



OPEN

Assortative mixing of opinions about COVID-19 vaccination in personal networks

Marian-Gabriel Hâncean^{1,2✉}, Jürgen Lerner^{3,4}, Matjaž Perc^{5,6,7,8,9}, José Luis Molina¹⁰ & Marius Geantă¹¹

Many countries worldwide had difficulties reaching a sufficiently high vaccination uptake during the COVID-19 pandemic. Given this context, we collected data from a panel of 30,000 individuals, which were representative of the population of Romania (a country in Eastern Europe with a low 42.6% vaccination rate) to determine whether people are more likely to be connected to peers displaying similar opinions about COVID-19 vaccination. We extracted 443 personal networks, amounting to 4430 alters. We estimated multilevel logistic regression models with random-ego-level intercepts to predict individual opinions about COVID-19 vaccination. Our evidence indicates positive opinions about the COVID-19 vaccination cluster. Namely, the likelihood of having a positive opinion about COVID-19 vaccination increases when peers have, on average, a more positive attitude than the rest of the nodes in the network (OR 1.31, $p < 0.001$). We also found that individuals with higher education and age are more likely to hold a positive opinion about COVID-19 vaccination. With the given empirical data, our study cannot reveal whether this assortative mixing of opinions is due to social influence or social selection. However, it may nevertheless have implications for public health interventions, especially in countries that strive to reach higher uptake rates. Understanding opinions about vaccination can act as an early warning system for potential outbreaks, inform predictions about vaccination uptake, or help supply chain management for vaccine distribution.

Vaccination has been the paramount pharmaceutical intervention to halt the coronavirus disease pandemic (COVID-19)¹. However, despite the actions taken in Europe by the European Commission to ensure timely access to vaccines for the Member States, various mass vaccination campaigns have not realized their potential, especially in Eastern European countries. Therefore, understanding the social mechanisms underpinning vaccination willingness is pivotal to fighting against both the still ongoing COVID-19 pandemic and future other pandemics.

Vaccination acceptance and its associated determinants are multiplex². In the case of COVID-19, most of the literature has focused on individual-level predictors such as demographic characteristics (gender/sex, age, ethnicity/race, education, income, occupation), personal health history (medical conditions, personal experience with COVID-19), and beliefs (perceptions about the harms or efficiency of the vaccine). Significantly fewer studies have addressed supra-individual level factors such as healthcare and societal determinants³.

Scientists agree that vaccine acceptance is a complex decision-making process influenced by "experience, risk perception, culture, confidence in authorities and medicine"⁴. However, as we move from one study to another, many of the reported empirical findings are mixed or unclear (especially concerning race, age group, gender, employment status, and education)⁴. This inconsistency in the results may suggest that researchers have

¹Department of Sociology, University of Bucharest, Panduri, 90-92, 050663 Bucharest, Romania. ²The Research Institute of the University of Bucharest, University of Bucharest, Panduri, 90-92, 050663 Bucharest, Romania. ³Department of Computer and Information Science, University of Konstanz, 78457 Konstanz, Germany. ⁴Human Technology Center, RWTH Aachen University, 52062 Aachen, Germany. ⁵Faculty of Natural Sciences and Mathematics, University of Maribor, Koroška Cesta 160, 2000 Maribor, Slovenia. ⁶Department of Medical Research, China Medical University Hospital, China Medical University, Taichung 404332, Taiwan. ⁷Community Healthcare Center Dr. Adolf Drolc Maribor, Vošnjakova Ulica 2, 2000 Maribor, Slovenia. ⁸Complexity Science Hub Vienna, Josefstädterstraße 39, 1080 Vienna, Austria. ⁹Department of Physics, Kyung Hee University, 26 Kyungheedaero, Dongdaemun-Gu, Seoul, Republic of Korea. ¹⁰GRAFO - Department of Social and Cultural Anthropology, Universitat Autònoma de Barcelona, 08193 Bellaterra (Cerdanyola del Vallès), Barcelona, Spain. ¹¹Center for Innovation in Medicine, Th. Pallady 42J, 032266 Bucharest, Romania. ✉email: gabriel.hancean@sas.unibuc.ro

overlooked some predictors. Notably, the potential role of human networks in forming, reinforcing, or spreading opinions about COVID-19 vaccination has not been fully considered.

Disregarding network data is striking. Evidence shows that social networks affect health outcomes^{5–7} (e.g., the spread of obesity, COVID-19 infections, smoking, health screening, HPV vaccination uptake, happiness, depression, sleep, or loneliness). People do not live in isolation, and their behavior is not detached from the behavior of others. Health is a social network outcome. Individuals are interconnected, so their health is interconnected (health preferences, decisions, or habits). Research has already illustrated *assortativity* (connected individuals tend to share traits and behavior⁸) as an essential property of human networks⁹. For example, previous work has revealed the association between node characteristics (behavior) and network structure in the case of influenza vaccination¹⁰, local and global COVID-19 spreading^{11,12}, sexually transmitted infections¹³, alcohol consumption¹⁴, and, generally, in epidemiologic studies⁸. Additionally, theoretical demonstrations¹⁵ claim that opinions about vaccination are generally not randomly distributed in human networks but clustered.

Surprisingly, to our knowledge, analyzing opinions about COVID-19 vaccination using a social network perspective has proved nonexistent in the literature. Therefore, our paper examines the role of human networks in understanding and predicting opinions about COVID-19 vaccination. Specifically, our main research objective in this study is to assess whether assortativity positively contributes to predicting COVID-19 opinions. In this direction, we regard *personal networks* (individuals, their direct social contacts, and the interconnections among them) as the immediate social contexts embedding the individuals¹⁶.

We analyze the personal networks of 443 individuals (*egos*), their social contacts (*alters*), and the tie configurations embedding the alters and surrounding the egos (we collected the information between March 16 and March 30, 2022, in Romania). These networks are of an equal number of alters (ten alters per network), amounting to 4430 alters. We examine if the opinions of the social contacts are clustered (by assortativity) or randomly distributed. Specifically, our study aims to assess the assortative mixing of opinions among the alters of the same ego. Thus, alters' opinion about the COVID-19 vaccination is the dependent variable. We use multi-level regression models to cope with the fact that alters are clustered by ego (some egos may be in communities with more positive/negative opinions; different egos may have a different opinion of what constitutes a positive/negative opinion, etc.).

The dataset comprises socio-demographic variables for egos and alters (individual attributes: sex, education, income, and age) and network variables (characteristics of the ego-alter ties and the alter-alter tie configurations). This unique collection of personal networks is also significant because it comes from a population (Romania) that has exhibited the second-lowest COVID-19 vaccination among European Union (EU) countries. As of March 13, 2023, according to the European Centre for Disease Prevention and Control, Romania (an Eastern European EU country) reported 42.6% of the population with at least one dose uptake and only 9.2% with the first booster uptake. Our findings suggest that opinions about COVID-19 vaccination are clustered in personal networks. Specifically, we illustrate that accounting for information about social contacts (ego's alters) brings new insights and allows for predicting COVID-19 vaccination opinions.

Assortativity by COVID-19 vaccination opinions

We claim that current mainstream research can benefit from linking people to their surrounding social context¹⁷. In this fashion, we aim to detect assortativity in the social organization of opinions about COVID-19 vaccination. On the one hand, existing evidence¹⁸ advocates the role of *social contagion* (*influence*) in adopting innovation. Adopting an idea, a vaccine, or a technology is dependent on the proportion of surrounding people who have already adopted it¹⁹. Further, actors mutually influence and inform each other, increasing homogeneity within structural subgroups²⁰. Human networks can be addressed as conduits for the circulation of intangible or tangible resources: from COVID-19 infections^{11,21}, opinions²² and ideas^{23,24} to goods and other objects^{25,26}. Thus, the pattern of networks is essential for understanding the flow of information²⁷. Acquiring a trait may result from interacting with only one source (e.g., SARS-COV-2 spreading) or multiple active sources (e.g., opinion formation)¹⁹. Scholars argue that close friends and relatives are critical for complex contagion, whereas acquaintances are instrumental for the circulation of information over long social distances²⁸.

On the other hand, human networks are not fixed but the object of innovations from the part of their embedded members. Social interactions are governed by *social selection*. Individuals tend to prefer to interact with others similar in a space of socio-demographic multi-dimensions^{29,30} (*homophily*). Additionally, *contextual influence* (e.g., sharing the same environment: country, community, etc.) can be pivotal in acquiring specific traits³¹.

According to the literature³², social influence (contagion), social selection (homophily), and contextual influence (confounding) can lead to assortative mixing. Namely, people live most of their time in clusters of similar peers wherein opinions are formed and reinforced.

Assortativity is the key variable in our models. Therefore, we limit detecting the positive contribution of assortativity to predicting opinions about COVID-19 vaccination (outcome variable). (Disentangling the factors responsible for assortativity is beyond the scope of our study.) We observe the behavior of our key variable by controlling for some other potentially relevant aspects of our real-world networks. We account for *network composition* (actor features such as age, education, sex, and income). Also, we employ *betweenness centrality* to measure a given actor's importance for the information flow between pairs of nodes³³. This property identifies the nodes that control the circulation of opinions about vaccination in a network.

Next, the *network density* (how many possible ties are observed) gives information about the speed of opinion circulation. We expect high-density networks to exert a higher social control over their members. Furthermore, consequently, to enhance a specific opinion. Low scores lead to brokerage positions (people connecting social circles (peers) that otherwise remain disconnected). Then, the number of *components* (parts of the network completely disconnected from one another) indicates divisions in the structure and potential lines of cleavages

(e.g., pro and against vaccination sub-groups). Last, *network centralization* (the tendency of a single node to be more central than all the other nodes) shows whether positional advantages have unequal distributions in personal networks. This structural analysis of the local neighborhoods shows how strategic network positions can be related to vaccination opinions (how the social texture affects individuals' opinions).

When predicting alters' opinions on vaccination, we employ multilevel analysis with random ego-level ("group-level") intercepts to account for two properties of our data. First, alters connected to the same ego will likely share unobserved variables (e.g., political attitudes, trust in institutions), which may influence their vaccination opinions. Second, all alters' variables are reported by their respective egos—and we could suspect that different egos have a different understanding of what constitutes a positive or negative opinion. By allowing the expected ratio of pro-vaccination alters to vary randomly by ego, our analysis essentially seeks to predict the difference of opinion among alters in the same personal network.

Results

Tables 1 and 2 provide detailed descriptive statistics about the variables of interest. In Table 1, we note that our 443 respondents (egos) are preponderantly pro-vaccination ($f = 319$; 72.0%), females ($f = 335$; 75.6%), have higher education ($f = 285$; 64.3%), and a monthly income between the national minimum and median salary ($f = 193$; 43.6%). Overall, in their social contexts, the interviewees are surrounded by females ($f = 2717$; 61.3%) and contacts who reportedly have pro-vaccination opinions ($f = 2895$; 65.4%). Further, people with higher education are slightly prevalent in egos' networks ($f = 2226$; 50.3%). Table 2 shows that respondents are relatively young (*Median* = 34.0 years old, *Range* = 56.0) and embedded in personal networks of durable relationships (*ego-alter tie duration: Mdn* = 23.0 years; $R = 77.0$). Their networks have only one component (*components: Mdn* = 1.0, $R = 9.0$), with a low degree of centralization (*centralization: Mdn* = 0.3, $R = 0.7$) and moderate density (*density: Mdn* = 0.4, $R = 1.0$).

Looking at the 4430 alters (Table 2), their median age is similar to the one of the egos (*Mdn* = 40.0; $R = 72.0$), which may express an age selection effect³⁴. Given the high number of alters in each network (*Mean* = 6.5, *Std. Dev.* = 3.0, *Mdn* = 7, $R = 10$), the number of ties to peers that are in favor of COVID-19 vaccination ('ties to pro-vaccination peers', $M = 2.9$, $SD = 2.2$, *Mdn* = 3.0, $R = 9.0$) is higher than the one to peers that are against ('ties to anti-vaccination peers', $M = 1.0$, $SD = 1.5$, *Mdn* = 0.0, $R = 9.0$). Further, alters have, on average, 4.2 ties ($SD = 2.5$, *Mdn* = 4.0, $R = 9.0$) and display low betweenness scores ($M = 2.1$, $SD = 4.8$, *Mdn* = 0.0, $R = 34.0$). We emphasize that the maximum number of ties that an alter can have in a personal network (the degree centrality of an alter) is nine (all personal networks have ten alters). Visualizations of the descriptive statistics are available in the Supplementary Material.

In Table 3, we report the results of the multilevel logistic regression models fitted to predict alters' opinions about COVID-19 vaccination. And specifically, to detect evidence of possible assortativity effects in personal networks.

	Egos		Alters	
Opinions about COVID-19 vaccination				
Very good & good	319	72.01%	2895	65.35%
Very bad & bad	87	19.64%	1055	23.81%
Missing	37	8.35%	480	10.84%
Total	443	100.00%	4430	100.00%
Sex				
Male	108	24.38%	1713	38.67%
Female	335	75.62%	2717	61.33%
Missing	0	0.00%	0	0.00%
Total	443	100.00%	4430	100.00%
Higher education studies				
Yes	285	64.33%	2226	50.25%
No	158	35.67%	2204	49.75%
Missing	0	0.00%	0	0.00%
Total	443	100.0%	4430	100.00%
Income				
Less than minimum wage	78	17.61%		
In-between minimum & median wage	193	43.57%		
In-between median wage & median wage plus one minimum wage	145	32.73%		
More than median wage plus one minimum wage	27	6.09%		
Missing	0	0.00%		
Total	443	100.00%		

Table 1. Descriptive statistics. Categorical variables of interest.

Egos						
	Age	Betweenness	Constraint	Centralization	Density	Components
Mean	36.47	24.07	0.32	0.31	0.47	1.67
Std.Dev	11.20	8.90	0.04	0.15	0.20	1.28
Min	19.00	0.00	0.10	0.00	0.00	1.00
Median	34.00	25.00	0.33	0.31	0.44	1.00
Max	75.00	45.00	0.38	0.73	1.00	10.00
Skewness	0.66	- 0.56	- 1.89	0.06	0.57	3.07
SE.Skewness	0.04	0.12	0.12	0.12	0.12	0.12
Kurtosis	- 0.17	0.24	5.40	- 0.17	0.24	12.56
N.Valid	443.00	443.00	443.00	443.00	443.00	443.00
Alters						
	Age	Betweenness	Ties to anti-vaccination peers	Ties to pro-vaccination peers	Degree	Ego alter tie duration (years)
Mean	41.71	2.14	1.04	2.90	4.19	22.82
Std.Dev	15.05	4.82	1.48	2.23	2.49	14.99
Min	18.00	0.00	0.00	0.00	0.00	1.00
Median	40.00	0.00	0.00	3.00	4.00	23.00
Max	90.00	34.00	9.00	9.00	9.00	78.00
Skewness	0.57	3.17	1.81	0.74	0.32	0.34
SE.Skewness	0.04	0.04	0.04	0.04	0.04	0.04
Kurtosis	- 0.29	10.74	3.49	0.00	- 0.74	- 0.66
N.Valid	4430.00	4430.00	4212.00	4212.00	4430.00	4430.00

Table 2. Descriptive statistics. Numeric variables of interest.

	Model 0 ('intercept')		Model 1 ('attributes')		Model 2 ('network')		Model 3 ('full')	
	Log-odds (95% CI)	OR (p)	Log-odds (95% CI)	OR (p)	Log-odds (95% CI)	OR (p)	Log-odds (95% CI)	OR (p)
alter sex			0.08 (- 0.12; 0.27)	1.08 (0.435)			0.11 (- 0.08; 0.31)	1.12 (0.260)
alter edu			0.46 (0.25; 0.67)	1.58 (<0.001)			0.44 (0.23; 0.65)	1.55 (<0.001)
alter age			0.15 (0.04; 0.26)	1.16 (0.005)			0.16 (0.04; 0.28)	1.17 (0.008)
ego sex			0.08 (- 0.27; 0.42)	1.08 (0.658)			0.06 (- 0.29; 0.41)	1.06 (0.739)
ego edu			- 0.28 (- 0.61; 0.05)	0.75 (0.094)			- 0.28 (- 0.62; 0.06)	0.75 (0.102)
ego income			0.18 (- 0.01; 0.38)	1.20 (0.069)			0.19 (- 0.01; 0.39)	1.21 (0.065)
ego age			0.05 (- 0.11; 0.20)	1.05 (0.561)			0.04 (- 0.12; 0.20)	1.04 (0.612)
ego covid			2.14 (1.80; 2.48)	8.48 (<0.001)			2.12 (1.77; 2.46)	8.30 (<0.001)
ego alter duration					0.05 (- 0.05; 0.15)	1.05 (0.324)	0.00 (- 0.12; 0.12)	1.00 (0.997)
assortativity					0.30 (0.21; 0.38)	1.34 (<0.001)	0.27 (0.19; 0.36)	1.31 (<0.001)
alter betw					0.07 (- 0.03; 0.17)	1.07 (0.156)	0.10 (0.00; 0.20)	1.10 (0.058)
comp					0.03 (- 0.18; 0.25)	1.04 (0.758)	0.09 (- 0.09; 0.27)	1.10 (0.312)
dens					0.05 (- 0.19; 0.28)	1.05 (0.696)	0.09 (- 0.10; 0.28)	1.10 (0.345)
centraliz					0.03 (- 0.18; 0.24)	1.03 (0.809)	0.03 (- 0.14; 0.20)	1.03 (0.736)
Intercept	1.46 (1.28; 1.65)	4.32 (<0.001)	- 0.69 (- 1.17; - 0.20)	0.50 (0.005)	1.47 (1.28; 1.66)	4.36 (<0.001)	- 0.66 (- 1.16; - 0.17)	0.51 (0.008)
Num. obs	3588		3588		3588		3588	
Num. groups: ego_id	401		401		401		401	
ICC	0.414		0.260		0.412		0.265	
AIC	3616.287		3462.61		3576.006		3428.7	
BIC	3628.658		3524.464		3625.489		3527.665	
Log Likelihood	- 1806.144 (df=2)		- 1721.305 (df=10)		- 1780.003 (df=8)		- 1698.35 (df=16)	
Var (SD): ego_id (Intercept)	2.323 (1.524)		1.153 (1.074)		2.305 (1.518)		1.183 (1.088)	

Table 3. Multilevel logistic regression models explaining alters' opinions about vaccination. Significant values are in bold.

We find that alters with a higher level of education (*Model 1*, OR 1.58, 95% CI 1.28, 1.95, $p < 0.001$; *Model 3*, OR 1.55, 95% CI 1.26, 1.91, $p < 0.001$) and older alters (*M1*, OR 1.16, 95% CI 1.05, 1.29, $p = 0.005$; *M3*, OR 1.17, 95% CI 1.04, 1.32, $p = 0.008$) are more likely to have a positive opinion on COVID-19 vaccination. The only significant effect among the network-based covariates is the assortativity variable, which is consistently positive (*M2*, OR 1.34, 95% CI 1.24, 1.46, $p < 0.001$; *M3*: OR 1.31, 95% CI 1.21, 1.43, $p < 0.001$). This indicates that alters with similar opinions cluster together. Namely, people are more likely to have a positive attitude if their neighbors, on average, have a more positive attitude than the average attitude in the network. We also find that ego's opinion about vaccination makes a statistically significant contribution in predicting alters' opinions (*M1*, OR 8.48, 95% CI 6.03, 11.94, $p < 0.001$; *M3*: OR 8.30, 95% CI 5.87, 11.73, $p < 0.001$). We note that our multilevel models are mixed-effects models since they contain both random effects (namely the ego-specific intercept) and fixed effects (all other parameters in Table 3).

We assess the robustness of our results (Table 3) by fitting standard logistic regression models without any multilevel structure (Table 4). In these models, the additional variable *proportion of alters that are pro-vaccination (excluding the alter of reference)*, i.e., *prop vacc ex alter*, controls for the average opinion of all other alters in the network. Interestingly, the two families of models (the multilevel logistic and standard logistic regression models) qualitatively yield the same results for almost all effects. The only qualitative difference in the standard logistic regression models, compared to the multilevel models, is that in the "joint" model (*M3*, in Table 4), alters whose ego has a higher level of education are less likely to have a positive opinion (*M1*, OR 0.79, 95% CI 0.65–0.96, $p = 0.021$; *M3*, OR 0.80, 95% CI 0.65–0.98, $p = 0.030$). While this might seem strange at first glance, we have to consider that all alter data is reported by ego. The negative effect of ego's education could reveal a social prejudice that higher educated people think more often that their alters have a negative attitude toward vaccination. Similarly, in the multilevel logistic regression models (Table 3), ego's education has a negative effect. However, this is not statistically significant (*M1*, OR 0.75, 95% CI 0.54, 1.05, $p = 0.094$; *M3*, OR 0.75, 95% CI 0.54, 1.06, $p = 0.102$).

In Table 4, the additional control variable *prop vacc ex alter* (proportion of alters that are pro-vaccination, excluding the alter of reference), giving the average opinion among the alters in the same network (minus the alter of reference), has a strong positive effect on the attitude of the alter of reference. While this was expected (in fact, everything else would be a surprise), it underlines the need to control for the average opinion in the network. The multilevel models control for varying average opinions via the random ego-level intercepts.

Discussion

Our study suggests that people with similar opinions about COVID-19 vaccination tend to cluster together (or be partitioned) in personal networks. We find assortativity to be a positive statistically significant effect (which directly contributes to our research objective). Namely, assortative mixing of opinions among the alters of the same ego. The likelihood of having a positive opinion about the COVID-19 vaccination increases when peers (neighbors) have, on average, a more positive attitude than the rest of the nodes in the network. Further, we

	Model 0 ('intercept')		Model 1 ('attributes')		Model 2 ('network')		Model 3 ('full')	
	Log-odds (95% CI)	OR (p)	Log-odds (95% CI)	OR (p)	Log-odds (95% CI)	OR (p)	Log-odds (95% CI)	OR (p)
alter sex			0.05 (– 0.13; 0.23)	1.05 (0.568)			0.09 (– 0.09; 0.28)	1.10 (0.318)
alter edu			0.43 (0.25; 0.62)	1.54 (< 0.001)			0.41 (0.22; 0.60)	1.51 (< 0.001)
alter age			0.14 (0.04; 0.23)	1.15 (0.006)			0.14 (0.03; 0.25)	1.15 (0.013)
ego sex			0.02 (– 0.19; 0.23)	1.02 (0.845)			– 0.00 (– 0.22; 0.21)	1.00 (0.988)
ego edu			– 0.24 (– 0.44; – 0.04)	0.79 (0.021)			– 0.23 (– 0.43; – 0.02)	0.80 (0.030)
ego income			0.02 (– 0.10; 0.14)	1.02 (0.719)			0.02 (– 0.10; 0.14)	1.02 (0.731)
ego age			– 0.03 (– 0.13; 0.07)	0.97 (0.549)			– 0.04 (– 0.15; 0.06)	0.94 (0.424)
ego covid			0.72 (0.50; 0.94)	2.05 (< 0.001)			0.67 (0.44; 0.89)	1.95 (< 0.001)
ego alter duration					0.04 (– 0.05; 0.12)	1.04 (0.412)	0.01 (– 0.09; 0.12)	1.01 (0.793)
assortativity					0.28 (0.20; 0.35)	1.32 (< 0.001)	0.26 (0.18; 0.33)	1.30 (< 0.001)
alter betw					0.08 (– 0.01; 0.18)	1.09 (0.081)	0.10 (0.00; 0.19)	1.10 (0.050)
comp					0.05 (– 0.06; 0.16)	1.05 (0.358)	0.08 (– 0.03; 0.19)	1.08 (0.175)
dens					0.07 (– 0.04; 0.18)	1.07 (0.225)	0.09 (– 0.02; 0.21)	1.10 (0.107)
centraliz					0.03 (– 0.06; 0.13)	1.04 (0.491)	0.04 (– 0.06; 0.14)	1.04 (0.467)
prop vacc ex alter	1.01 (0.93; 1.10)	2.74 (< 0.001)	0.82 (0.72; 0.92)	2.71 (< 0.001)	1.01 (0.93; 1.10)	2.75 (< 0.001)	0.84 (0.73; 0.94)	2.31 (< 0.001)
Intercept	1.27 (1.18; 1.36)	3.56 (< 0.001)	0.56 (0.23; 0.88)	3.27 (< 0.001)	1.28 (1.19; 1.37)	3.61 (< 0.001)	0.61 (0.28; 0.94)	1.84 (< 0.001)
Num. obs	3588		3588		3588		3588	
AIC	3428.97		3377.527		3380.57		3338.17	
BIC	3441.34		3439.38		3430.053		3437.136	
Log Likelihood	– 1712.485 (df=2)		– 1678.763 (df=10)		– 1682.285 (df=8)		– 1653.085 (df=16)	
Deviance	3424.97		3357.527		3364.57		3306.17	

Table 4. (Standard) logistic regression models explaining alters' opinions about vaccination. Significant values are in bold.

discover a social correlation between the opinions held by egos (respondents) and alters (social contacts). This positive association suggests a possible ego-alter contagion effect or, alternatively, a social selection effect in the sense that egos tend to select alters with the same opinion. Unfortunately, our data do not permit a detailed examination of the causes of this assortative mixing. Future research and longitudinal data are needed to distinguish between social selection, contagion (or social influence), and confounding.

Our models also control for the attributes of the actors (education, sex, age) and their structural positions in the networks. We discover that alters with higher education³⁵ and older³⁶ are more likely to hold a positive opinion. Additionally, educated respondents (egos) think more often than those less educated that their social contacts (alters) have a negative opinion about vaccination. This result may indicate social prejudices (perceptions about the virtues of vaccination associated with education). Also, it might be a sign that egos with different levels of education have a different understanding of what constitutes a positive opinion. At the same time, the structural positions (node-level betweenness) and the organization of the relationships in personal networks (components, density, centralization) do not significantly contribute to predicting COVID-19 vaccination opinions.

Our readers should note several things regarding the interpretation of our results. An inherent feature of any network research design is that *egos* (respondents) report information about *alters* (their social contacts). Potentially, this can create biases percolating through the models and interpretations: false consensus effect³⁷ (the tendency to see one's own choices as relatively common and appropriate) and inaccurate reports³⁸. We included the duration of ego-alter ties (years) in our models (75% of all ego-alter ties have a duration of at least eight years). Also, we asked our respondents to provide information about those people with whom they communicate most often. At the same time, we employed two different statistical procedures (multilevel and standard logistic regression models) that eventually yielded similar findings. These remedies should counteract these biases' effects and improve the quality of the collected data. That is, we expect people who frequently interact over longer intervals to have more accurate data on their peers. Longitudinal cohort (balanced panel) data can improve control of the magnitude of these biases in the future. However, it cannot completely filter them out. Therefore, as a caveat, our findings may reflect the perceived clustering of vaccination opinions within networks as reported by the egos, which is relevant in its own right as perceptions can influence behavior. We acknowledge this limitation and suggest it as a direction for future research, proposing a design that includes the direct collection of vaccination opinions from both egos and alters. Also, future work can replicate our design in diverse settings and gather longitudinal data that may be beneficial in investigating related topics such as the evolution of opinion polarization. Additionally, when employed alongside Computer-Assisted Web Interviewing (CAWI), different data collection methodologies can provide valuable insights into potential biases inherent in relying solely on CAWI for data gathering.

In sum, our study claims that assortativity impacts COVID-19 vaccination opinions. Thus, we align to the stream of work that has already shown the role of assortativity in vaccination dynamics^{8,39} and status⁴⁰ or disease spread⁴¹. In the special case of COVID-19 vaccination, in our sample, people with positive opinions declare having in their social proximity peers holding rather similar positive opinions. We suspect these results can be generalized to the whole population. We build on the growing evidence affirming not only the opinion clustering^{42,43} or social selection in the adoption of health behavior⁴⁴ but also the clustered vaccination adoption (childhood vaccination refusals^{45,46}, seasonal influenza vaccine uptake⁴⁷, the imitation of vaccination behavior⁴⁸). However, only future work could reveal if fixed network size leads to a biased mix of favorable and unfavorable vaccination opinions⁴⁹. Our study can have broader implications that go beyond the specificity of the COVID-19 context and reach future pandemics. Understanding people's opinions on vaccination can act as an early warning system for potential outbreaks, inform predictions about vaccination uptake, or help supply chain management for vaccine distribution. Public health officials may address the challenges posed by assortativity in social networks and educational disparities by employing various strategies. For example, targeted communication strategies (e.g., social media campaigns tailored to specific social networks or communities), educational outreach programs (e.g., focused on areas with low levels of educational attainment), data-driven approaches and community engagement initiatives.

Methods

Study design, size, and selection of participants

We performed a real-world cross-sectional study and employed a personal network research design⁵⁰ (*penet*). Using computer-assisted web interviewing, we collected questionnaire data from a random panel of 30,000 individuals (the panel was deemed representative of the Romanian population). Individuals were at least 18 years old when filling out the questionnaire and could speak the Romanian language (the questions were formulated in Romanian). We initiated the data collection process on March 16, 2022, and halted it on March 30, 2022 (we stopped due to the lack of new respondents). The actual data collection process was outsourced (yet the research team outlined and created the content of the research design). We sent invitations to all panel members to participate in the study (an invitation and two additional reminders via email). The invitations informed each potential participant about the research objectives and granted anonymity and the possibility of opting out at any moment (even after submitting the filled-out questionnaire). The email invitations also included the link to the questionnaire. The panel members participated in the study on a completely voluntary basis. The data collection process was single-stage. Individuals received an invitation to participate in the study and could choose to accept or decline. Those who agreed to participate were given immediate access to the self-administered questionnaire. This questionnaire included prompts, such as '*Please use acronyms or nicknames when mentioning alters*' to guide responses. An availability sample resulted in valid answers (questionnaires) from 896 respondents (dubbed *egos* according to the *penet* terminology). We note that the socio-demographic structure of our sample did not replicate the structure of the Romanian population.

We followed the conventional *penet* practice and organized the questionnaire into five components. (1) First, we addressed questions referring to the egos (socio-demographic items and opinions about COVID-19 vaccination). Then, (2) we included a *generator of alters* (persons connected to the ego). Each ego was required to elicit ten social contacts or alters (maximum five relatives and five friends). Namely, the egos were supposed to nominate people with whom they communicate most frequently. We restrained the number of alters to ten to avoid the respondent burden⁵¹. For example, a respondent (an ego) nominating ten alters must assess 45 pairs of alter-alter relationships, equating to 45 questions, i.e., $(10 \times 9)/2 = 45$ pairs. If the number of nominated alters were to increase to 20, the number of required assessments will rise to 190 pairs $((20 \times 19)/2 = 190)$. To maintain manageability and ensure high-quality data, we opted for a fixed number of alters, thus preventing undue burden on respondents, e.g., higher numbers of alters could be too time-consuming, mentally taxing, or otherwise overwhelming, potentially leading to lower response rates, incomplete data, or compromised data quality due to respondent fatigue or disengagement. Afterward, (3) we applied a *name interpreter*. We asked egos to report socio-demographic information about each of their nominated alters. Further, participants were asked about alters' opinions on COVID-19 vaccination. (4) We assessed alter-alter ties by inquiring whether each ego was aware of communication occurring between their respective alters in their absence (*Do [Alter X] and [Alter Y] contact each other independently of you?*). Namely, that was a question about the existence of relationships among alters as perceived by the respondent. Lastly, (5) we measured *ego-alter ties* in terms of duration (*For how many years do you know this alter?*). Notably, we instructed the egos to use acronyms or nicknames when eliciting information about alters (social contacts). In this way, we avoided disclosing the identity of the alters and harming them in any way. Additionally, we coached the respondents to devise the acronyms in such a way that it would allow them to respond to the alter-alter-tie questions.

Out of the 896 egos with valid responses, only 443 elicited the theoretical maximum number of ten alters. We kept this sub-sample of 443 egos for the statistical analysis and modeling. We excluded smaller ego-networks (network size less than ten) to ensure methodological consistency and comparability. This approach mitigated the potential for confounding effects that could arise from analyzing networks of varying sizes. (For instance, as shown elsewhere¹⁹, complex contagion is a function of the number of individuals that act as active social reinforcement sources; in our case, either pro or against COVID-19 vaccination. Researchers should be aware of several risks before replicating or performing a similar network data collection. Notably, we refer to low response rates, technical selection biases leading to non-representation (e.g., some groups may have limited access to internet), difficulties in capturing complex network relationships through standard survey methods, and the risk of skewed results due to response bias.

Figure 1 provides an example of the personal networks that resulted after the administration of the questionnaire. First, we underline the multilevel (hierarchical) organization of the data: alters (*the first level*) are grouped by respondents (*the second level*). Next, we stress the existence of *within-network dependencies*: various patterns display how alters are interconnected. Also, we highlight the attributes describing *the nodes* (both egos and alters) and *the ties* (both ego-alter and alter-alter relationships). The variables included in Fig. 1 are only a selection for expository purposes.

Variables

In this Section, we narratively describe our variables of interest, while in the following Section (*'The assortativity variable'*), we present them in a summarized list format. We start the presentation of the variables used in our study with the *actor-level variables*, and then we continue with the *network-level variables*. We collected socio-demographic data referring to *sex* (0: males, 1: females), *education* (0: no higher education, 1: higher education), *age* (numerical, ≥ 18 years old), and *income* (0: less than minimum wage, 1: between minimum & median wage, 2: one minimum wage over the median wage, 3: more than one minimum wage over the median wage; the thresholds reflect Romanian national wages at the moment of data collection). For clarification purposes, the *education*

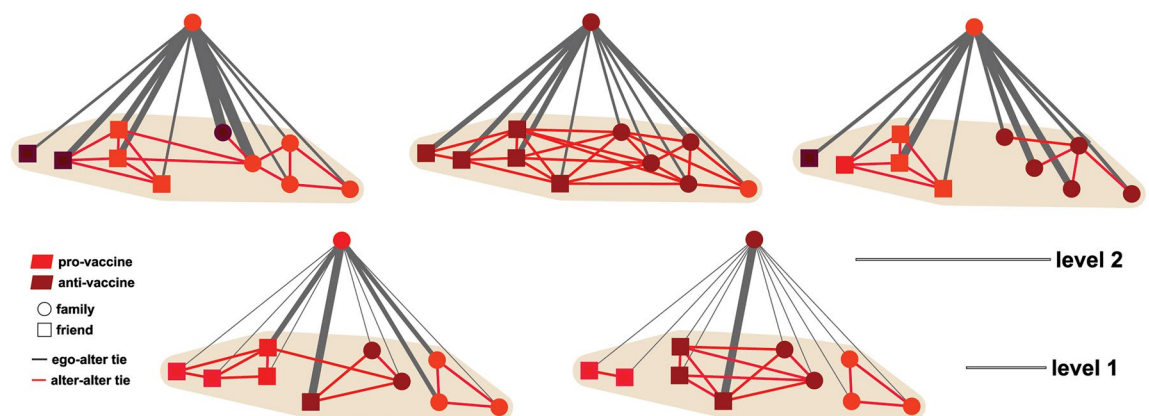


Figure 1. Networks resulted from the personal network research design. Ego-alter tie duration (in years) is marked by the tie thickness. Alters (level 1) are nested in egos (level 2). The horizontal lines indicating the data levels (level 1 and level 2) serve solely to denote these levels without reflecting any actual personal network connections.

variable in our study categorizes individuals into two groups: those who do not have a university degree, representing 'no higher education,' and those who possess a university degree, representing 'higher education.' We also measured *actors' opinions about COVID-19 vaccination*. Each study participant was required to answer the following question: *What opinion do you have about COVID-19 vaccination?* As stated in the presentation of the study design, we did not interview the alters (this is a common practice in personal network research studies). For this reason, we used a *proxy* to capture information about them. Namely, we requested egos to inform us about the socio-demographics (sex, age, education) and the opinions of each of their nominated alters about vaccination (*What opinion does Alter X have about COVID-19 vaccination?*). The questions concerning COVID-19 vaccination had the following pre-defined answers: *very bad, bad, good, and very good*. We intentionally omitted a neutral point given several methodological arguments, e.g., avoiding ambiguity, establishing a clear direction (either positive or negative), better discrimination among participants, and increasing engagement in providing answers. Later, we re-coded the responses into binary variables: either an ego (alter) has a *positive* (good or very good) or *negative* (bad or very bad) opinion vis-à-vis COVID-19 vaccination.

In terms of network-level measurements, first, we accounted for the *composition of the personal networks*. We computed the proportion of female alters, alters with higher education, alters with a positive opinion about COVID-19 vaccination, and the average age of the alters. We derived these variables from the actor-level variables (see the paragraph above). Then, we summarized *the properties of the ego-alter ties* in each personal network: the mean duration (in years). Further, we computed *node-level properties* for all the nodes (both egos and alters), such as Freeman's betweenness centrality (i.e., it quantifies the number of times a node acts as a bridge along the shortest path between two other nodes^{33,52}). We also scrutinized the maximum scores of Freeman's betweenness centrality among the alters with a positive opinion about vaccination (i.e., people in favor of vaccination that connect social circles that otherwise would remain disconnected).

Lastly, we looked at the alter-alter ties and described the *overall structure of each personal network*. We calculated the density (the ratio between the number of observed ties and the number of theoretically possible ties), the centralization (i.e., the tendency of the ties in a network to revolve around a node; as the centralization score increases, this tendency becomes more pronounced⁴³), and the number of components (i.e., a subset of the network's nodes (individuals, in our case) and the ties (relationships or interactions) connecting them, such that every node in the component is reachable from every other node within the same component, but not from nodes in different components). To accurately calculate the number of components containing alters within a personal network, we removed the ego. This is because the ego is connected to all alters; if we include egos in the analysis, their personal networks will consist of only one component⁵⁰. A detailed discussion of the network measurements is available for the interested readership in the Supplementary Material.

The 'assortativity' variable

Overall, for the analysis, we used data on $n_e = 443$ egos and $n_a = 4430$ alters. For alter i , let $j[i]$ denote the ego of i . Note that there is no overlap among the alters of different egos, so each alter is assigned to exactly one ego.

For each ego j , the binary variable $y_j^{(e)}$ indicates j 's vaccination opinion (1 for positive opinion; 0 for negative opinion). Likewise, for each alter i , the binary variable $y_i^{(a)}$ indicates i 's vaccination opinion (1 for positive opinion; 0 for negative opinion). In our analysis, we estimated models explaining alters' opinions $y_i^{(a)}$ and egos' opinions $y_j^{(e)}$ (dependent variables). Given the objectives of our study, we report here only the models predicting alters' opinions $y_i^{(a)}$. However, the models predicting egos' opinions $y_j^{(e)}$ are available in the Supplementary Material. Interested readers may consult these *ego models* if they look for further insights to contextualize the results reported in the body of the paper.

For ego j , $j = 1, \dots, n_e$, we have a vector of covariates u_j , comprising the following covariates:

Attribute-level

- u_{j1} : ego's education ('ego edu')
- u_{j2} : ego's sex ('ego sex')
- u_{j3} : ego's income ('ego income')
- u_{j4} : ego's age ('ego age')

Network-level

- u_{j5} : proportion of female alters ('prop fem')
- u_{j6} : proportion of alters with higher education studies ('prop edu')
- u_{j7} : mean age of alters ('mean age')
- u_{j8} : proportion of alters in favor of vaccination ('prop vacc')
- u_{j9} : ego-alter mean duration ('mean duration')
- u_{j10} : ego-betweenness ('ego betw')
- u_{j11} : highest betweenness of an alter that is in favor of vaccination ('alter betw max pro')
- u_{j12} : centralization ('centraliz')
- u_{j13} : network components ('comp')

For alter i , $i = 1, \dots, n_a$, we have a vector of covariates x_i , comprising the following covariates:

Attribute-level

- x_{i1} : alter's education ('alter edu')
- x_{i2} : alter's age ('alter age')
- x_{i3} : alter's sex ('alter sex')
- moreover, x_i contains all the ego covariates $u_{j[i]}$ and ego's opinion $y_j^{(e)}$ (note that these variables are well-defined since each alter i is assigned to exactly one ego $j[i]$).

Network-level

- x_{i4} : proportion of alters in favor of vaccination, except for the alter of reference ('prop vacc ex alter')
- x_{i5} : duration of ego-alter ties ('ego alter tie duration')
- x_{i6} : alter covariate assortativity ('assortativity')
- x_{i7} : alter-betweenness ('alter betw')
- x_{i8} : number of components in a personal network ('components')
- x_{i9} : network density ('density')
- x_{i10} : network centralization ('centraliz')

Of particular relevance (and requiring additional explanation) is the alter covariate *assortativity variable* ('assortativity'), which tests whether alter's opinion is likely to be influenced by the opinions of those alters to which they are connected. Quantitatively, the assortativity variable indicates whether those alters connected to alter i have, on average, a higher or a lower vaccination opinion than all of $j[i]$'s alters, different from i . More precisely, for alter i , let $A_{j[i]}^{(i)}$ be the set of alters of ego $j[i]$, different from i . (Note that in our data, these are always exactly nine alters since the size of all personal networks is ten.) Let N_i be the *neighbors* of alter i , that is, those other alters of ego $j[i]$ who are connected to i by an alter-alter tie. Then, for an alter i with $|N_i| > 0$ (that is, excluding the isolated alters), we define the assortativity variable to be the difference between the average opinion of i 's neighbors and the average opinion among all of $j[i]$'s alters, different from i . In formulas,

$$\text{assortativity}_i = \frac{\sum_{i' \in N_i} y_{i'}^{(a)}}{|N_i|} - \frac{\sum_{i' \in A_{j[i]}^{(i)}} y_{i'}^{(a)}}{|A_{j[i]}^{(i)}|}$$

For alters with no neighbors (that is, isolated alters with $|N_i| = 0$), we consider this variable as undefined and drop the respective alter from the analysis.

To provide an example, assume that among the nine alters of ego $j[i]$ that are different from i , six have a positive opinion. Further, assume that alter i has alter-alter ties to four alters, among which two have a positive opinion. Then i 's assortativity variable would equal to $\frac{2}{4} - \frac{6}{9} = -\frac{1}{6}$. Indeed, i 's neighbors have a below average opinion for this network, justifying the negative value.

We emphasize that the definition of the assortativity variable for alter i does not take into account i 's vaccination opinion to avoid circular dependency in the data.

The assortativity variable tells us whether i 's neighbors are more or less positive about vaccination than all the other alters of ego $j[i]$. The normalization obtained by subtracting the average opinion over the other alters is necessary since, without this normalization, the effect of this variable would be confounded by the overall ratio of positive opinion in $j[i]$'s network.

In a network with a high ratio of positive opinion, we would expect by chance alone that the average opinion among the neighbors of every alter is likely to lean on the positive side—and in addition, most alters in this network themselves are expected to have a positive opinion. Thus, without normalization, we would expect a positive correlation (taken over all the alters in our data) between i 's opinion and the average opinion over i 's neighbors—even if there is no assortativity of opinion present. (As a simple exercise, we estimated models with the average opinion of i 's neighbors as a covariate and found a strong positive effect—which we claim to be a useless finding since we cannot tell whether that effect indicates assortativity or just a correlation of opinion caused by varying positive ratios over the egos.) By the normalization, we capture in the assortativity variable whether i 's neighbors are more or less positive than the rest of the alters in the same network.

Statistical models

We estimated models explaining alters' opinion about COVID-19 vaccination, $y_i^{(a)}$, via multilevel logistic regression with random ego-level intercepts⁵³. The multilevel approach is necessary to account for two characteristics of our data. First, alters are clustered within egos (see the *study design* previously presented). Therefore, alters of the same ego might have similar values in unobserved covariates, e.g., political attitude, trust in institutions, etc. Second, alters' opinions are always reported by ego, and it is questionable whether all egos have the same understanding of "positive" or "negative" opinions. In fact, by the multilevel approach, we do not attempt to explain the absolute level of opinion of the alters—but rather whether alters have a more or less positive opinion compared to the other alters in the same network. An alternative approach using fixed (rather than random) ego-level intercepts is not feasible since there are egos in whose networks no alter or all alters have a positive opinion—which would theoretically lead to ego-level intercepts that are minus infinity or infinity, respectively. Formally, the multilevel logistic regression models specify, for each alter $i = 1, \dots, n_a$, the probability that i has a positive opinion via

$$\text{logitPr}(y_i^{(a)} = 1) = \alpha_{j[i]} + \beta x_i$$

$$\alpha_j \sim N(\mu_e, \sigma_e^2)$$

In the first equation, the probability that alter i has a positive opinion is the logistic transformation of the ego-level intercept $\alpha_{j[i]}$ plus the sum over the values in the covariate vector x_i , multiplied with the values in the parameter vector β . In the second equation, the ego-level intercepts α_j , for $j = 1, \dots, n_e$, are assumed to be drawn from a normal distribution with mean μ_e and variance σ_e^2 . Given the data, we estimate $\mu_e, \sigma_e^2, \alpha_j$, for $j = 1, \dots, n_e$, and the parameter vector β with the function *glmer* from the R package *lme4*⁵⁴.

As an alternative to the multilevel models described, we fit standard logistic regression models that have no ego-level intercept but instead include the average opinion over all other alters $\frac{\sum_{i' \in A_{j[i]}^{(i)}} y_{i'}^{(a)}}{|A_{j[i]}^{(i)}|}$ (compare with the discussion of the assortativity variable) as an additional predictor for explaining the alter's opinion $y_i^{(a)}$. This additional variable *prop vacc ex alter* controls for the average opinion of all other alters in the network (see “Results” Section). By doing so, we explain the difference between the opinion of alter i and that of all other alters. In other words, we explain why some alters have a higher or lower opinion than the other alters in the same network. We believe the multilevel models are a better and more principled approach for our data. On the other hand, the pure logistic regression models (without any multilevel structure) are easier to understand and probably more common. In any case, it is interesting to compare the results yielded by the two classes of models.

Additionally, we estimated standard logistic regression models to predict egos' opinions about COVID-19 vaccination, $y_j^{(b)}$. These models are not directly relevant to the objectives of our study. Yet, it may be useful for further insights into the general topic of opinions about vaccination. For this reason, these models are included and presented in the Supplementary Material.

The dataset, the entire code, and all the statistical analyses (diagnostics, further information about the variables included in the models and detailed results) are open access⁵⁵ and available for replication (see also Supplementary Material). Moreover, we annotated the code (R script) with explanatory comments and spatially arranged it for human readability. In the code, we made the statistical models exactly reproducible by pseudo-random number generators, which are explicitly seeded.

Statement

All methods were carried out in accordance with the relevant guidelines and regulations (specifically, those provided by the Romanian Sociologists Society, i.e., the Professional Association of Romanian Sociologists).

All the experimental protocol was approved by a named institutional/licensing committee. Specifically, the research protocol was approved by the Center for Innovation in Medicine Ethics Committee (2022/7/2 D) (Bucharest, Romania).

Informed consent was obtained from all subjects.

Ethical declaration

The conducted research was performed in compliance with the Ethics Code of the University of Bucharest. The data collection process was outsourced. The authors analyzed raw data that were already anonymized. The research team received the data set from MEDNET Research based on the written accord no. 1 / 27.07.2022.

Data availability

The dataset analyzed in the current study is available in the *figshare* repository <https://doi.org/https://doi.org/10.6084/m9.figshare.22309174.v1>.

Received: 21 March 2023; Accepted: 5 February 2024

Published online: 09 February 2024

References

1. Prosser Scully, R. COVID vaccines slash risk of spreading Omicron—and so does previous infection. *Nature* <https://doi.org/10.1038/d41586-022-02328-0> (2022).
2. Wang, Y. & Liu, Y. Multilevel determinants of COVID-19 vaccination hesitancy in the United States: A rapid systematic review. *Prev. Med. Rep.* **25**, 101673. <https://doi.org/10.1016/j.pmedr.2021.101673> (2022).
3. Lazarus, J. V. *et al.* A multinational Delphi consensus to end the COVID-19 public health threat. *Nature* **611**, 332–345. <https://doi.org/10.1038/s41586-022-05398-2> (2022).
4. Piltch-Loeb, R. *et al.* Determinants of the COVID-19 vaccine hesitancy spectrum. *PLoS ONE* **17**, e0267734. <https://doi.org/10.1371/journal.pone.0267734> (2022).
5. Christakis, N. A. & Fowler, J. H. Social contagion theory: examining dynamic social networks and human behavior. *Stat. Med.* **32**, 556–577. <https://doi.org/10.1002/sim.5408> (2013).
6. Cheng, S., Pain, C. C., Guo, Y.-K. & Arcucci, R. Real-time updating of dynamic social networks for COVID-19 vaccination strategies. *J. Ambient Intell. Humaniz. Comput.* <https://doi.org/10.1007/s12652-023-04589-7> (2023).
7. Saunders, H. A. & Schwartz, J. M. COVID-19 vaccination strategies depend on the underlying network of social interactions. *Sci. Rep.* **11**, 24051. <https://doi.org/10.1038/s41598-021-03167-1> (2021).
8. Zivich, P. N., Volfovsky, A., Moody, J. & Aiello, A. E. Assortativity and bias in epidemiologic studies of contagious outcomes: A simulated example in the context of vaccination. *Am. J. Epidemiol.* **190**, 2442–2452. <https://doi.org/10.1093/aje/kwab167> (2021).
9. Newman, M. E. J. Mixing patterns in networks. *Phys. Rev. E* **67**, 026126. <https://doi.org/10.1103/PhysRevE.67.026126> (2003).

10. Barclay, V. C. *et al.* Positive network assortativity of influenza vaccination at a high school: Implications for outbreak risk and herd immunity. *PLOS ONE* **9**, e87042. <https://doi.org/10.1371/journal.pone.0087042> (2014).
11. Hãncean, M.-G., Slavinec, M. & Perc, M. The impact of human mobility networks on the global spread of COVID-19. *J. Complex Netw.* **8**, cnaa041. <https://doi.org/10.1093/comnet/cnaa041> (2020).
12. Hãncean, M.-G. *et al.* The role of age in the spreading of COVID-19 across a social network in Bucharest. *J. Complex Netw.* **9**, cnab026. <https://doi.org/10.1093/comnet/cnab026> (2021).
13. Looker, K. J. *et al.* Evidence of synergistic relationships between HIV and Human Papillomavirus (HPV): Systematic reviews and meta-analyses of longitudinal studies of HPV acquisition and clearance by HIV status, and of HIV acquisition by HPV status. *J. Int. AIDS Soc.* **21**, e25110. <https://doi.org/10.1002/jia2.25110> (2018).
14. Cheadle, J. E., Stevens, M., Williams, D. T. & Goosby, B. J. The differential contributions of teen drinking homophily to new and existing friendships: an empirical assessment of assortative and proximity selection mechanisms. *Soc. Sci. Res.* **42**, 1297–1310. <https://doi.org/10.1016/j.ssresearch.2013.05.001> (2013).
15. Ogburn, E. L. & VanderWeele, T. J. Vaccines, contagion, and social networks. *Ann. Appl. Stat.* **11**, 919–948. <https://doi.org/10.1214/17-AOAS1023> (2017).
16. Vacca, R. Structure in personal networks: Constructing and comparing typologies. *Netw. Sci.* **8**, 142–167. <https://doi.org/10.1017/nws.2019.29> (2020).
17. Burt, R. S. Network items and the general social survey. *Soc. Netw.* **6**, 293–339. [https://doi.org/10.1016/0378-8733\(84\)90007-8](https://doi.org/10.1016/0378-8733(84)90007-8) (1984).
18. Borgatti, S. P. & Foster, P. C. The Network paradigm in organizational research: A review and typology. *J. Manag.* **29**, 991–1013. https://doi.org/10.1016/S0149-2063_03_0008 (2003).
19. Centola, D. *How Behavior Spreads: The Science of Complex Contagions* (Princeton University Press, 2018).
20. Lerner, J. & Lomi, A. A dynamic model for the mutual constitution of individuals and events. *J. Complex Netw.* **10**, cnac004. <https://doi.org/10.1093/comnet/cnac004> (2021).
21. Hãncean, M.-G., Perc, M. & Lerner, J. Early spread of COVID-19 in Romania: imported cases from Italy and human-to-human transmission networks. *R. Soc. Open Sci.* **7**, 200780. <https://doi.org/10.1098/rsos.200780> (2020).
22. Hãncean, M.-G., Perc, M., Gheorghită, A., Vega Yon, G. G. & Mihăilă, B.-E. The formation of political discussion networks. *R. Soc. Open Sci.* **9**, 211609. <https://doi.org/10.1098/rsos.211609> (2022).
23. Hãncean, M.-G., Perc, M. & Lerner, J. The coauthorship networks of the most productive European researchers. *Scientometrics* **126**, 201–224. <https://doi.org/10.1007/s11192-020-03746-5> (2021).
24. Hãncean, M.-G., Perc, M. & Vlăsceanu, L. Fragmented Romanian sociology: Growth and structure of the collaboration network. *PLOS ONE* **9**, e113271. <https://doi.org/10.1371/journal.pone.0113271> (2014).
25. Shirado, H., Iosifidis, G., Tassioulas, L. & Christakis, N. A. Resource sharing in technologically defined social networks. *Nat. Commun.* **10**, 1079. <https://doi.org/10.1038/s41467-019-08935-2> (2019).
26. Scala, A. & Delmastro, M. The explosive value of the networks. *Sci. Rep.* **13**, 1037. <https://doi.org/10.1038/s41598-022-26961-x> (2023).
27. Borgatti, S. P. Centrality and network flow. *Soc. Netw.* **27**, 55–71. <https://doi.org/10.1016/j.socnet.2004.11.008> (2005).
28. Guilbeault, D. & Centola, D. Topological measures for identifying and predicting the spread of complex contagions. *Nat. Commun.* **12**, 4430. <https://doi.org/10.1038/s41467-021-24704-6> (2021).
29. McPherson, J. M. & Ranger-Moore, J. R. Evolution on a dancing landscape: Organizations and networks in dynamic Blau space. *Soc. Forces* **70**, 19–42. <https://doi.org/10.2307/2580060> (1991).
30. Lerner, J., Lomi, A., Mowbray, J., Rollings, N. & Tranmer, M. Dynamic network analysis of contact diaries. *Soc. Netw.* **66**, 224–236. <https://doi.org/10.1016/j.socnet.2021.04.001> (2021).
31. Zafarani, R., Abbasi, M. A. & Liu, H. *Social Media Mining: An Introduction* (Cambridge University Press, 2014).
32. Cohen-Cole, E. & Fletcher, J. M. Is obesity contagious? Social networks vs. environmental factors in the obesity epidemic. *J. Health Econ.* **27**, 1382–1387. <https://doi.org/10.1016/j.jhealeco.2008.04.005> (2008).
33. Freeman, L. C. Centrality in social networks conceptual clarification. *Soc. Netw.* **1**, 215–239. [https://doi.org/10.1016/0378-8733\(78\)90021-7](https://doi.org/10.1016/0378-8733(78)90021-7) (1978/1979).
34. Hoang, T. *et al.* A Systematic review of social contact surveys to inform transmission models of close-contact infections. *Epidemiology* **30**, 723–736. <https://doi.org/10.1097/EDE.0000000000001047> (2019).
35. Robertson, E. *et al.* Predictors of COVID-19 vaccine hesitancy in the UK household longitudinal study. *Brain. Behav. Immun.* **94**, 41–50. <https://doi.org/10.1016/j.bbi.2021.03.008> (2021).
36. Steinmetz, L. Sociodemographic predictors of and main reasons for COVID-19 vaccine hesitancy in eastern Oslo: A cross-sectional study. *BMC Public Health* **22**, 1878. <https://doi.org/10.1186/s12889-022-14261-y> (2022).
37. Ross, L., Greene, D. & House, P. The “false consensus effect”: An egocentric bias in social perception and attribution processes. *J. Exp. Soc. Psychol.* **13**, 279–301. [https://doi.org/10.1016/0022-1031\(77\)90049-X](https://doi.org/10.1016/0022-1031(77)90049-X) (1977).
38. Feld, S. L. & Carter, W. C. Detecting measurement bias in respondent reports of personal networks. *Soc. Netw.* **24**, 365–383. [https://doi.org/10.1016/S0378-8733\(02\)00013-8](https://doi.org/10.1016/S0378-8733(02)00013-8) (2002).
39. Chang, S. L., Piraveenan, M. & Prokopenko, M. Impact of network assortativity on epidemic and vaccination behaviour. *Chaos Solitons Fractals* **140**, 110143. <https://doi.org/10.1016/j.chaos.2020.110143> (2020).
40. Burgio, G., Steinegger, B. & Arenas, A. Homophily impacts the success of vaccine roll-outs. *Commun. Phys.* **5**, 70. <https://doi.org/10.1038/s42005-022-00849-8> (2022).
41. Klaus, C. *et al.* Assortative mixing among vaccination groups and biased estimation of reproduction numbers. *Lancet Infect. Dis.* **22**, 579–581. [https://doi.org/10.1016/S1473-3099\(22\)00155-4](https://doi.org/10.1016/S1473-3099(22)00155-4) (2022).
42. McPherson, M., Smith-Lovin, L. & Cook, J. M. Birds of a feather: Homophily in social networks. *Annu. Rev. Sociol.* **27**, 415–444. <https://doi.org/10.1146/annurev.soc.27.1.415> (2001).
43. Salathé, M. & Bonhoeffer, S. The effect of opinion clustering on disease outbreaks. *J. R. Soc. Interface* **5**, 1505–1508. <https://doi.org/10.1098/rsif.2008.0271> (2008).
44. Centola, D. An experimental study of homophily in the adoption of health behavior. *Science* **334**, 1269–1272. <https://doi.org/10.1126/science.1207055> (2011).
45. Lieu, T. A., Ray, G. T., Klein, N. P., Chung, C. & Kulldorff, M. Geographic clusters in underimmunization and vaccine refusal. *Pediatrics* **135**, 280–289. <https://doi.org/10.1542/peds.2014-2715> (2015).
46. Parker, A. A. *et al.* Implications of a 2005 Measles outbreak in Indiana for sustained elimination of measles in the United States. *N. Engl. J. Med.* **355**, 447–455. <https://doi.org/10.1056/NEJMoa060775> (2006).
47. Edge, R., Keegan, T., Isba, R. & Diggle, P. Observational study to assess the effects of social networks on the seasonal influenza vaccine uptake by early career doctors. *BMJ Open* **9**, e026997. <https://doi.org/10.1136/bmjopen-2018-026997> (2019).
48. Ndeffo Mbah, M. L. *et al.* The impact of imitation on vaccination behavior in social contact networks. *PLoS Comput. Biol.* **8**, e1002469. <https://doi.org/10.1371/journal.pcbi.1002469> (2012).
49. Wambua, J. *et al.* The influence of COVID-19 risk perception and vaccination status on the number of social contacts across Europe: Insights from the CoMix study. *BMC Public Health* **23**, 1350. <https://doi.org/10.1186/s12889-023-16252-z> (2023).
50. McCarty, C., Lubbers, M. J., Molina, J. L. & Vacca, R. *Conducting Personal Network Research: A Practical Guide* (The Guilford Press, 2019).

51. Stadel, M. & Stulp, G. Balancing bias and burden in personal network studies. *Soc. Netw.* **70**, 16–24. <https://doi.org/10.1016/j.socnet.2021.10.007> (2022).
52. Freeman, L. C. A set of measures of centrality based on betweenness. *Sociometry* **40**, 35–41. <https://doi.org/10.2307/3033543> (1977).
53. Gelman, A. & Hill, J. *Data Analysis Using Regression and Multilevel/Hierarchical Models* (Cambridge University Press, 2007).
54. Bates, D., Mächler, M., Bolker, B. & Walker, S. Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* **67**, 1–48. <https://doi.org/10.18637/jss.v067.i01> (2015).
55. Bunaciu, D. A. *et al.* Replication data for: Assortative mixing of opinions about COVID-19 vaccination in personal networks. Figshare <https://doi.org/10.6084/m9.figshare.22309174.v1> (2023).

Acknowledgements

We thank Bianca-Elena Mihăilă, Iulian Oană, David-Andrei Bunaciu, and Laura Pascal (MEDNET Research) for the support provided to this study.

Author contributions

All authors made substantial contributions to the study and approved the submitted version. M.-G.H. prepared the study, wrote the manuscript, designed the work, interpreted the data, created the visualizations, and performed the analysis. J.L. contributed to the manuscript's writing and revised it, and designed the analysis. M.P., J.L.M., and M.G. contributed to the manuscript's writing and revised it, designed the work, contributed to the preparation of the study, and made suggestions for the interpretation of the data.

Funding

M.-G.H. was supported by The Research Institute of the University of Bucharest (ICUB) (Grant No. 8405 / 22.07.2023) and The Executive Unit for Financing Higher Education, Research, Development and Innovation (UEFISCDI) (PN-III-P4-ID-PCE-2020-2828); J.L. was supported by Deutsche Forschungsgemeinschaft (DFG) (321869138); M.P. was supported by the Slovenian Research Agency (Javna agencija za raziskovalno dejavnost RS) (Grant No. P1-0403).

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-024-53825-3>.

Correspondence and requests for materials should be addressed to M.-G.H.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024