Title: The collaboration research networks of misinformation researchers and their research topics.

Authors: Jaihyun Park, Alton Chua

Affiliation: Wee Kim Wee School of Communication and Information, Nanyang Technological University

Contact: {jay.park, altonchua}@ntu.edu.sg

Abstract: The information age has opened up opportunities for users to make informed decisions based on the abundance of information around them. However, not every piece of content is trustworthy. In fact, the prevalence of online falsehood has increased significantly in recent years. In response to this problem, much research attention has been devoted to better detect and mitigate the effects of misinformation. This study aims to explore (1) collaboration networks among researchers in the misinformation field and (2) the research topics to understand their research agenda. A total of 1,570 papers were collected from arXiv[1], a preprint research paper repository by searching with the keyword "misinformation." Analysis shows that the average clustering coefficient of the co-authorship network is 0.89, indicating that the field of misinformation is highly collaborative (See Fig 1). Additionally, through embedding-based topic modeling (See Table 1), 14 topics related to themes such as politics, vaccines, and large language models were uncovered. The code for this study can be found at https://github.com/park-jay/bibliometric-misinfo

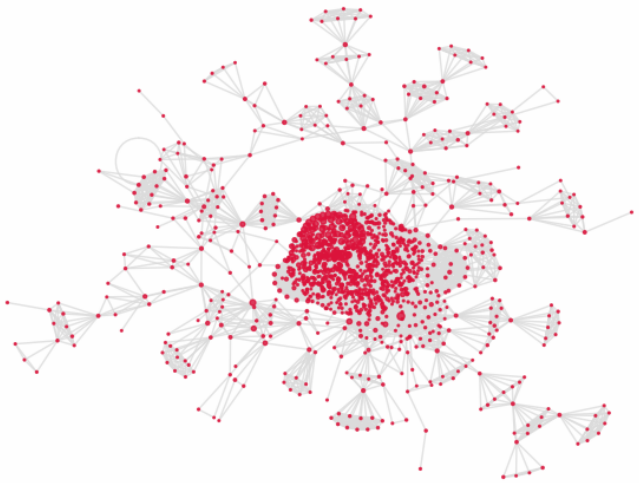

Fig 1. The co-authorship network of researchers in the field of misinformation.

[1] https://arxiv.org/

0	1	2	3	4	5	6
information	datasets	networks	impacts	news	infodemic	llms
misinformation	dataset	algorithms	misinformation	credible	epidemic	datasets
disinformation	credible	distribution	media	sources	misinformation	llm
datasets	data	network	influence	propaganda	disinformation	methodology
data	evidence	methodology	impact	disinformation	information	multilingual
infordemic	sources	datasets	disinformation	credibility	datasets	sources
impacts	methodology	dataset	communities	misinformation	dataset	multimodal
media	reliability	nodes	users	media	media	reliability
dataset	information	computational	propaganda	fake	data	resource
impact	factual	propagation	information	information	viral	models
7	8	9	10	11	12	13
datasets	multilingual	artificial	twitter	deepfakes	vaccines	bots
algorithms	languages	technologies	tweets	deepfake	vaccine	bot
models	datasets	intelligence	election	detectors	vaccination	automated
generative	dataset	ethical	impacts	fake	epidemic	twitter
representations	linguistic	technology	influence	algorithms	viral	datasets
dataset	language	generative	propaganda	detection	propaganda	tweets
deepfake	data	automated	media	detecting	spread	effectiveness
detection	disinformation	risks	political	technologies	population	artificial
methodology	information	vulnerable	campaigns	technology	disinformation	algorithms
technologies	accuracy	implications	tweet	techniques	spreading	dataset

Table 1. The top 10 words for 14 topics