CARMA 2024

6th Int. Conf. on Advanced Research Methods and Analytics
Universitat Politècnica de València, València, 2024
DOI: https://doi.org/10.4995/CARMA2024.2024.17691

Vaccine voices in the digital sphere: a multilayer network analysis of online forum discussion in Taiwan

Jason Dean-Chen Yin
School of Public Health, University of Hong Kong, Hong Kong, China.

How to cite: Yin, J. D-C. 2024. Vaccine voices in the digital sphere: a multilayer network analysis of online forum discussion in Taiwan. In: 6th International Conference on Advanced Research Methods and Analytics (CARMA 2024). Valencia, 26-28 June 2024. https://doi.org/10.4995/CARMA2024.2024.17691

Abstract

New spatiotemporal considerations of the network allow surveillance of vaccine opinion online. The study uses a multilayer network – with each layer representing vaccine opinion discussion – to examine how the structure and timing of engagement in online communities may affect the spread of COVID-19 vaccine opinions. The aim is to improve public health messaging management during health crises and contribute to WHO's growing infodemic research agenda. The study finds that online discussions on COVID-19 vaccines are dominated by a few highly connected nodes within power-law structured communities. Vaccine-hesitant and pro-vaccine discussions are more engaged with, and more frequently posted earlier, but overall, less densely connected than the fewer, but highly clustered anti-vaccine discussions. Temporally, this trend increases over time for anti-vaccine discussions, suggesting insular communication (and potential echo chambering) happens gradually. The findings suggest proactive information management with consistent vaccine advocacy in online communities is crucial in low-activity periods, as dense anti-vaccination networks may pose misinformation risks.

Keywords: *multilayer network; vaccine hesitancy; infodemiology; social network analysis*

1. Introduction

The architecture and dynamics of network structure are influential in shaping the trajectory of information flow. These dynamics and structures have been explored in the digital realm in the age of the internet from various perspectives of network level. At the micro-network level, Bakshy et al. in their study of Facebook found that weak ties play a dominant role in disseminating information, highlighting the importance of users (a "node") in the network. At the meso-network level, the importance of community formation also has implications for information spread (Girvan & Newman, 2002). High degrees of clustering in communities can often form echo-chambers with limited exposure to other information, and facilitate stronger information dissemination in those communities (Moody & White, 2003). At the macro-network level, the structure of node-edge connection describes the mechanisms for how the network grows. One such example is identifying if the network follows a power law distribution (Barabási, 2009). Those following power-law distributions are characterized by a few nodes having a disproportionate number of edges. Analyses of social media platforms have found that many follow power-law distributions (Mislove et al., 2007).

In the public health space, network studies have only recently become popular. For vaccine opinion in particular, studies have used network analysis to identify anti-vaccination themes (Featherstone et al., 2020; Lutkenhaus et al., 2019), study polarization in vaccine ideology (Jiang et al., 2021), or even track opinions on vaccines (Boucher et al., 2021; Gunaratne et al., 2019). While the COVID-19 pandemic has prompted more studies on networks in vaccine hesitancy, it is still a relatively understudied area.

One particularly overlooked area is in the spatiotemporal conceptualization of the network. Networks in vaccine hesitancy are usually represented as a single-layer network graph. However, a multilayer network offers a different lens to study information dissemination across different vaccine opinions. Conceptualizing the network as a multilayer would allow comparisons across structural differences between opinions and their dynamics. While relatively common in biology and physics, their use is limited in studying vaccine opinions (one such exception is their use in disease prediction (Fügenschuh & Fu, 2023; García et al., 2022).

Another overlooked area is incorporating the temporal dimension into network analysis. Networks are not static entities; rather, they change over time. This evolution can influence how information spreads across the network and shape the trajectory of dissemination. One relevant concept here is that of the "first-mover advantage" (Lieberman & Montgomery, 1988). In discussion on vaccines, first narratives may set the tone for ensuing conversations on vaccines. This timing may be important for managing misinformation, since early "psychological inoculation" can strengthen resistance against vaccine misinformation (Compton, 2013). Recent calls from vaccine hesitancy researchers to better incorporate elements of time into vaccine hesitancy research emphasize the salience of the temporal component (Larson et al., 2022).

The current study addresses the following questions: what are the patterns, observations, and implications of online community formation and their timing of online engagement on the spread of COVID-19 vaccine opinions? Specifically, it addresses the following two sub questions: 1) What are implications of differences in the structure and cohesion of different layers on vaccine opinion transmission; and 2) What are the implications of differences in timing on discussion engagement between the different opinions on information transmission? The findings provide insight on the targeted management and monitoring of online public health messaging for current and future public health crises.

2. Methods

2.1. Data

PTT is a terminal-based bulletin board system in Taiwan. It is a free and open forum and non-commercial, where users post and discuss a variety of topics. Often termed Taiwan's "Reddit", it is one of the most active forums in Taiwan. From July 2022 to July 2023, the average users per day was 56,000 (PTT, 2023).

The web-based version of PTT has the following structure. In the forum, there are *boards* and *board masters,* which are the same as subforums and moderators. Posts within each board can be done by creating a new post or a reply post. Within each post, users may leave follow-up comments in text that are also embedded with a sentiment: they can *like, boo,* or have a *neutral reaction*. These are like upvotes or likes, downvotes, or neutral replies, respectively.

All posts from the "Gossiping" board – the most active and popular on PTT – are collated from January 1, 2021 to December 1, 2022: dates that capture the vaccine stockpiling and administration during COVID-19. To find vaccine-related boards, the filter word *"vaccine"* in Chinese is used. For each discussion board, two independent labelers assigned a label of either "pro-vaccine", "vaccine hesitant", or "anti-vaccination" to delineate the sentiment of the post. These labels were classified using a combination of two main criteria. First is WHO's scope of "vaccine hesitant" to encapsulate a broad spectrum of reasons for non-vaccination (Larson et al., 2022; MacDonald et al., 2015). Second is consulting health psychology theories like Health Belief Model (Champion & Skinner, 2008) that include inaction due to self-efficacy reasons. Irrelevant posts mentioning vaccines but not related to vaccine discussions were discarded. The goal was a target inter-rater agreement of 85% and above, with discrepancies resolved by the main author. These three layers constitute the three vaccine opinions of the multilayer network. To construct the network, the embedded sentiment is used to assign the connection to a given layer. For those who "like" or leave neutral comments, they are assigned into the same layer as the sentiment of the post. For those who "boo", they will randomly be assigned into one of the two other layers. Given this randomness, the sample will be bootstrapped, and all measurements (elaborated below) calculated across bootstrap samples.

2.2. Data analysis

The multilayer M consists of three layers $M = \{G^P, G^H, G^A\}$, each being a directed, weighted graph that represents the aggregate links between commenters and authors for vaccine stances *pro-vaccination, vaccine hesitant,* and *anti-vaccination.* Each layer $l \in M$ consists of all interactions of the set of nodes V^l and set of edges E^l, and is represented as $G^l = (V^l, E^l)$, with each node being a user, and each edge being a comment on that user's post. These three layers represent the sentiments towards vaccination.

Mathematically, for the evolving network, assume that we observe the network over a finite time T, with starting point $t_s = 0$ and ending point $t_e = T$. Each layer in M is defined as $G^l_{0,T} = (V, E_{0,T})$ on a time interval $[0, T]$ which consists of a set of nodes or vertices V and a set of temporal edges $E_{0,T}$. The evolving multilayer network is thus $M_{0,T} = \{G^P_{0,T}, G^H_{0,T}, G^A_{0,T}\}$. This multilayer, temporal network is observed at discrete time points $t_1 t_2, \ldots, t_{n-1}, t_n$. At any time point t_n, an instantiation of the multilayer, M_n, is observed, whereby each G^l_n contains the set of temporal edges E^l_n such that $(u, v)^l{}_{t_n} \in E^l_{0,T}$ with edges between nodes u, v contained within the period $t_n = [\, t_{n_s}, t_{n_e}]$ such that $t_{n_s} \leq T$ and $t_{n_e} \geq t_{n_s} \geq 0$ (*i.e.* the instantiation time is between the start time t_s and end time t_e, and the end time t_e is later than the start t_s). The graphs, being directed, are also non-mirrored such that $(u, v) \neq (v, u)$). Each of the following measures below are written for one snapshot t_n, but are calculated temporally.

Degree (average)

The degree of a node is the number of edges connected to it. The average degree of the layer is the average of degrees of all nodes in the network layer. For a directed graph $G(V, E)$, the average degrees $\bar{d}_{tot}$ of all nodes is:

$$\bar{d}_{tot} = \frac{\bar{d}_{in} + \bar{d}_{out}}{2} \tag{1}$$

where $\bar{d}_{in}$ and $\bar{d}_{out}$ represent the average indegrees and outdegrees of a given network.

Density

The density of a network layer measures the connectedness of the graph. For any network, the density $d(G)$ of a graph $G(V, E)$ is the number of edges divided by the theoretical possible number of edges. In a directed graph, the density is calculated as $\frac{E}{2V(V-1)}$, where E is the number of edges and V is the number of vertices (nodes). The density of a network ranges from 0 to 1, with 0 representing no edges, and 1 representing all possible edges present.

Power law distribution

A network is considered a power law network when the probability distribution of degree d, $p(d)$, follows a power law $p(d) \propto d^{\delta}$, where δ is the exponential parameter of the power law distribution, usually falling between $1 \leq \delta \leq 3$. Delta values around 2 indicate a power law network. This will be tested with the "powerlaw" library in Python.

Modularity

Modularity is defined as the difference between the actual number of edges within communities from the expected number of edges given the *degree distribution*, or, the number of edges connected to a node. Mathematically, modularity Q is defined as:

$$Q = \frac{1}{2m} \sum_i \sum_j \left[\frac{A_{ij} - k_i^{out} * k_j^{in}}{2m} \right] * \Delta(c_i, c_j) \tag{2}$$

where A_{ij} is the weight of the edge from node i to node k, k_i^{out} is the sum of weights of the edges leaving node i , k_j^{in} is the sum of the weights entering node j, m is the total weight of all edges in the graph, and $\Delta(c_i, c_j)$ is the Kronecker delta function, equalling 1 if nodes i and j belong to the same community, and 0 if not (boolean operator for contributing to modularity score).

Active communities

An "active community" is one that has a significant number of nodes compared to other communities in the same layer. If n^l and c^l denote the number of nodes and communities in layer l, and n_c^l represents the number of nodes in community c of layer l, then the existence of an active community (AC) in layer l is given as Boolean classifier:

$$AC^l(c) = \begin{cases} true, & if\ n_c^l \geq (\frac{n^l}{c^l}) \\ false, & else \end{cases} \tag{3}$$

Percent constitution of network layer by active community

After identifying the active communities, the number of nodes in the active communities is divided by all the nodes in the layer to calculate the percent constitution of network layer (PC_{AC}^l) by active communities with the following formula:

$$PC_{AC}^l = \frac{\sum_{c \in AC^l} n_c^l}{n^l} \tag{4}$$

Survival analysis

To analyse engagement, I fit the time-to-posting of each layer to a Cox Proportional Hazards Model, with predictor as layer. Hazard ratios of each possible pairwise comparison indicate the relative rate at which the groups experience the two events above, assuming proportional hazards.

3. Results

I find that all three layers have delta values – the power law parameter –significantly around two, with p-values above 0.05 ($G^P: \delta = 2.2, p = 0.067 \quad G^H: \delta = 2.3, p = 0.068 \quad G^A: \delta = 2.3, p = 0.09$). This confirms these layers are power-law networks, and all three layers generally revolve around a few highly connected nodes.

Discussions on the vaccine hesitant are the most active, followed by the pro-vaccine layer (Table 1). The layer with the least users and interactions is anti-vaccination. This is also reflected in average degrees. These findings suggest more hesitant and positive discussions are more engaged with than anti-vaccination ones. However, density is in the opposite direction. The anti-vaccination layer has the highest density suggesting that, despite its small size, the discussions are more tightly connected as a network. Since degree and density are network size-dependent, the results in Table 1 should be interpreted cautiously. When comparing just the pro-vaccination and hesitant layers, which are more similar in size, the hesitant layer has more connections per node (higher degree) but is overall more dispersed (lower density), suggesting a more scattered network of information hubs.

Table 1. Properties of the multilayer network by layer (bootstrap n=1,000)

Layer	Number of posts	Nodes (average)	Edges (average)	Density (10^{-4})	Average degree	Modularity	Active communities	Active communities' constitution of network
G^P	1,283	11,087	23,504	1.91 (1.89 – 1.94)	4.24 (4.21 – 4.27)	0.574 (0.570 – 0.578)	46.0 (39.0 – 53.0)	79.6 (72.7 – 86.7)
G^H	1,322	14,037	33,520	1.70 (1.68 – 1.72)	4.77 (4.75 – 4.80)	0.543 (0.538 – 0.548)	34.0 (28.0 – 41.0)	75.9 (69.7 – 81.3)
G^A	387	5,477	8,696	2.90 (2.85 – 2.95)	3.18 (3.15 – 3.20)	0.699 (0.691 – 0.704)	21.0 (17.0 – 25.0)	57.4 (51.2 – 63.4)

On modularity, the pro-vaccination and vaccination communities are similar, while the anti-vaccination network is higher, indicating stronger clustering. Interpreted in tandem to density, the high density and modularity of the anti-vaccination layer suggests many localized and highly intra-connected communities that communicate amongst themselves. The pro and hesitant communities, alternatively, connect more sparsely in the communities but more strongly in the layer. The difference in the number of communities and their percent constitution of the overall network further illustrates this. The percent constitution of the entire layer is much lower for the anti-vaccination layer, suggesting that the active communities are dominant and main contributors to high density, while the other half of the layer is very inactive and separate from these tight communities.

Temporally, Figure 1 illustrates that anti-vaccination community cohesion grows over time, indicating insular information transmission hubs. On the other hand, the pro-vaccination and hesitant layers have steadier community cohesion over time. The gradual tapering in all three layers in modularity and percentage suggests that new participants are isolated from these conversations, with this more noticeable in the anti-vaccination network.

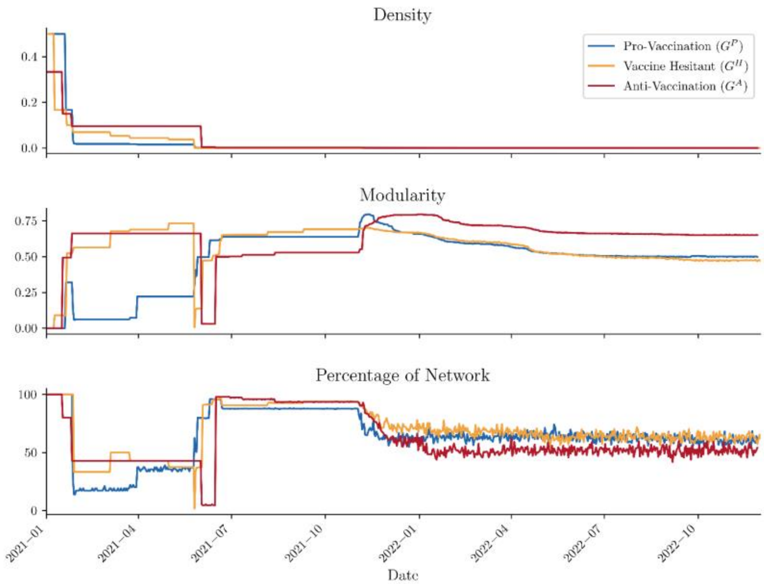

Figure 1. Temporal changes in density, modularity, and percentage of network occupied by active communities

Differences between topic starting were tested for each layer using survival curves and a Cox proportional hazards model to test earliness to conversation. Overall, the hesitant layer and anti-vaccination layers posted less than the pro-vaccination layer (0.87, p=0.005; 0.75, p=0.01, respectively, Table 2). Hazards were proportional in each layer. However, the discussion around vaccines only escalated much later around September 2021 after vaccines were procured, and the vaccination campaign was in effect on the island (Figure 2). In the period prior, the forum paid little attention to COVID-19 vaccinations.

Table 2. Hazard ratios for time-to-thread posting for each layer

Measure	Hazard ratio	p-value	Coefficient	Standard error	Z
Pro-vaccination (baseline)	-	-	-	-	-
Vaccine hesitant	0.87 (0.78 – 0.96)	0.005	-0.144	0.05	-2.79
Anti vaccination	0.75 (0.60 – 0.94)	0.014	-0.27	0.11	-2.41

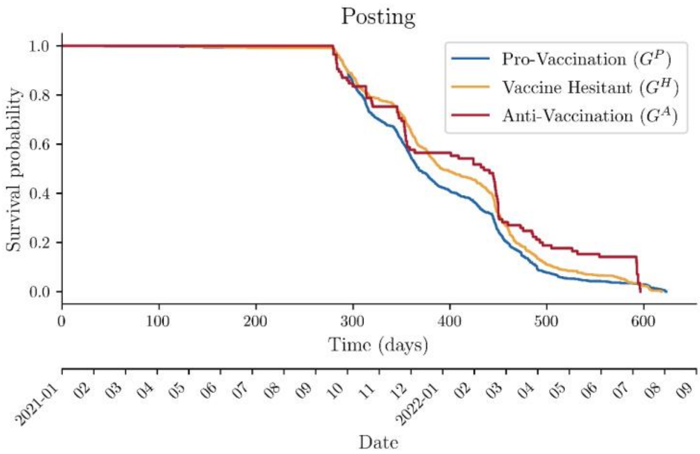

Figure 2. Thread-posting survival curves for the three layers

4. Discussion

In this paper, I have proposed and explored using a multilayer network to describe the formation of discussions online. I found that the anti-vaccination discussions are fewer, but more densely interconnected, indicating a higher risk of anti-vaccination misinformation in this group. This trend was only later in the time series, and not true initially. In addition, vaccine-hesitancy posts are made less often, but more sporadically. These findings underscore the importance of consistent vaccination advocacy in 'downtime' periods. This advocacy can target tightly knit anti-vaccination discussion communities that potentially form echo chambers. Findings in this study should be compared – both within and across platforms domestically, and internationally – to better triangulate findings.

References

Barabási, A. L. (2009). Scale-free networks: A decade and beyond. Science, 325(5939), 412–413.

Boucher, J.-C., Cornelson, K., Benham, J. L., Fullerton, M. M., Tang, T., Constantinescu, C., Mourali, M., Oxoby, R. J., Marshall, D. A., Hemmati, H., Badami, A., Hu, J., & Lang, R. (2021). Analyzing Social Media to Explore the Attitudes and Behaviors Following the Announcement of Successful COVID-19 Vaccine Trials: Infodemiology Study. JMIR Infodemiology, 1(1), e28800. https://pubmed.ncbi.nlm.nih.gov/34447924/

Champion, V. L., & Skinner, C. S. (2008). The Health Belief Model. In Health Behavior and Health Education: Theory, Research, and Practice (4th ed.). Jossey-Bass.

Compton, J. (2013). Inoculation theory. In The SAGE handbook of persuasion: Developments in theory and practice (pp. 220–236). Sage Publications, Inc. https://psycnet.apa.org/record/2013-39243-014

Featherstone, J. D., Ruiz, J. B., Barnett, G. A., & Millam, B. J. (2020). Exploring childhood vaccination themes and public opinions on Twitter: A semantic network analysis. Telematics and Informatics, 54, 101474.

Fügenschuh, M., & Fu, F. (2023). Overcoming Vaccine Hesitancy by Multiplex Social Network Targeting. Studies in Computational Intelligence, 1077 SCI, 576–587. https://link.springer.com/chapter/10.1007/978-3-031-21127-0_47

García, Y. E., Mery, G., Vásquez, P., Calvo, J. G., Barboza, L. A., Rivas, T., & Sanchez, F. (2022). Projecting the impact of Covid-19 variants and vaccination strategies in disease transmission using a multilayer network model in Costa Rica. Scientific Reports 2022 12:1, 12(1), 1–9. https://www.nature.com/articles/s41598-022-06236-1

Girvan, M., & Newman, M. E. J. (2002). Community structure in social and biological networks. Proceedings of the National Academy of Sciences of the United States of America, 99(12), 7821–7826. https://www.pnas.org/doi/abs/10.1073/pnas.122653799

Gunaratne, K., Coomes, E. A., & Haghbayan, H. (2019). Temporal trends in anti-vaccine discourse on Twitter. Vaccine, 37(35), 4867–4871. https://pubmed.ncbi.nlm.nih.gov/31300292/

Jiang, X., Su, M. H., Hwang, J., Lian, R., Brauer, M., Kim, S., & Shah, D. (2021). Polarization Over Vaccination: Ideological Differences in Twitter Expression About COVID-19 Vaccine Favorability and Specific Hesitancy Concerns: Social Media and Society, 7(3). https://journals.sagepub.com/doi/full/10.1177/20563051211048413

Larson, H. J., Gakidou, E., & Murray, C. J. L. (2022). The Vaccine-Hesitant Moment. New England Journal of Medicine, 387(1), 58–65. https://www.nejm.org/doi/full/10.1056/nejmra2106441

Lieberman, M. B., & Montgomery, D. B. (1988). First-mover advantages. Strategic Management Journal, 9(S1), 41–58. https://onlinelibrary.wiley.com/doi/full/10.1002/smj.4250090706

Lutkenhaus, R. O., Jansz, J., & Bouman, M. P. A. (2019). Mapping the Dutch vaccination debate on Twitter: Identifying communities, narratives, and interactions. Vaccine: X, 1, 100019.

MacDonald, N. E., Eskola, J., Liang, X., Chaudhuri, M., Dube, E., Gellin, B., Goldstein, S., Larson, H., Manzo, M. L., Reingold, A., Tshering, K., Zhou, Y., Duclos, P., Guirguis, S., Hickler, B., & Schuster, M. (2015). Vaccine hesitancy: Definition, scope and determinants. Vaccine, 33(34), 4161–4164. https://pubmed.ncbi.nlm.nih.gov/25896383/

Mislove, A., Marcon, M., Gummadi, K. P., Druschel, P., & Bhattacharjee, B. (2007). Measurement and analysis of online social networks. Proceedings of the ACM SIGCOMM Internet Measurement Conference, IMC, 29–42. https://dl.acm.org/doi/10.1145/1298306.1298311

Moody, J., & White, D. R. (2003). Structural cohesion and embeddedness: A hierarchical concept of social groups. American Sociological Review, 68(1), 103–127.

PTT. (2023). PTT Online User Statistics. https://www.ptt.cc/bbs/Record/search?q=上站人次統計