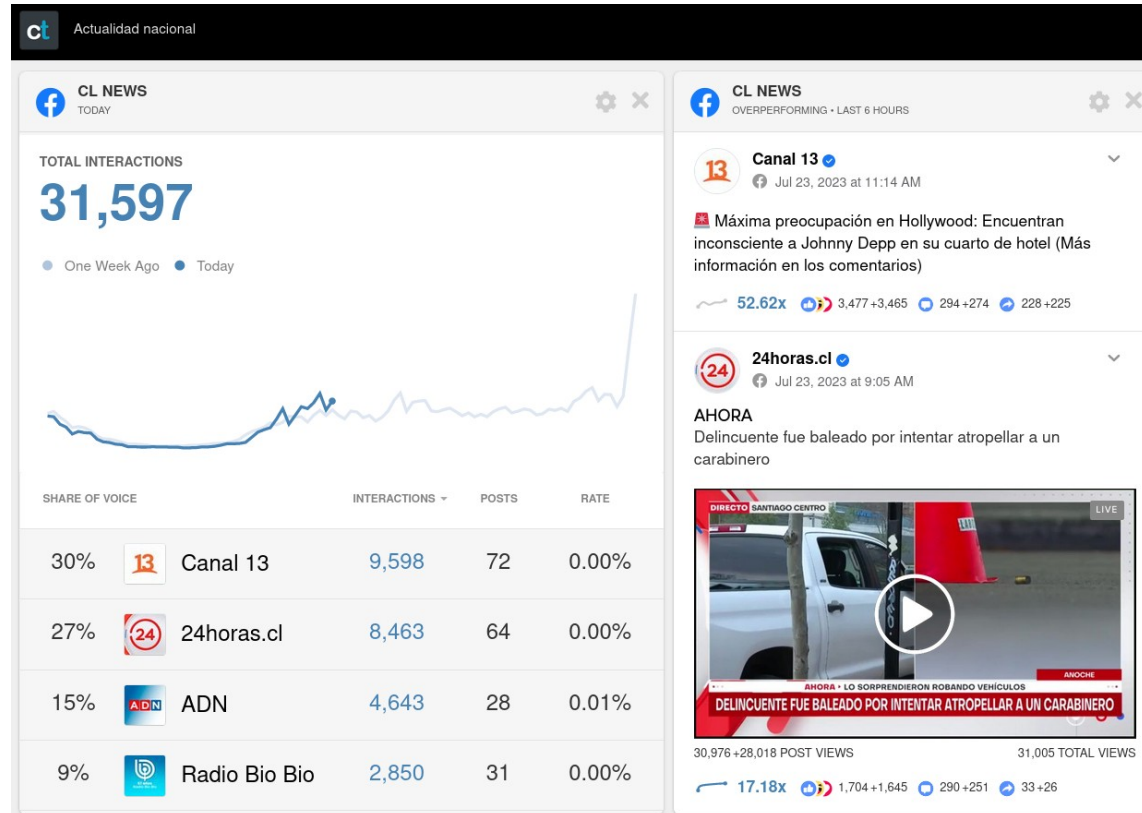
Definitions and basic concepts

Online social networks: Digital social networks where you can write and comment. Some networks establish specific mechanisms for interaction such as likes, retweets, emojis, and so forth.



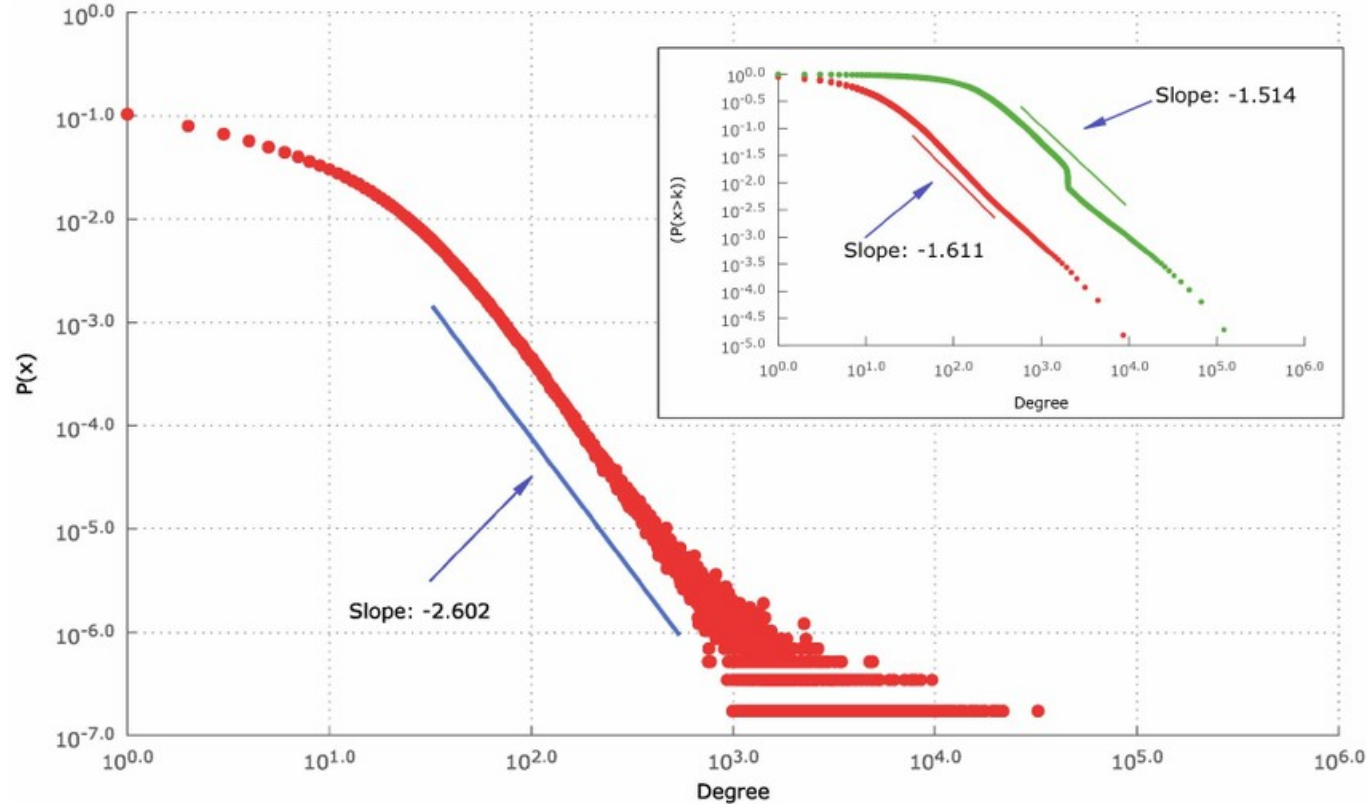
Definitions and basic concepts

Online social networks: They serve as a massive source of information.



Definitions and basic concepts

Structure: node degrees can vary widely.



Definitions and basic concepts

Hierarchy of relevance: Certain accounts are significantly more visible than others (e.g., celebrities, influencers, media outlets, ...).

Presidente Boric defiende reconocimiento a Garzón y delimita gesto a su labor "contra la impunidad"

El Mandatario señaló que la medalla en cuestión no apunta a "otras gestiones que él ya ha realizado como jurista".

18 de Julio de 2023 | 14:07 | Por María Luisa Cisternas, Emol.



Cerrando su paso por Bélgica, el Presidente, **Gabriel Boric**, realizó un punto de prensa en el frontis del edificio Atomium de Bruselas.

Allí, el Mandatario defendió el polémico reconocimiento que entregó en Madrid al ex juez Baltasar Garzón, apuntando que el gesto **"es producto del trabajo que él ha hecho en contra de la impunidad en materia global y en particular en el caso que todos conocemos del juicio a Pinochet, y no por otras gestiones que él ya ha realizado como jurista"**.

NOTICIAS RELACIONADAS



Cumbre Celac-UE: Boric dice que lo de Ucrania "es una guerra de agresión imperial inaceptable" y pide destabar declaración

508

Macaya (UDI) y condecoración a Garzón: "Inconforme a su canceler" y "ofende a una parte de Chile"

540

sectores", agregó.

Bajo esa consideración indicó que "invito a que si alguien tiene un pronunciamiento al respecto, se pronuncie respecto a ese caso que fue el motivo por el cual se entregó a él y a Joan Manuel Serrat una medalla de conmemoración por los 50 años del quiebre de la democracia en Chile".

"Creo que es un momento importante para reflexionar en conjunto sobre la importancia y el valor que le damos a la democracia ante los riesgos que esta enfrenta desde todos los

508

EL COMENTARISTA OPINA

Chile y la meteorología extrema

2

Martin Jacques Coper

CHILE 1973 - 2023



El MAPU, una escisión de la DC que quería ir más allá del "velvetismo"

4



Patricia Arancibia: "Los partidos de derecha no aprendieron las ventajas de conformar un solo partido fuerte"

3



Columna de Carlos Pella: El largo peregrinaje de los partidos

3

RECOMENDADOS EMOL



Repaso semanal junto a Zúñiga, Bertelsen y Zapata

Definitions and basic concepts

Hierarchy of relevance: Certain accounts are significantly more visible than others (e.g., celebrities, influencers, media outlets, ...).

Presidente Boric defiende reconocimiento a Garzón y delimita gesto a su labor "contra la impunidad"

El Mandatario señaló que la medalla en cuestión no apunta a "otras gestiones que él ya ha realizado como jurista".

18 de Julio de 2023 | 14:07 | Por María Luisa Cisternas, Emol.



Cerrando su paso por Bélgica, el Presidente, **Gabriel Boric**, realizó un punto de prensa en el frontis del edificio Atomium de Bruselas.

Allí, el Mandatario defendió el polémico reconocimiento que entregó en Madrid al ex juez Baltasar Garzón, apuntando que el gesto **"es producto del trabajo que él ha hecho en contra de la impunidad en materia global y en particular en el caso que todos conocemos del juicio a Pinochet, y no por otras gestiones que él ya ha realizado como jurista"**.

NOTICIAS RELACIONADAS



Cumbre Celac-UE: Boric dice que lo de Ucrania "es una guerra de agresión imperial inaceptable" y pide destabar declaración

598

Macaya (UDI) y condecoración a Garzón: "Inconforme a su cancelación" y "ofende a una parte de Chile"

540

sectores". agregó.

Bajo esa consideración indicó que "invito a que si alguien tiene un pronunciamiento al respecto, se pronuncie respecto a ese caso que fue el motivo por el cual se entregó a él y a Joan Manuel Serrat una medalla de conmemoración por los 50 años del quiebre de la democracia en Chile".

"Creo que es un momento importante para reflexionar en conjunto sobre la importancia y el valor que le damos a la democracia ante los riesgos que esta enfrenta desde todos los

EL COMENTARISTA OPINA

Chile y la meteorología extrema

2 2 4



Martín Jacques Coper

CHILE 1973 - 2023



El MAPU, una escisión de la DC que quería ir más allá del "velvetismo"

4



Patricia Arancibia: "Los partidos de derecha no aprendieron las ventajas de conformar un solo partido fuerte"

3



Columna de Carlos Pella: El largo peregrinaje de los partidos

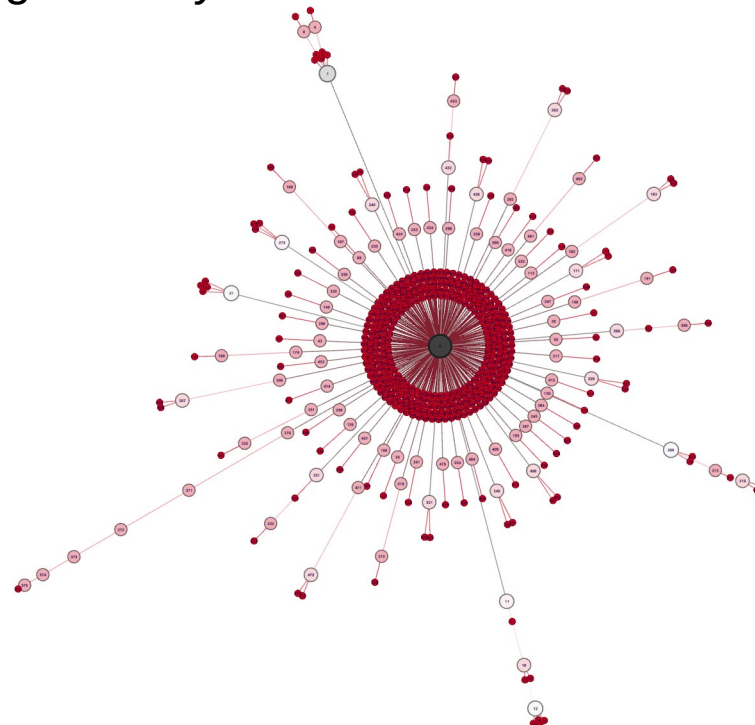
3

RECOMENDADOS EMOL



Reposo semanal junto a Zúñiga, Bertelsen y Zapata

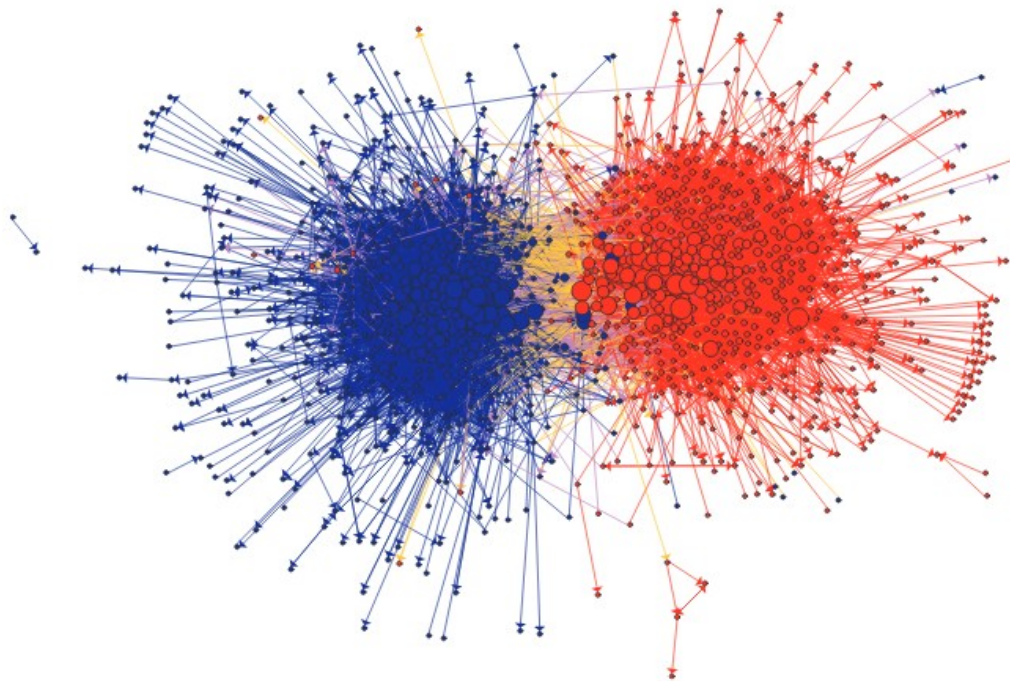
AGENCIA CONSTITUCIONAL



Definitions and basic concepts

Echo chambers: homophily drives towards echo chambers.

Who is speaking —→
with whom

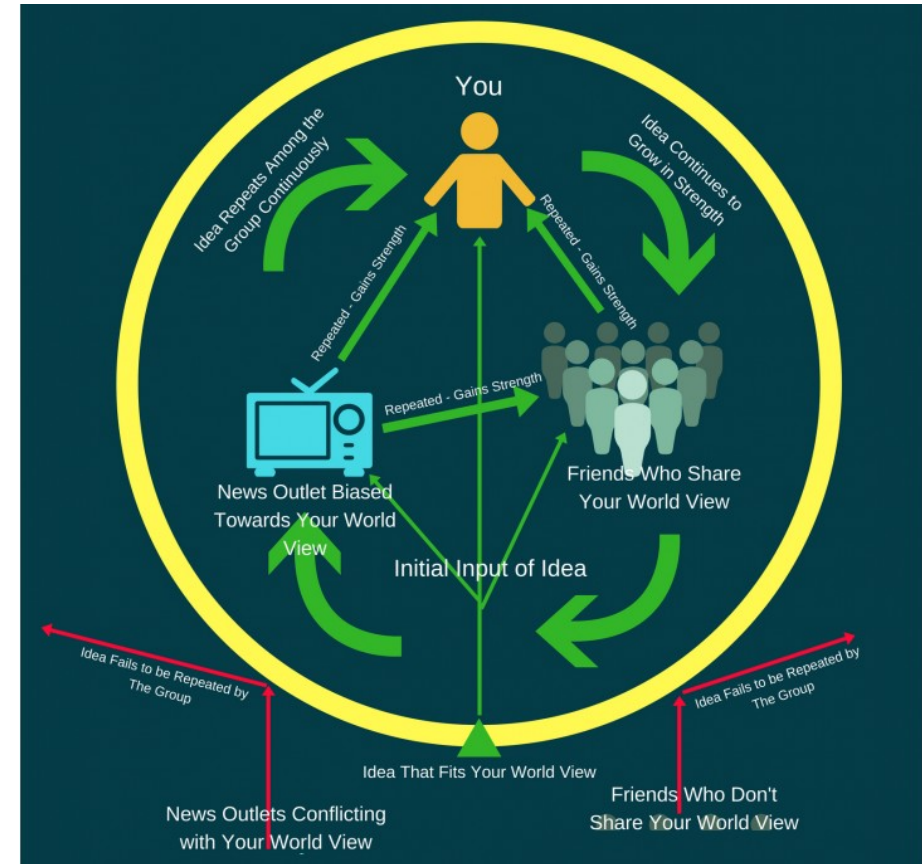


Adamic, L., & Glance, N. (2005). The political blogosphere and the 2004 u.s. election: Divided they blog. In *3rd International Workshop on Link Discovery, LinkKDD 2005 - in conjunction with 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 36–43).

Definitions and basic concepts

Echo chambers: homophily drives towards echo chambers.

The Impact of
an Echo
Chamber



Definitions and basic concepts

Misinformation: "A piece of information whose content contradicts the epistemic consensus achieved through the systematic application of a methodology" [1].

Disinformation: "A specific type of misinformation aimed at manipulating public opinion" [2].

Rumor: "A piece of information whose truthfulness has not been verified at the time of publication" [3].

Fake News: "False news published on a digital and/or traditional information medium" [2].



[1] Swire-Thompson B, Lazer D (2020) Public health and online misinformation: challenges and recommendations. *Annu Rev Public Health* 41(1):433–451.

[2] Zhou X, Zafarani R (2020) A survey of fake news: fundamental theories, detection methods, and opportunities. *ACM Comput Surv (CSUR)* 53(5):1–40.

[3] Zubiaga A, Aker A, Bontcheva K, Liakata M, Procter R (2018) Detection and resolution of rumours in social media: a survey. *ACM Comput Surv* 51(2):32:1–32:36.

Definitions and basic concepts

Fake News: a fabricated content designed to disinform.



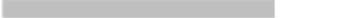












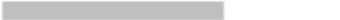
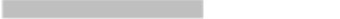
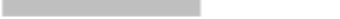



Definitions and basic concepts

A story about influence. Accounts with the highest visibility have a greater impact on social networks. Malicious actors gain traction through fake followers, often automated (bots) or hired (hyper-partisan individuals).

Numerous studies have aimed to curb the influence of malicious actors



rank	feature	proposed in	normalized score	
1	friends/(followers ²) ratio	[8]	1.000	
2	age	[2, 28, 4, 1]	0.919	
3	number of tweets	[1, 4, 8, 12, 14]	0.816	
4	profile has name	[12]	0.782	
5	number of friends	[32, 8, 14, 4]	0.781	
6	has URL in profile	[12]	0.768	
7	following rate	[2]	0.765	
8	default image after 2 months	[14]	0.755	
9	belongs to a list	[12]	0.752	
10	profile has image	[12]	0.751	
11	friends/followers ≥ 50	[14]	0.736	
12	bot in biography	[11]	0.734	
13	duplicate profile pictures	[11]	0.731	
14	$2 \times \text{followers} \geq \text{friends}$	[11]	0.721	
15	friends/followers $\simeq 100$	[11]	0.707	
16	has address	[12]	0.677	
17	no bio, no location, friends ≥ 100	[14]	0.664	
18	has biography	[12]	0.602	
19	number of followers	[12, 4, 3]	0.594	



Cresci, S. Fame for sale: efficient detection of fake Twitter followers, 2015.

Definitions and basic concepts

The Ecosystem of bots:

Astroturfers: These are bots used to support political campaigns. They are employed for electoral propaganda and are managed by organizations or companies that control botnets.

Fake followers: These bots increase the visibility of an account by retweeting its posts.

Cashbots: These are bots that assist fundraising campaigns, mainly dealing with bitcoins.

Spammers: These bots generate automated content, typically delivering repetitive messages.

Self-declared: These are bots that manage chatbot sessions and openly declare themselves as such. Most of these are benign.



Cresci, S.: A decade of social bot detection. [Commun. ACM63\(10\)](#): 72-83 (2020).

Definitions and basic concepts

The Ecosystem of users.

Groups: An institutional mechanism designed to cultivate echo chambers. These groups often incite and escalate radicalization.



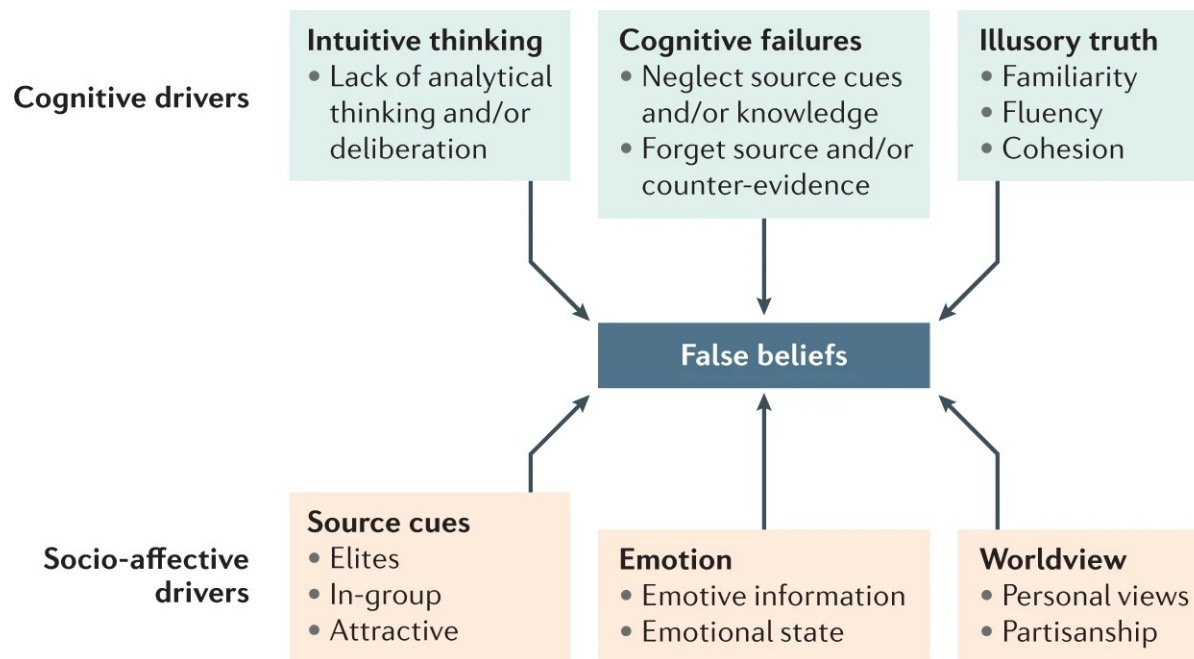
Definitions and basic concepts

Confirmation bias: The veracity of information isn't as relevant as the extent to which it supports your own beliefs and viewpoints.



Definitions and basic concepts

The psychology of fake news: There's an assumption that fake news exacerbates polarization. But it might be the case that polarization exacerbates fake news.



Ecker, U. et al. The psychological drivers of misinformation belief and its resistance to correction, Nature reviews psychology, 2022.

Definitions and basic concepts

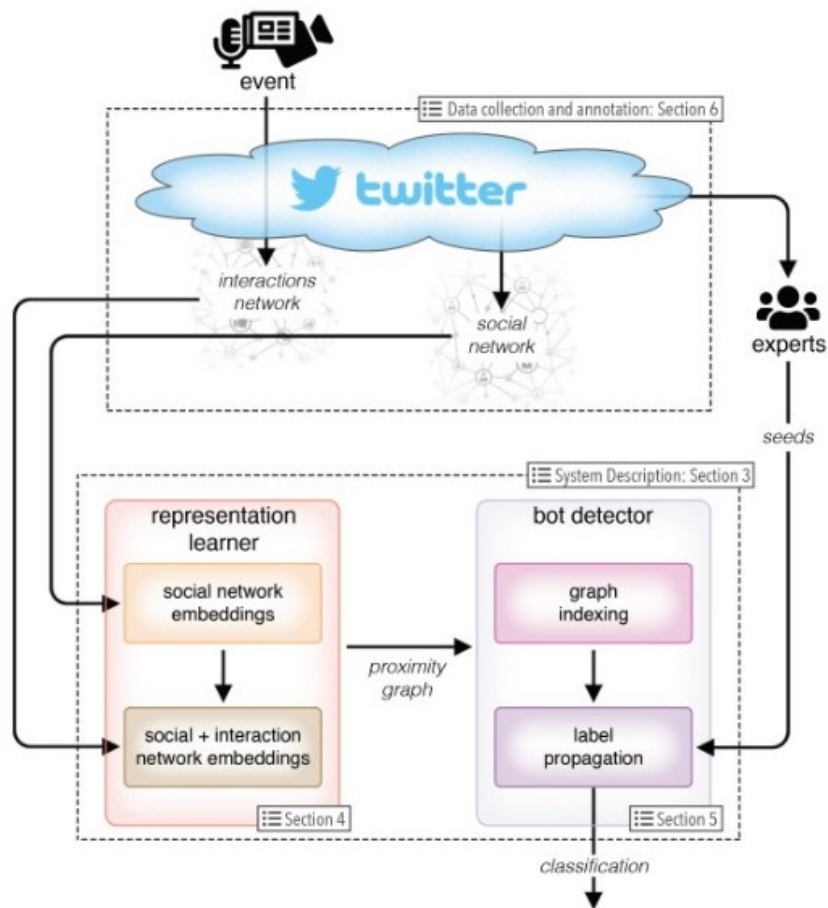
A disinformation ecosystem:



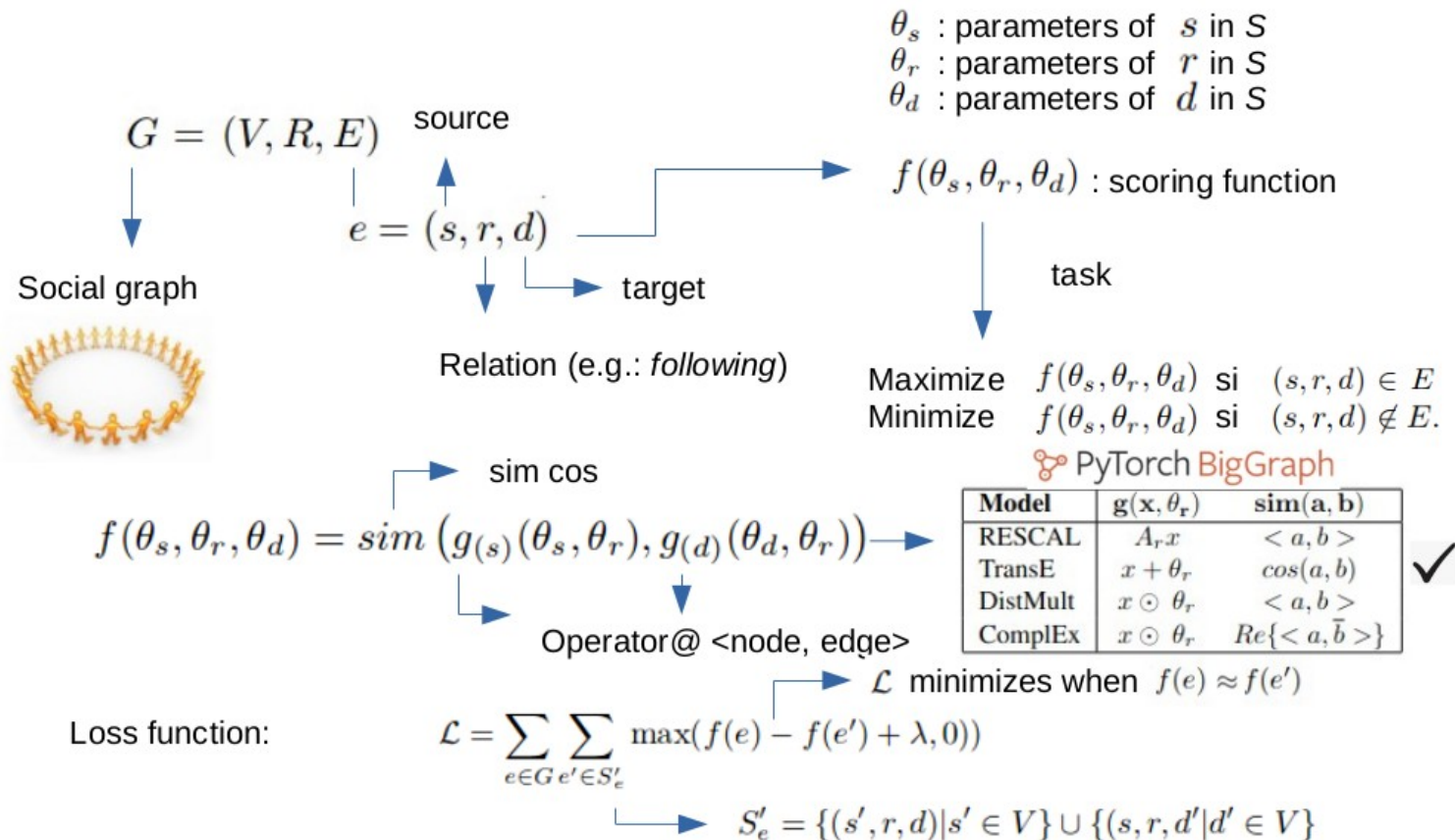
Providel, E., Mendoza, M. (2021). Misleading information in Spanish: a survey, Social Network Analysis and Mining, 11:36

- Bot detection -

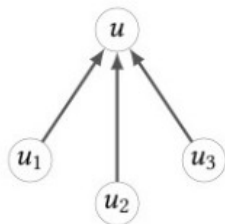
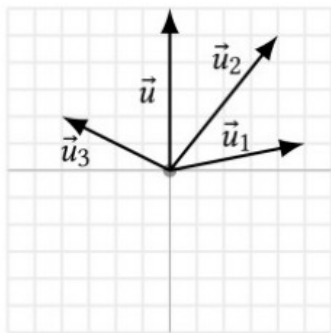
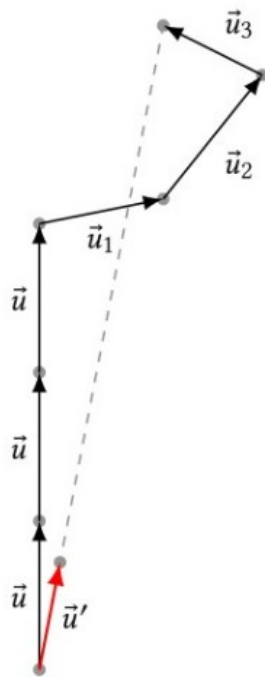
Bots



Bots



Bots

(a) Interaction network of u .(b) Social network embeddings of u , u_1 , u_2 , u_3 .(c) Retro-fitting \vec{u} to u 's neighborhood in the interaction network. \vec{u}' (red-colored) is obtained by combining the social network embedding of u with those of the users with whom it interacted.

$$\vec{u}' = \frac{1}{2n_u} \cdot \left(n_u \cdot \vec{u} + \sum_{i=1}^{n_u} \vec{u}'_i \right).$$

$$\mathcal{F} = \sum_{u \in V} \left[n_u \cdot \|\vec{u}' - \vec{u}\|^2 + \sum_{i=1}^{n_u} \|\vec{u}'_i - \vec{u}\|^2 \right].$$

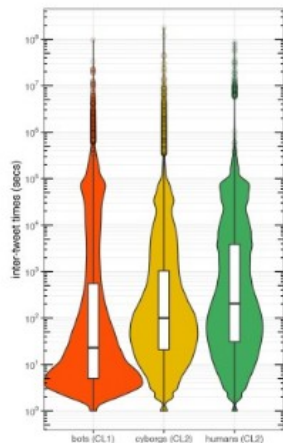
Bots

technique	type	evaluation metrics					
		precision	recall	accuracy	F1	MCC	AUC
comparisons							
Botometer [20, 64]	supervised	0.6951 [•]	0.3098 [•]	0.5830 [•]	0.4286 [•]	0.2051 [•]	0.6889 [•]
Social fingerprinting [12]	supervised	0.6562 [•]	<u>0.8978[•]</u>	0.7114 [•]	0.7582 [•]	0.4536 [•]	0.7501 [•]
HoloScope [39]	unsupervised	0.2857 [•]	0.0049 [•]	0.4908 [•]	0.0096 [•]	−0.0410 [•]	–
RTbust [41]	unsupervised	0.9304[◦]	0.8146 [•]	0.8755 [•]	<u>0.8687[•]</u>	0.7572 [•]	–
our contributions							
social network	semi-supervised	0.8461 [•]	0.8918 [•]	<u>0.8773[•]</u>	0.8684 [•]	<u>0.7658[•]</u>	<u>0.8263[•]</u>
social + interaction networks	semi-supervised	<u>0.9102</u>	0.9594	0.9386	0.9342	0.8778	0.9245

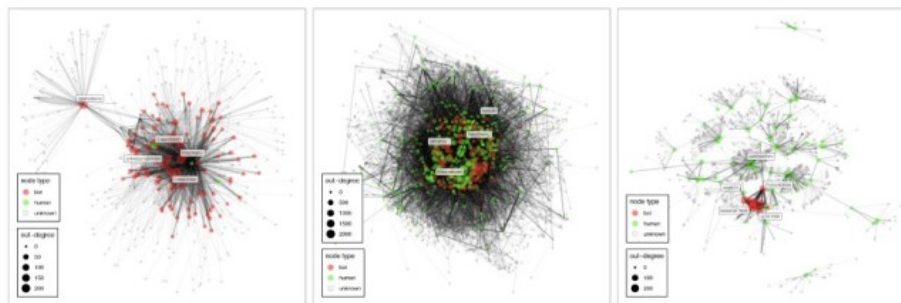
*: $p < 0.01$, °: $p \geq 0.1$ Best results in each evaluation metric are shown in **bold**, second-bests are underlined. Statistical significance results are related to differences between evaluation metrics computed for the best-performing technique (last row) with respect to all the others.

~ 7% de bots

Bots



Distribution of inter-tweet times for accounts labeled as bots in CL1 and CL2 and for accounts labeled as humans in CL2.



(a) Cluster CL1.

(b) Cluster CL2.

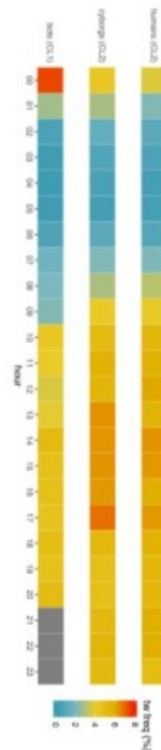
(c) Cluster CL3.

Top-10 Accounts with the Largest In-degree (Most Retweets Received), per Cluster

rank	CL1		CL2		CL3	
	account	label	account	label	account	label
1	@Valerio_Scanu	human	@BTS_ITALIA	bot	@IsabellaF1890	bot
2	@ArmataScanu	bot	@BTSItalia_twt	unknown	@RCCDM90	bot
3	@RaiDue	unknown	@taegijkook	unknown	@RITAPRR9099	unknown
4	@mammaraffy	bot	@GOT7_italia	unknown	@JNYROD	bot
5	@ItaliaCharleroi	unknown	@justignorance_	human	@MARCOFF3	bot
6	@Diletta123	bot	@BLVRRYTAEI	bot	@VFRR1962	bot
7	@OptiMagazine	unknown	@fjfhseason	unknown	@RAVAG68	bot
8	@dada_loi	unknown	@RaiRadio2	human	@JLSESI90	bot
9	@FrancescaDivs	bot	@BTS_ITALIA_ARMY	unknown	@lisaf881	bot
10	@GammaStereoRoma	unknown	@NonnaHaozi	bot	@ansla54	bot

The analysis of most retweeted accounts quickly reveals the goals of the different groups of bots.

Hourly tweet frequencies for accounts labeled as bots in CL1 and CL2 and for accounts labeled as humans in CL2.



Marcelo Mendoza, Maurizio Tesconi, Stefano Cresci.

[Bots in Social and Interaction Networks: Detection and Impact Estimation](#), ACM Transactions on Information Systems (TOIS), 39(1):5:1--5:32, 2020

- Disinformation dynamics -

Fact-checking

“Fact-checking, in the broadest sense, refers to any analysis that publicly challenges a given account or statement”.

Lucas Graves,
Deciding What’s true
The rise of political Fact-Checking,
American Journalism.

Fact-checking



Definition or validation protocols, standards, and methodologies for the accreditation of fact-checking agencies.

- Commitment to transparency.
- Commitment to fairness.
- Commitment to openly disclose sources.
- Commitment to transparently explain the methodology.
- Commitment to clearly indicate the agency's funding sources.
- Commitment to openly present corrections.

Fact-checking

Topic	Fast Check	Decodificador	FactCheckingUC	Total
Social outbreak	102	16	12	130
Covid-19	214	53	86	353
2021 Elections	67	12	0	79
Constitution	57	36	39	132
Other	216	52	38	306
Total	656	169	175	1000

Table 1: Topics per fact checker.

Fact-checking

Veracity	Fast Check	Decodificador	FactCheckingUC	Total
True	284	38	53	375
False	250	86	37	373
Imprecise	122	42	85	249
Unverifiable	0	3	0	3
Total	656	169	175	1000

Table 2: Veracity checks per fact checker.

Fact-checking

Topic	True	False	Imprecise
Social outbreak	20 (8.8%)	15 (0.0%)	7 (10.5%)
Covid-19	43 (17.6%)	37 (45.4%)	14 (21.0%)
2021 Elections	7 (8.8%)	19 (9.0%)	1 (15.7%)
Constitution	20 (17.6%)	23 (18.0%)	29 (15.7%)
Other	39 (41.2%)	19 (27.2%)	14 (36.8%)
Total	129	113	65

Table 4: Veracity per topic in Twitter.

Lingüistic analysis

Crisis literature →

Feature	Social out.	Covid-19	2021 Elect.	Constitution	Other
Length	↓ 2561	4179	4907	4815	4774
Words	↓ 422	690	807	783	788
Emoticons	0	0.02	0	0	0
Entropy	-4.24	-4.25	-4.29	-4.25	-4.28
Sentiment	5.68	5.87	5.94	6.03	6.09
Arousal	5.33	5.25	5.26	5.25	5.29
Dominance	5.05	5.12	5.15	5.16	5.17
Verbs	↓ 37.4	61.5	69.1	68.1	65.5
Dets	↓ 60.1	98.4	106.8	117.3	111.2
Nouns	↓ 89.5	143.7	159.6	161.1	162.7
Propns	↓ 54.2	79.4	121.1	94.1	104.8
Adps	↓ 78.8	121.9	148.2	138.1	146.7
Persons	6.31	7.39	↑ 18.9	10.7	11.1
Locations	↓ 6.28	12.1	11.3	9.3	13.1
Organizations	6.43	7.72	12.3	13.1	11.5
Miscs	10.7	16.7	20.5	20.1	19.6

Table 5: Content profiling (average over source and replies in Twitter). Dets: Determiners, Propns: Proper-nouns, Adps: adpositions, Miscs: Miscellaneous.

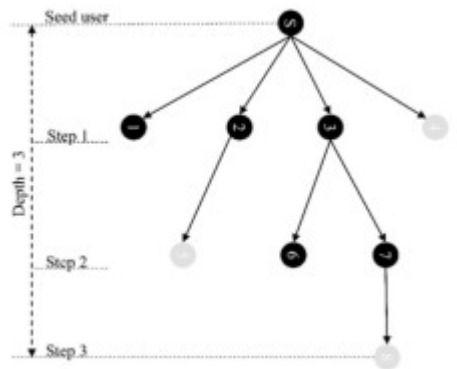
Readability indices

	Metric	False			Imprecise			True		
		mean	std	median	mean	std	median	mean	std	median
1	Number of characters	214	67.48	200	214.27	62.52	214	216.59	76.28	233
2	Number of words	31.8	11.67	34	31.94	11.01	33	31.54	12.9	34
3	Letters per word*	5.87	0.95	5.73	5.89	1.21	5.7	6.04	1.05	5.94
4	Number of sentences	3.42	2.69	3	2.53	1.29	2	2.76	1.4	2
5	Total sentences*	3.43	2.13	3	2.88	1.82	2	2.74	1.64	2
6	TTR	0.9	0.07	0.9	0.89	0.07	0.89	0.9	0.07	0.89
7	LWF	7.07	4.42	6	8.69	4.96	8	8.91	5.53	7.17
8	GFOG	25.3	5.14	23.99	26.25	4.98	25.48	26.75	5.62	27.24
9	DCRS	11.1	1.66	10.96	11.28	1.6	11.14	11.47	1.74	11.49
10	ARI*	15.2	5.65	13.7	16.3	6.5	15	17.28	6.33	16.4
11	FKG*	11.5	3.87	10.6	12.15	4.26	11.9	12.67	4.69	12.3
12	DW	12.9	4.34	13	13.16	4.54	13	13.15	5.03	13
13	CLI*	17.5	5.69	16.81	17.6	7.14	15.73	18.72	6.38	18.09
14	FRE	40.1	19.99	45.72	39.74	23.22	46.13	36.16	23.12	38.82
15	IFSZ	67.3	15.5	67.56	64	18.62	66.57	68.33	13.46	69.86

Differences in cascades

Depth	Mean	Std	Min	Max
True	↓ 6.92	8.54	1	78
False	9.13	11.54	1	88
Imprecise	9.92	10.10	1	60

Table 9: Depth of the propagation trees in Twitter (replies). KS-tests: false and true: $D = 0.168, p \sim 0.08$, false and imprecise: $D = 0.124, p \sim 0.70$, true and imprecise: $D = 0.275, p \sim 0.01$.

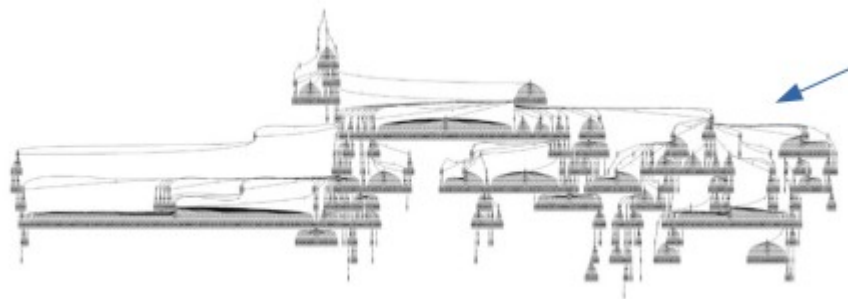


Truthful content often lacks the depth found in other types of content.

Differences in cascades

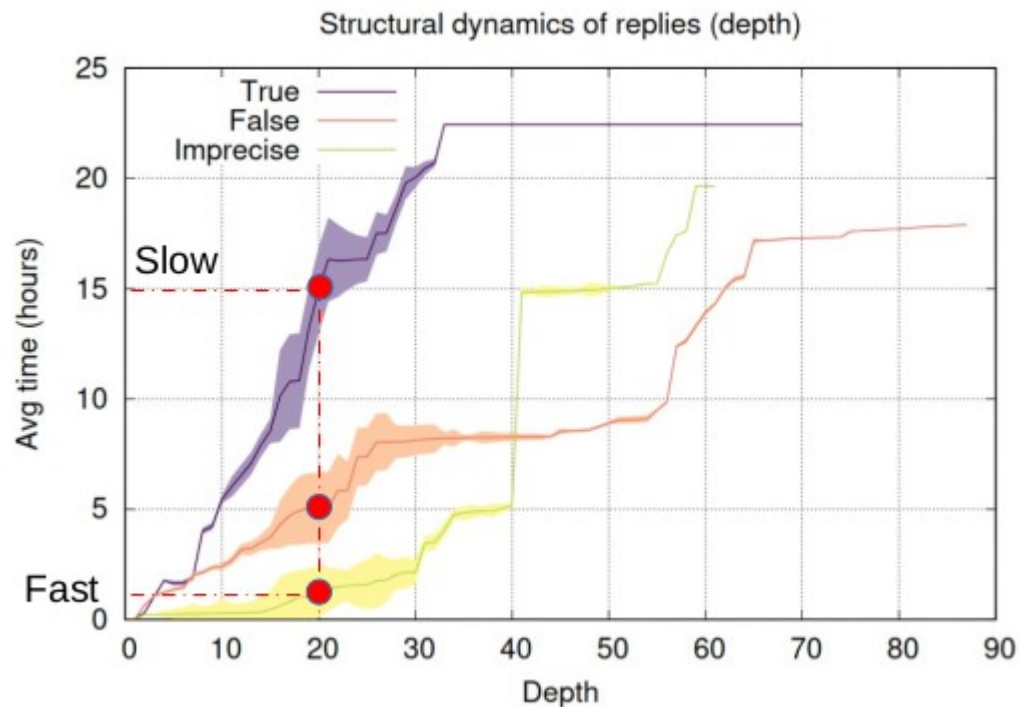
Size	Mean	Std	Min	Max
True	485	1701	3	14237
False	457	669	2	3570
Imprecise	↑ 679	1101	4	5321

Table 10: Size of the propagation trees in Twitter (replies). KS-tests: false and true: $D = 0.206, p \sim 0.01$, false and imprecise: $D = 0.152, p \sim 0.45$, true and imprecise: $D = 0.195, p \sim 0.17$.

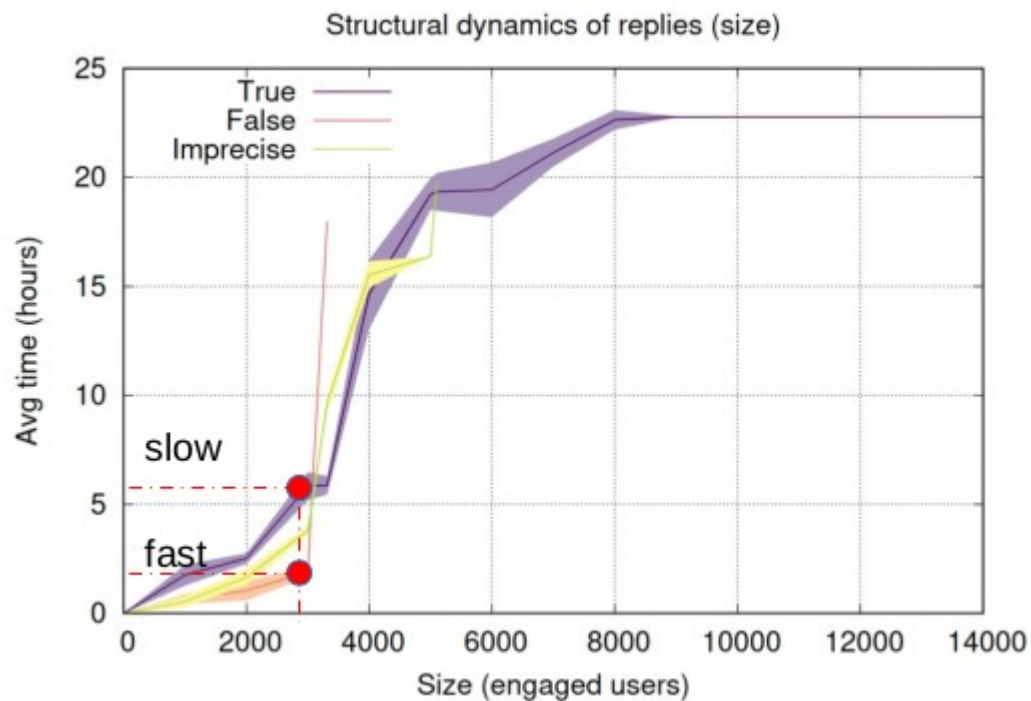


Imprecise content generates more replies than other types of content.

Differences in cascades



Differences in cascades



Differences in cascades

The study can be found at:



desinformacion.cl

It includes a free online course on the subject matter.

You can review the complete version of the study in the following citation:



Marcelo Mendoza, Sebastián Valenzuela, Enrique Núñez-Mussa, Fabián Padilla, Eliana Providel, Sebastián Campos, Renato Bassi, Andrea Riquelme, Valeria Aldana, Claudia López: A Study on Information Disorders on Social Networks during the Chilean Social Unrest and the COVID-19 Pandemic. Applied Sciences, Vol. 13, Issue 9, 2023.

- Perspectives in the era of ChatGPT -

Generative AI and disinformation

Generative AI, in the hands of malicious actors, opens up new possibilities for disinformation.



Generative AI and disinformation

Generative AI, in the hands of malicious actors, opens up new possibilities for disinformation.



SICSS 2023

Deep fakes

43

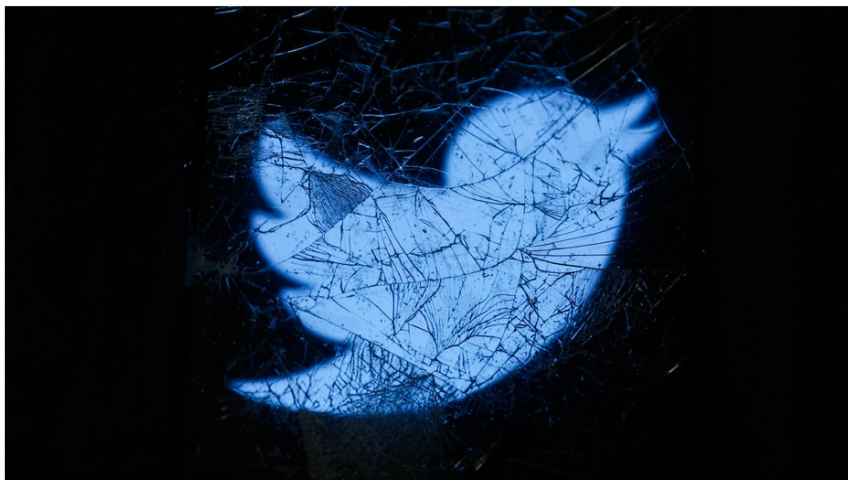
Generative AI and disinformation

Tech Social Media

Twitter's new API pricing is killing many Twitter apps that can't pay \$42,000 per month

"Hobbyists" can pay \$100. Have an app with more than a few users? You'll likely be paying Twitter \$42,000 per month.

By [Matt Binder](#) on March 30, 2023



Twitter appears to have put the nail in the coffin for any indie developer running a Twitter-based app. Credit: Jakub Porzycki/NurPhoto via Getty Images



Generative AI and disinformation

- There are alternatives to Twitter analysis, such as Facebook, Instagram, and TikTok via scraping.
- These are challenging times, filled with numerous hurdles for developing methods of detecting disinformation campaigns. With limited or highly expensive access to data, the gap in data accessibility widens instead of narrowing.
- The challenges are even greater due to the ease with which bad actors can generate content using generative AI. This capability allows them to paraphrase comments, making malicious interactions even more difficult to detect.



Mailto: marcelo.mendoza@uc.cl

Twitter: [@botcheckcl](https://twitter.com/botcheckcl) [@mmendozarocha](https://twitter.com/mmendozarocha)

Web: desinformacion.cl

Disinformation and network analysis

Marcelo Mendoza

Departamento de Ciencia de la Computación, Pontificia Universidad Católica de Chile

Instituto Milenio Fundamentos de los Datos
Centro Nacional de Inteligencia Artificial

