

# Birds of a Feather Flock Together Online: Digital Inequality in Social Media Repertoires

Social Media + Society  
October–December 2021: 1–14  
© The Author(s) 2021  
Article reuse guidelines:  
sagepub.com/journals-permissions  
DOI: 10.1177/20563051211052897  
journals.sagepub.com/home/sms  
 SAGE

Emőke-Ágnes Horvát<sup>1</sup> and Eszter Hargittai<sup>2</sup>

## Abstract

Communication has long been concerned with people's media repertoires, yet little of this approach has extended to the combination of social media platforms that people use. Despite their considerable popularity, research has found that people do not select into the use of social network sites (SNSs) randomly, which has implications for both whose voices are represented on them and where messaging can reach diverse people. While prior work has considered self-selection into one SNS, in this article we ask: how are different SNSs linked by user base? Using national survey data about 1,512 US adults' social media uses, we build networks between SNSs that connect SNS pairs by user base. We examine patterns by subgroups of users along the lines of age, gender, education, and Internet skills finding considerable variation in SNS associations by these variables. This has implications for big data analyses that depend on data from particular social media platforms. It also offers helpful lessons for how to reach different population segments when trying to communicate to diverse audiences.

## Keywords

platform associations, social media adoption, social network sites, media repertoires, survey, Facebook, Twitter, Pinterest, LinkedIn, Instagram

## Introduction

Communication scholarship has long been interested in which particular media channels people use to get information (e.g., Heeter, 1985; Phalen & Ducey, 2012) and to interact with others (e.g., Haythornthwaite, 2001) in a media-rich environment. Despite this long-held understanding that people turn to multiple media for their communication needs, most research on social media use examines use of one such platform in isolation from the use of other such platforms. Even when studies do look at the adoption of multiple social network sites (SNSs) (e.g., Blank & Lutz, 2017; Hargittai, 2007), they do not connect the dots from use of one to another. Although some work has suggested the need to differentiate by both diversity and frequency of platform usage (Hargittai & Hsieh, 2010), little empirical research has taken such an approach. In this article, we apply the media repertoire perspective to the adoption of social media platforms by using social network analysis, which has previously been used to study news consumption patterns (e.g., Mukerjee et al., 2018; Taneja et al., 2012), but not SNS usage.

With the majority of the population in the United States using social media (Pew Research Center, 2019), it is easy to assume that everyone uses such services. Yet the diffusion of SNSs is far from universal. For example, less than a quarter of Americans visit Twitter or Reddit, which is not the

impression one gets from the amount of attention they receive both in the popular press and in academic research (Blank & Lutz, 2017), the latter mostly due to ease of data collection rather than representativeness (Tufekci, 2014). In fact, research has found that people do not select into the use of SNSs randomly (Blank, 2017; Hargittai, 2015; Hargittai & Litt, 2011). In other words, different social media platforms have varying user bases with implications both for the generalizability of research that uses such sites as their sampling frame and for how those wishing to reach diverse audience can do so. For example, key findings about disease tracking (Paul & Dredze, 2021; Sarker et al., 2020), crisis response (Sutton et al., 2015; Vieweg et al., 2010), activism and political participation (Budak & Watts, 2015; Conover et al., 2011; González-Bailón & Wang, 2016; Hemphill & Schöpke-Gonzalez, 2020), and spreading of (mis)information (Bakshy et al., 2011; Huang et al., 2015; Jiang et al., 2020; Romero

<sup>1</sup>Northwestern University, USA

<sup>2</sup>University of Zurich, Switzerland

## Corresponding Author:

Emőke-Ágnes Horvát, Department of Communication Studies,  
Northwestern University, 2240 Campus Dr, Evanston, IL 60208-0001,  
USA.  
Email: a-horvat@northwestern.edu



et al., 2011; Starbird, 2017) rely largely on research that has been conducted on single SNSs.

While research efforts focused on one platform are helpful, their insights may have limited applicability at the time of important communication challenges like during natural disasters (Hargittai, 2020, p. 20) and pandemics (Barry, 2009). Such situations require effective and simultaneous communication with the entire population regardless of demographics and media use (Gans, 2020). The World Health Organization (2017) itself notes the importance of social media communication in crisis situations yet does not give concrete recommendations as to which channels to pursue for effective population-wide communication. This article adds to the existing literature on social media adoption and crisis communication by exploring how use of one SNS relates to the use of another, which has both scholarly implications for expanding understanding of SNS use across the population, and practical implications for advising organizations—whether governmental, nonprofit, or commercial—on how to reach diverse constituents.

Prior research on social media has shown that people select into their uses at different rates. Understanding patterns of social media adoption across sites is important for several reasons. First, methodologically speaking, many studies use just one SNS as their sampling frame (e.g., Twitter or Facebook), thereby biasing their samples (e.g., Bakshy et al., 2015; Romero et al., 2013; Ugander et al., 2012; Weng et al., 2013). By knowing which such services have overlaps in their user base and, importantly, which do not, researchers can diversify their samples without having to include in their sampling frame too many sites *per se*. Second, for digital inequality scholarship, it is important to go beyond understanding inequalities in the use of one site to see whether variations across population groups exist even when considering the use of more than one site. In other words, are people siloing themselves across platforms not just with respect to individual sites? Third, for campaigns—whether commercial (Liang & Turban, 2011), health-related (Neiger et al., 2012), political (Klinger & Svensson, 2015), or otherwise—it is significant to know how they can maximize reaching different types of people on social media (Thackeray et al., 2012). If platform A and platform B have largely overlapping user bases, then there is not much point in spending resources on both; rather, reaching out to platforms whose user bases do not overlap so much is a better way to reach a broader public.

We address this gap in the literature by analyzing a national survey about American adults' social media uses to see how the user base of various such platforms may overlap. We answer this question by identifying pairs of SNSs that have a significant shared user base. Our computational approach is rooted in work on product associations such as “customers who bought product *A* also bought product *B*” (Linden et al., 2003). Comparable methods based on social network analysis techniques have been employed to study

audience overlap in the consumption of digital news (Mukerjee et al., 2018) as well as different online and offline media (Olsen, 2020). The basic idea behind such studies is that two media outlets are connected if they share a fraction of their audiences. Here, we employ a similar approach and connect SNSs if they share a fraction of their user base. Then, we examine differences in the resulting networks by age, gender, education, and Internet skills as these are variables that the literature reviewed below has identified to be important in distinguishing users of different sites.

We rely on data about 1,512 Internet users and their experiences with 10 popular SNSs at the time of data collection in 2016. We show that women and men inhabit different social media universes when going beyond the use of just one such service, which is what most research tends to do. We also find similar variations across age, educational levels, and Internet skills. We end by discussing the implications of these findings both for research methodologies and for substantive questions ranging from digital inequality scholarship to communication that concerns reaching diverse publics.

## Media Repertoires and Differentiated SNS Adoption

Both with respect to how people get their information (Kim, 2016; Mukerjee et al., 2018; Taneja, 2017; Taneja & Webster, 2016; Taneja et al., 2012; Wu et al., 2020) and how they communicate with others (Haythornthwaite, 2001; Hsieh, 2012), communication research has highlighted that it is important to recognize that people do not use one or another medium in isolation of others. Despite this uncontested premise, relatively little empirical work takes a media repertoire perspective beyond the topical focus of news consumption (Dvir-Gvirsman, 2020; Yuan, 2011). While theoretically it is a convincing proposition, empirically it is certainly more complex to investigate the use of multiple channels of communication than it is to focus on one. In work on social media, most scholarship tends to focus on just one platform with relatively rare exceptions (e.g., Blank & Lutz, 2017; Hargittai, 2015; 2020), and such work has rarely used network analytical methods with no work known to us doing so to map the shared user base of different platforms. Platforms differ in their affordances (van Dijck, 2013), which means that although part of the same genre of media, they nonetheless represent diverse options. To this end, there is value in seeing how much their user bases overlap or diverge.

In another stream of communication scholarship, digital inequality research has long studied how user background relates to what people do online. This interest in identifying variations in usage by sociodemographics expanded to SNSs as those started gaining traction. Haight and colleagues (2014) identified differences by gender and education when considering use of SNSs generally speaking without disaggregating the question to specific platforms. Other work has explored differences by specific service.

When MySpace and Facebook were the most popular such sites, boyd (2011) and Hargittai (2007) both documented socioeconomic differences in their adoption finding that those from less privileged backgrounds were more likely to be on MySpace, whereas those from higher socioeconomic status (SES, measured by parental education and income) adopted Facebook at higher rates. As Twitter gained traction, using panel data about a diverse group of young adults collected in 2009 and then again in 2010, Hargittai and Litt (2011) showed that African American young adults were more likely to start using the site than young people from other racial and ethnic backgrounds. They also found socioeconomic differences (measured by parental education) whereby those from the least privileged backgrounds were considerably less likely to adopt the platform than the most privileged (p. 836). A few years later, Blank (2017) investigated this same question on data collected in 2013 from a national sample of British and American adults also finding that those from higher SES were more likely to adopt the service. As social media platforms started to proliferate, the studies about their adoption expanded to additional services. A study of undergraduate students in 2014–2015 at a US university found that women were considerably more likely to use Instagram than men (Sheldon & Bryant, 2016).

Although Hargittai's (2007) early paper of SNSs included several platforms, this was likely viable at the time due to that study's focus on college students. No other work at the time compared several platforms. This changed as social media diffused to the larger population. Blank and Lutz (2017) analyzed data from the 2013 Oxford Internet Surveys to examine how various factors were linked to the use of Facebook, LinkedIn, Twitter, Pinterest, Google+ and Instagram. They found differences by age, gender, and income, which is similar to what Hargittai (2015) found analyzing panel data about a group of 25- to 26-year-olds with 2012 data about SNS use. Unfortunately, neither of these papers report regression results on SNS adoption before controlling for a host of factors beyond sociodemographics that may be explaining differences by such characteristics. That is, Blank and Lutz (2017) control for autonomy of use, Internet skills, self-efficacy, and privacy concerns, while Hargittai also controls for autonomy of use, Internet skills, and frequency of use and number of use years, so it is impossible to tell whether there are differences across population groups concealed due to the model specifications.

Analyzing data from the 2015 British Election Study, Mellon and Prosser (2017) find age, gender, and education differences between Facebook and Twitter users compared with the general population. Survey data about Belgian adults from 2017, 2018, and 2019 also showed that age, gender, and other factors were related to the propensity of using Facebook, Twitter, and Instagram (Hellemans et al., 2020). Gazit and colleagues (Gazit et al., 2019) examined whether the use of Facebook, Instagram, Twitter, and

WhatsApp varied by gender among Israeli college students in 2017–2018. They found large variation in the popularity of these platforms with WhatsApp being the most popular, Twitter the least, and popularity varied by gender whereby women used WhatsApp and Instagram more whereas men used Twitter more.

In sum, while the specific sociodemographics that matter may differ, the overwhelming consensus of this work is that people do not select into the use of such sites randomly; rather, various socioeconomic factors—especially age, gender, and education—relate to who ends up on which site. In addition, those with higher Internet skills are more likely to use social media (Hargittai, 2020). While the above work has been helpful in pointing out such differences, the literature so far has only looked at site adoption in isolation of other site adoption. That is, while papers may have compared the users of various SNSs, they have not looked at shared users. That is the gap in the literature this article fills. We add an important new dimension to the literature on SNS use by asking the research questions: Which SNSs share a user base? What user characteristics are associated with pairs of SNSs?

We draw on methods used in the fields of data mining, information retrieval, and network science to do this. The approach mirrors techniques used for recommender systems such as “Customers who bought book *X*, also bought book *Y*” or “People who liked this movie also liked these others” (Park et al., 2012; Resnick et al., 1994). Recommender systems require knowledge-discovery techniques to make personalized recommendations for information, products, and services. One common approach is to generate recommendations from item-to-item similarities, which can be obtained using a cosine similarity, which we describe in the next section on methods (Deshpande & Karypis, 2004; Sarwar et al., 2001).

## Data and Methods

We draw on a national survey study to explore SNS associations. The data set is based on a national sample of US adults 18 years old and over. Data collection took place online in summer 2016 through the University of Chicago's NORC research unit using their AmeriSpeak panel. The panel is representative of the US population using “area probability sampling and includes additional coverage of hard-to-survey population segments such as rural and low-income households that are underrepresented in surveys relying on address-based sampling” (NORC, n.d.). After pretesting the survey with 23 respondents and updating it based on the results in early May 2016, we ran the survey from 25 May to 5 July 2016. (We note the change in popularity of the various services since 2016 when we discuss the results below.) For survey quality, we included an attention-check question and only analyze responses from participants who passed this question. The 1,512 participants reflect a 37.8% survey response rate.

### Measures: Independent Variables

Background variables about respondents were supplied by NORC based on their earlier data collection about the AmeriSpeak panel. In this article, we analyze SNS associations by age, gender, education, and Internet skills, variables that past work on both digital inequality and SNS adoption has identified as important correlates of SNS adoption. Gender was collected and coded as binary. The survey asked respondents their date of birth, which was used to calculate their age. We created three education categories: high school or less, some college, and college degree or more. We collected data on Internet skills through a widely used measure (Hargittai & Hsieh, 2012) that asks respondents their level of understanding of various Internet-related terms on a 1–5 scale; averaging these responses creates an index score (Cronbach's  $\alpha = .94$ ).

### Measures: Dependent Variable

To measure whether people use SNSs, we asked respondents whether they visit various sites with the following answer options: “No, have never visited it,” “Yes, have visited it in the past, but do not visit it nowadays,” “Yes, currently visit it sometimes,” “Yes, currently visit it often.” We recoded the answers into a binary no (for people who picked the first and second response options) and yes for those who use it currently (respondents who picked the third and fourth options). We collected this information for the following social media platforms, chosen for inclusion due to their popularity during or before the time of data collection: Facebook, GooglePlus, Instagram, LinkedIn, MySpace, Pinterest, Reddit, Snapchat, Tumblr, and Twitter.

### Sample Descriptives

Table 1 shows sample descriptives including figures on the right side of the table from the Pew Research Center's 2015 Spring Tracking survey to offer comparisons of the sample composition. Almost the same number of men and women participated in this study, the average age was 48.7. Just under 30% of respondents were ethnic and racial minorities (11.7% Hispanic, 11.5% African American, 3.1% Asian American, and 1.7% Native American). A quarter had no more than a high school education, just under a third (31.4%) completed some college, and 43.5% had at least a college degree. The average household income was US\$71,478. Just under two-thirds were employed, and 13.4% lived in a rural area. Overall, it is a diverse sample although somewhat less educated, less rural and more racially diverse than Internet user figures according to the Pew Research Center.

Regarding their online experiences, respondents on average have been using the Internet for just over 11 years, have 4.8 locations where they can go online, and spend 14.7 hr on

the Web weekly. Their Internet skills are varied; on a 1–5 scale, they average a 3.4 score.

### Analysis

We generate the association between two SNSs based on whether they have a similar set of users. Specifically, to compute the association between SNS  $A$  and SNS  $B$ , we normalize the set of users who use both of them ( $u_{A \& B}$ ) with the square root of the product of the number of users who use  $A$  ( $u_A$ ) and the number of users who use  $B$  ( $u_B$ ) as follows

$$\cos(A, B) = \frac{|u_{A \& B}|}{\sqrt{|u_A| \cdot |u_B|}},$$

where  $|X|$  denotes the number of unique elements in the set  $X$ . A similarity score of 0 indicates no shared users, while 1 corresponds to  $A$  and  $B$  having exactly the same users. This score is known as the *cosine similarity* in data mining and information retrieval (Leskovec et al., 2014; Linden et al., 2003; Salton & McGill, 1986). The cosine similarity is a powerful measure as it accounts for the stark differences in the prevalence of SNSs. Figure 1 shows that Facebook's adoption rate in our sample is above 75%, while other sites like Reddit and Tumblr are around 10%. Given this variation in the size of the user base of individual platforms, defining the association simply based on the number of users the two platforms share would be bounded by the prevalence of the smaller SNS and it would thus not be informative. Through the normalization factor (i.e., the specific division), the cosine similarity considers the size of the user bases of the individual platforms. While alternate approaches to this normalization exist, based on a wide range of applications, cosine similarity represents a sensible choice given the accuracy and technical sophistication trade-off. Note that we do not create null models, that is, expected networks of base rates of SNS use with which to compare observed associations, given (1) that there are several competing options for generating such networks (Fosdick et al., 2018; Gotelli & Graves, 1996; Lü & Zhou, 2011; Newman, 2010), (2) the stark impact of the choice of null models on the result (Schlauch et al., 2015), and (3) the observation that null-model-based approaches and conditional networks with thresholds do not necessarily yield better results than the cosine similarity (Spitz et al., 2016).

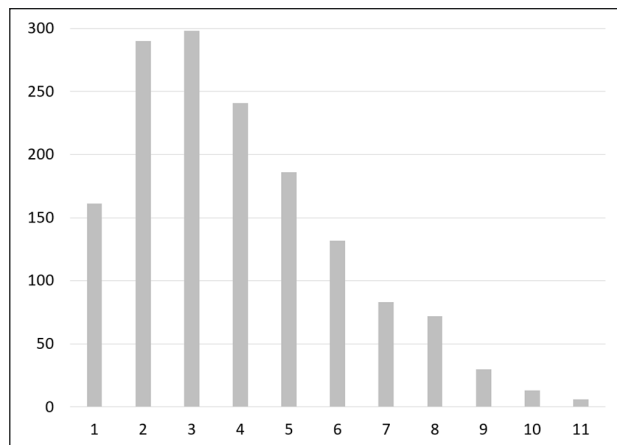
To develop an understanding of the similarity score, consider the following examples from our data set. LinkedIn and Instagram have comparable numbers of users (501 vs 517). Of these people, 245 used both platforms, with a similarity score of .48. Had there been 500 people who used both platforms, the similarity score would be nearly perfect (.98), whereas if the number of shared users had been small, say 50, then the similarity score would drop to .08. Looking next at two SNSs with very different popularity such as Facebook



**Table 1.** Sample Descriptives With Pew 2015 Data as Comparison.

	Present study				Pew 2015			
	Percent	Mean	SD	N	Percent	Mean	SD	N
Age (18–94)		48.7	16.9	1,512		48.6	17.9	1,957
Income in US\$1,000s (2.5–225)		71.5	54.4	1,512		70.2 <sup>a</sup>	50.8	1,696
Female	50.9			1,512	50.7			2,001
Employed	62.2			1,512	61.5			1,991
Rural resident	13.4			1,512	16.2			2,001
Education								
High school or less	25.9			1,512	25.0			1,981
Some college	31.6			1,512	27.6			1,981
Bachelor's or higher	42.5			1,512	47.5			1,981
Race and ethnicity								
White	71.1			1,511	76.2			1,947
Hispanic	11.7			1,511	11.1			1,982
Black	11.5			1,511	11.4			1,947
Asian	3.1			1,511	3.0			1,947
Native American	1.7			1,511	.9			1,947
Internet experiences								
Internet use years		11.1	2.8	1,512				
Internet autonomy (0–9)		4.8	2.3	1,512				
Internet use frequency (0–42)		14.7	10.8	1,491				
Internet skills (1–5)		3.4	1.1	1,512				

<sup>a</sup>Income was measured differently whereby Pew's highest category was considerably lower than NORC's.

**Figure 1.** Number of social network sites people use.

and Reddit (1,201 vs 182 users), with 159 shared users their similarity score is .34. In this case we would reach a maximal similarity score if all 182 Reddit users were also on Facebook. Yet, this maximum value would only be .39 since generally speaking, Reddit has considerably fewer users. If the two platforms had shared 18 users, the similarity would become .04. Note that 18 users in the latter case means 10% of the maximum 182 possible, just like 50 shared users means 10% of the maximum possible in the LinkedIn–Instagram case. However, the penalty was higher as indicated by the resulting similarity values .04 versus .08 due to the different user base sizes of the different platforms being compared.

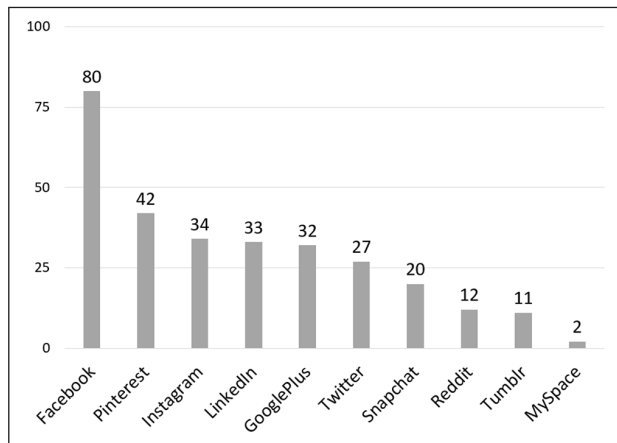
We computed the cosine similarity and created the networks using custom scripts written in the statistical programming language R. We made our full code publicly available on the GitHub repository under the link: <https://github.com/LINK-NU/SMS21-Horvat-and-Hargittai>. For network visualization, we used a freely available graph editor called yEd.<sup>1</sup> This software offers a wide variety of layout algorithms including ones based on the force directed layout paradigm, and it allows users to tune several parameters such as desired edge length, edge routing, and node labeling. To ensure that our network visualizations are not cluttered, we choose a cut-off of 10 SNS associations to show per network such that figures and tables convey information only about the strongest associations. This cut-off is not inherently motivated by the analytic method and can be freely changed based on application and visualization needs.

## SNS Use

Eleven percent of respondents do not use any social media while an additional 19% use only one. These participants do not provide any relevant data for examining SNS associations and thus are excluded from the analyses that follow. Figure 1 shows the total number of SNSs people use indicating that use of multiple platforms is common. Figure 2 displays the SNSs in order of popularity with over three-quarters of respondents using the most popular platform, Facebook, at the time of data collection and only 2% using the least popular MySpace. Given that some time has elapsed since

these data were collected, we note here changes in popularity of these sites based on data collected by the Pew Research Center (2019) as comparison. (Because their figures are based on all American adults rather than Internet users only like our study, we are only noting changes over time, not use percentages.) According to Pew, Facebook use has been stable since 2016 as has Pinterest with LinkedIn and Twitter gaining just a few percentage points. Instagram has seen the most significant gains. Pew did not measure the popularity of the remaining services until 2018 or later so comparisons are hard to show, but they all remain below LinkedIn in diffusion.

SNS associations quantified by the cosine similarity can be used to create networks in which SNSs are linked according to the similarity of their user bases. Figure 3 shows the top 10 SNS associations for the entire sample. The darker and thicker the line between two sites the higher their similarity score, that is, the higher their shared user base. A

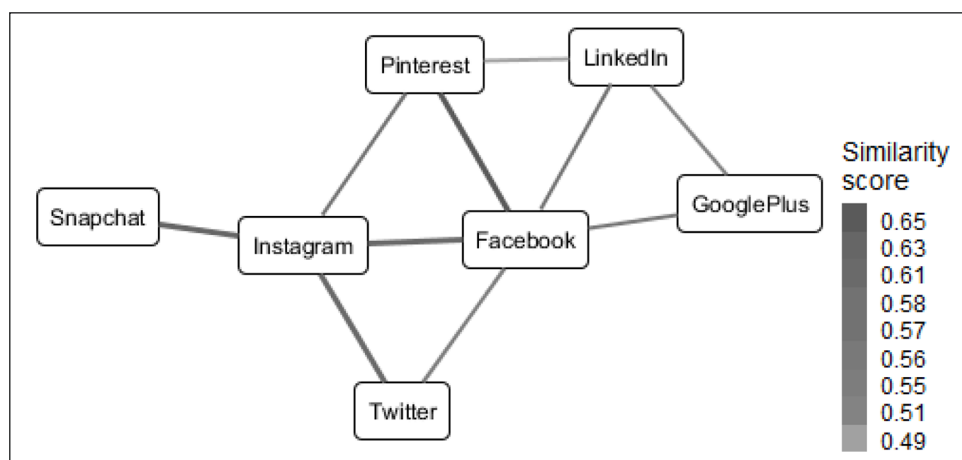


**Figure 2.** Percentage of respondents who use various social network sites.

similarity score of 1 would indicate that two platforms have the exact same users, while a score of 0 would mean that the sites do not share any users. We find that the similarity scores in our full data sample range from 0.43 to 0.65. The site pairs Facebook–Pinterest (similarity=0.65), Instagram–Snapchat (0.63), Facebook–Instagram (0.61), and Twitter–Instagram (0.61) are the strongest by association, sharing the most users. Conversely, Twitter–GooglePlus (similarity=0.43) and Tumblr–Reddit (.44) are the least similar based on their shared user base in the top 10. For a list of associations between all possible pairs of SNSs, see the table in the Appendix. This table indicates that the overall weakest associations all involve MySpace (similarities between 0.1 and 0.22). The next smallest shared user bases are registered between GooglePlus and Reddit (similarity=0.28), and Pinterest and Reddit (similarity=0.28).

This initial, full-sample analysis is helpful for identifying general patterns of pairs. However, because we know from prior literature that people of different backgrounds select into the use of social media platforms at different rates, it is also important to examine SNS pairings by user type. Next, we look at whether the association between SNSs is stronger or weaker for certain types of users. We do this by calculating the association scores by age, gender, education, and Internet skills to show to what extent different types of people use different mixes of platforms.

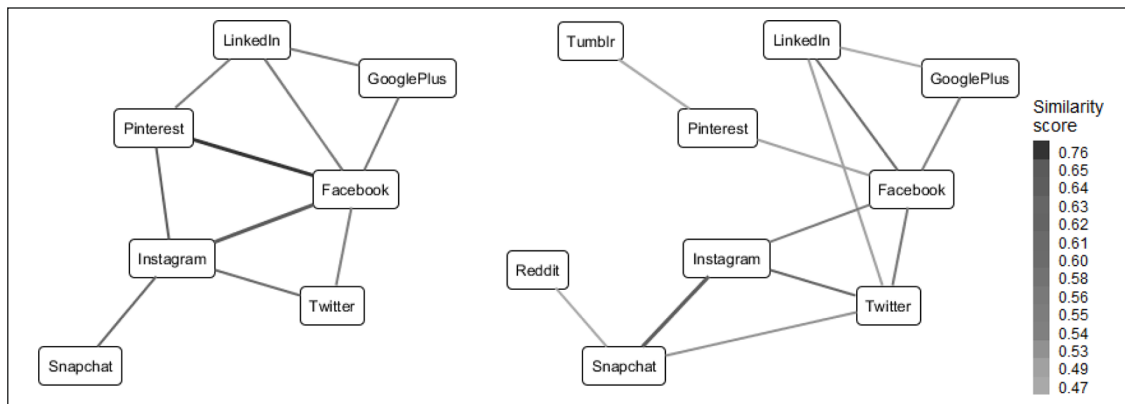
Table 2 lists the top site associations for men and women, while Figure 4 shows them graphically. We find that the user base of various SNSs very much varies by gender. While some SNS associations are similar for both men and women (e.g., Facebook–Instagram, Twitter–Instagram), others are mostly relevant for just one or the other group. For example, Reddit–Snapchat have the 12th highest association among men, while that association is not among the top 20 for women. In contrast, the highest similarity score (0.76) for women is Facebook–Pinterest, which is only in the eighth



**Figure 3.** Network of the top 10 SNS associations. SNS: social network site.

**Table 2.** SNS Similarity Scores by Gender (Female on Left, Male on Right).

SNS pair—FEMALE	Similarity score	SNS pair—MALE	Similarity score
Facebook–Pinterest	0.76	Instagram–Snapchat	0.64
Facebook–Instagram	0.65	Twitter–Instagram	0.63
Pinterest–Instagram	0.64	Facebook–LinkedIn	0.61
Instagram–Snapchat	0.62	Facebook–Twitter	0.56
Twitter–Instagram	0.6	Facebook–Instagram	0.56
Facebook–GooglePlus	0.58	Facebook–GooglePlus	0.54
Facebook–LinkedIn	0.56	Twitter–Snapchat	0.53
Pinterest–LinkedIn	0.56	Facebook–Pinterest	0.49
LinkedIn–GooglePlus	0.55	Twitter–LinkedIn	0.49
Facebook–Twitter	0.54	Tumblr–Pinterest	0.47
		LinkedIn–GooglePlus	0.47
		Reddit–Snapchat	0.47

**Figure 4.** Network of the top SNS associations by gender (female on left, male on right). SNS: social network site.

spot with a considerably lower score (0.49) for men. These numbers arise from the fact that while 86.5% of women use Facebook and 58.8% use Pinterest, 54.1% use both SNSs. At the same time, 72.1% of men use Facebook, 23.1% use Pinterest, and 19.9% use both. Accordingly, the similarity between Facebook and Pinterest in the full sample (0.65) is driven mostly by the female sub-population. This is an important nuance that the original analysis would have not revealed. Our gender-based SNS association analysis also allows identifying the pairs of platforms that could be used in a communication campaign to target the two groups while minimizing redundancy. For instance, relying on Pinterest and Snapchat would minimize the male users who are exposed to the message twice (23.1% men use Pinterest, 17% use Snapchat, only 6.9% use both, similarity is 0.35), while using Snapchat and LinkedIn would minimize the overlap in the female sample (31.2% women use LinkedIn, 23.3% use Snapchat, 8.8% use both, similarity is 0.33).

Next, we investigate SNS associations by age. When comparing the youngest (age 18–33 years) and oldest (age 62–94 years) respondents (chosen as the lowest and highest

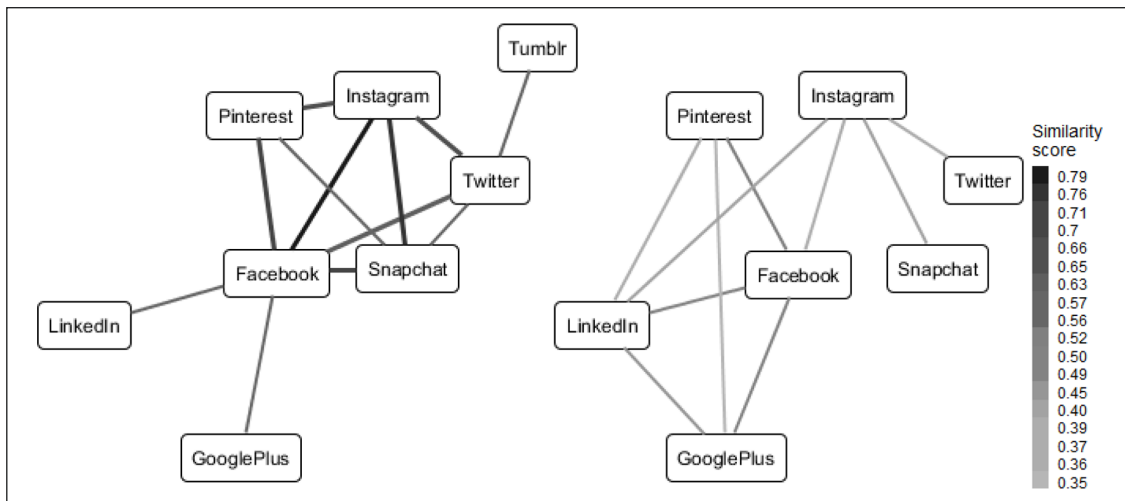
quartiles in the sample's age distribution), we find that the younger sample has systematically higher associations ranging between 0.79 and 0.56 than the older sample with associations between 0.52 and 0.35 (see Table 3). Among younger people, the most strongly associated SNS pair is Facebook–Instagram, which here is followed by the Instagram–Snapchat pair (similarity is 0.76). Among the older age group, the strongest association is between Facebook and Pinterest, a pair that is in top positions for several user types as we will see in the following. Figure 5 shows how the two groups vary in their SNS associations.

Table 4 lists the top SNS associations by level of education (also see Figure 6). Again, we observe considerable variation by user background. For example, while Facebook and LinkedIn are the most highly associated for those with a college degree or more (78.4% use Facebook, 48.2% use LinkedIn, 43.2% use both, similarity is 0.7), it is not in the top 10 associations among those with no more than a high school education (76.5% use Facebook, 14.5% use LinkedIn, 12.5% use both, similarity is 0.37). For the latter group, the highest association is for the Instagram–Snapchat pair

**Table 3.** SNS Similarity Scores by Age.

SNS pair—YOUNGER	Similarity score	SNS pair—OLDER	Similarity score
Facebook–Instagram	0.79	Facebook–Pinterest	0.52
Instagram–Snapchat	0.76	Facebook–LinkedIn	0.5
Facebook–Pinterest	0.71	Facebook–GooglePlus	0.49
Facebook–Snapchat	0.7	LinkedIn–GooglePlus	0.45
Pinterest–Instagram	0.66	Instagram–Snapchat	0.4
Twitter–Instagram	0.65	LinkedIn–Instagram	0.4
Facebook–Twitter	0.63	Twitter–Instagram	0.39
Pinterest–Snapchat	0.57	Pinterest–LinkedIn	0.37
Facebook–LinkedIn	0.57	Facebook–Instagram	0.36
Twitter–Snapchat	0.56	Pinterest–GooglePlus	0.35
Twitter–Tumblr	0.56		
Facebook–GooglePlus	0.56		

SNS: social network site.

**Figure 5.** Network of the top SNS associations by age.

SNS: social network site.

Younger sample on left and older sample on right.

(30.6% use Instagram, 21.4% use Snapchat, 16.6% use both, similarity is 0.65), which is also top 2 among those with some college education (34.6% use Instagram, 23.3% use Snapchat, 18.4% use both, similarity is 0.65), but only in sixth place for those with a college degree (36.1% use Instagram, 17.1% use Snapchat, 14.9% use both, similarity is 0.6). There is one association that is in a top position regardless of education, however, Facebook–Pinterest.

Finally, we look at how SNS associations may vary by Internet skills (see Table 5 and Figure 7). Here, we find that the Facebook–Instagram association is especially strong between users with high Internet skills (87.8% use Facebook, 54.4% use Instagram, 51.6% use both, similarity is 0.75). The associations between Facebook–LinkedIn (0.72), Facebook–Pinterest (0.7), and Twitter–Instagram (0.7) are also high and actually the top eight associations are stronger

than any associations based on the full sample. This indicates that there is a high agreement among Internet-savvy users in terms of the SNSs they use. In contrast, SNS associations between users with the lowest Internet skills are significantly weaker, that is, they are equal to or *lower* than 0.52. The latter similarity characterizes the Instagram–Snapchat pair (12.8% use Instagram, 7.4% use Facebook, and only 5.1% use both). The Facebook–Instagram pair leading the associations for users with high Internet skills is on rank 5 for users with low Internet skills (68.6% use Facebook, 12.8% use Instagram, 11.2% use both, similarity is 0.38).

## Discussion and Conclusion

Inspired by the media repertoire and media multiplexity perspectives, this article takes as its premise that most people



**Table 4.** SNS Similarity Scores by Level of Education.

SNS pair—LOW edu.	Similarity score
Instagram–Snapchat	0.65
Facebook–Pinterest	0.63
Twitter–Instagram	0.61
Facebook–Instagram	0.6
Facebook–GooglePlus	0.54
Twitter–Snapchat	0.52
Facebook–Snapchat	0.51
Facebook–Twitter	0.49
Pinterest–Instagram	0.48
LinkedIn–GooglePlus	0.48
Instagram–GooglePlus	0.48
SNS pair—MEDIUM edu.	Similarity score
Facebook–Pinterest	0.66
Instagram–Snapchat	0.65
Facebook–Instagram	0.62
Facebook–GooglePlus	0.6
Twitter–Instagram	0.59
Facebook–Twitter	0.56
Facebook–LinkedIn	0.54
Pinterest–Instagram	0.54
Tumblr–Reddit	0.52
Facebook–Snapchat	0.51
SNS pair—HIGH edu.	Similarity score
Facebook–LinkedIn	0.7
Facebook–Pinterest	0.66
Pinterest–Instagram	0.64
Twitter–Instagram	0.63
Facebook–Instagram	0.61
Instagram–Snapchat	0.6
Facebook–Twitter	0.57
Pinterest–LinkedIn	0.56
Facebook–GooglePlus	0.55
Twitter–LinkedIn	0.55
LinkedIn–GooglePlus	0.55

SNS: social network site.

use a combination of social media, and thus research looking at their adoption should do the same. While previous research has shown that different people select into the use of various SNSs at different rates, such work has only looked at the adoption of sites in isolation of considering adoption of other such sites (e.g., Blank & Lutz, 2017; Hargittai, 2020; Hargittai & Litt, 2011). Knowing that certain types of people are more or less likely to use various social media does not allow us to determine whether this selection also applies to SNS pairs. This article contributes to the literatures on media multiplexity and on SNS use by showing the shared user bases of various social media platforms by user attributes.

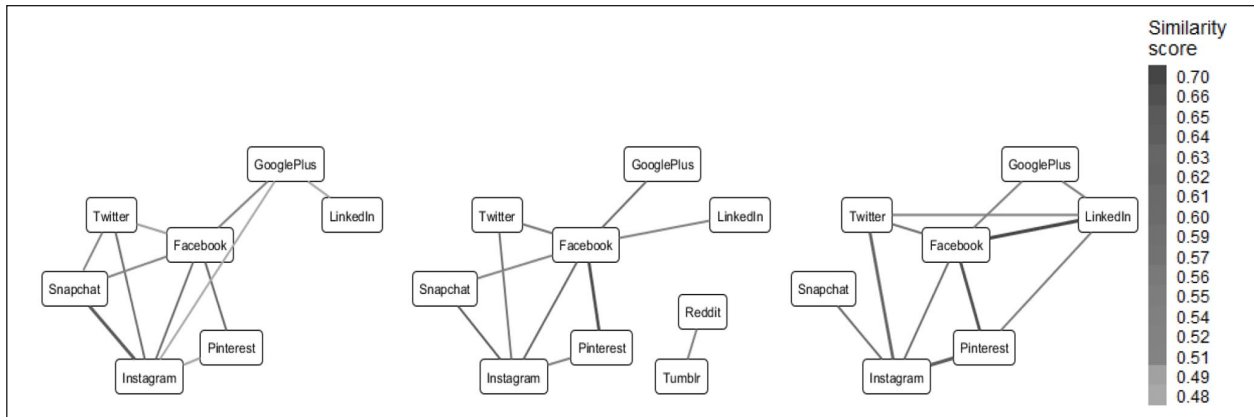
The networks mapped out in this study reflect how dissimilar SNSs are in terms of appeal and affordances to the

same pairs of users. Our research highlights the fact that more platforms do not necessarily mean more diverse voices represented on platforms if they are being used by similar people. The findings heighten potential concerns associated with questions of social inclusion, segregation and empowerment through social media (e.g., Costanza-Chock, 2020; Jackson et al., 2020; Tufekci, 2017), because they show that not using one platform can be non-trivially connected to not using a suite of other platforms.

This study thus adds a new dimension of applying social network analysis to media use scholarship about media repertoires and the multiplexity of people's media uses by presenting connections between social media platforms based on their usage patterns. While prior work has done this in the realm of news and journalism (e.g., Mukerjee et al., 2018; Taneja & Webster, 2016), we know of no other work to have done so for SNS associations. Our networks map out the structural underpinnings of SNS use. Moreover, we investigate and find striking differences in the associations between SNSs based on user attributes. More broadly, our work emphasizes how individual SNS uses are compounding to system-level SNS networks with clusters that end up being more or less accessible for people with different backgrounds.

In addition to the theoretical implications of the findings, there are also practical ones. For example, if an organization wishes to reach people across the population such as at the time of a crisis or for political campaigning purposes, it may think that targeting different social media platforms will meet its needs. However, as we show, which particular mix of social media it targets is crucial to reaching its goals. For example, basing such outreach on Twitter and Snapchat would oversample men without guaranteeing to reach women just as going for a mix of Facebook and Pinterest would leave out a considerable portion of men. Rather, our results indicate that the communication strategy would be more successful by choosing only one of the SNSs from the highly similar pairs to save resources and avoid disseminating information to the same people multiple times. Extending prior work, our results also suggest that not including multiple SNSs in information campaigns might lead to a large-scale and systematic exclusion of certain demographic groups which in turn can increase harmful information gaps (Gans, 2020). It is essential to target users efficiently through multiple SNSs and our results inform such efforts by highlighting which SNS combinations will increase coverage of different demographic groups the most.

Our findings also have important implications for data-collection projects that rely on SNSs as their sampling frames. Prior work has already noted that basing a study's sampling frame on only one platform poses major limitations as the data will bias against populations that are systematically less likely to be on said platforms (Blank & Lutz, 2017; Hargittai, 2015; Tufekci, 2014). What our study shows is that simply casting a wider net to include more than one platform



**Figure 6.** Network of the top SNS associations by education.

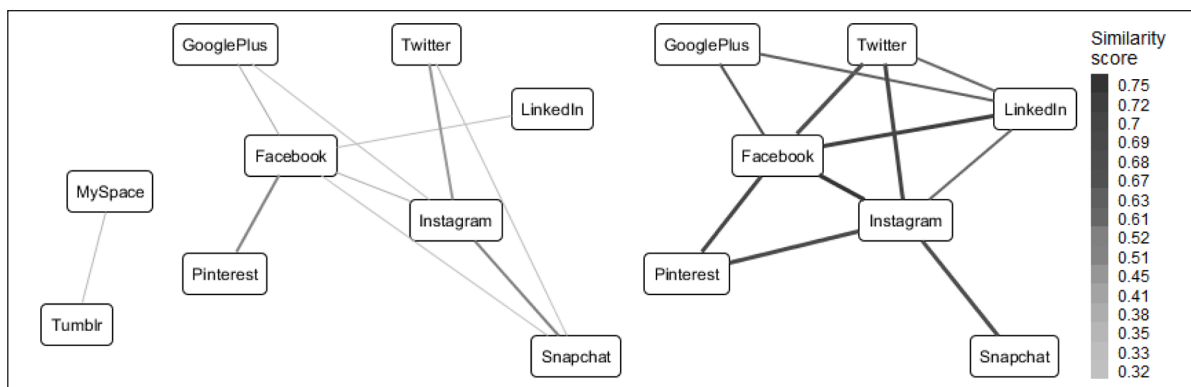
SNS: social network site.

Low education on left, medium education in middle, and high education on right.

**Table 5.** SNS Similarity Scores by Internet Skills.

SNS pair—LOW skills	Similarity score	SNS pair—HIGH skills	Similarity score
Instagram–Snapchat	0.52	Facebook–Instagram	0.75
Facebook–Pinterest	0.51	Facebook–LinkedIn	0.72
Twitter–Instagram	0.45	Facebook–Pinterest	0.7
Facebook–GooglePlus	0.41	Twitter–Instagram	0.7
Facebook–Instagram	0.38	Pinterest–Instagram	0.69
MySpace–Tumblr	0.35	Facebook–Twitter	0.68
Facebook–LinkedIn	0.33	Instagram–Snapchat	0.68
Twitter–Snapchat	0.33	Facebook–GooglePlus	0.67
Facebook–Snapchat	0.32	LinkedIn–GooglePlus	0.63
GooglePlus–Instagram	0.32	Twitter–LinkedIn	0.61
		LinkedIn–Instagram	0.61

SNS: social network site.



**Figure 7.** Network of the top SNS associations by Internet skills.

SNS: social network site.

Low skills on left and high skills on right.

may not in and of itself address the sampling biases of individual platforms. If projects target sites that have high associations for particular users then they will likely reach the

same people rather than a diverse set. Our findings suggest that not only are individual SNSs biased in whose voices they represented, so are SNS pairs.

Future work could extend the used method to larger clusters of SNSs than dyads. A generalized approach focusing on clusters of three or more SNSs would allow communication campaigns to find their audiences more flexibly depending on the size of their effort (i.e., the number of SNSs whose users they can target) and the expected overlaps between those sites' user bases.

While the article makes important novel contributions to the literature on social media adoption and optimal communication strategy using such channels, it also has limitations. Our measure of social media use is a basic one simply gauging whether someone visits a platform at least sometimes or not. What people do on such sites, how they approach the content they see, and how it fits into their broader information-seeking repertoires are factors we are not able to consider. In addition, people's use of various platforms fluctuates over time, and so findings about specific sites and their associations with others may be different as people abandon some sites and join others (Lazer et al., 2021). Nonetheless, the general point about platform associations holds regardless of specific usage levels.

With these limitations in mind, the most important contribution of this article is the proof of concept regarding bias by SNS associations not which specific associations may be more or less popular across sample groupings at any particular time. Theoretically, the findings support the importance of the media repertoire and media multiplexity perspectives showing that there is important insight to be gained from focusing on more than one platform when studying people's communication practices. Given that we show considerable variation in SNS dyads across population groups, it is important that future research with up-to-date platform popularity also investigates such divergences so as to avoid the biases that can stem from relying on sites that cater to the same user groups.

### Authors' Note

A previous version of this article was presented at the 2019 International Communication Association annual meetings in Prague.

### Acknowledgements

The authors acknowledge the support of Merck (known as MSD outside the United States and Canada) and the Robert and Kaye Hiatt Fund at Northwestern University. They thank Aaron Shaw for collaborating on the original data collection.

### Declaration of Conflicting Interests



The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

### Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This

work has been partially funded by NSF Faculty Early Career Development Program Grant IIS-1943506.

### ORCID iDs

Emőke-Ágnes Horvát  <https://orcid.org/0000-0001-7709-1172>  
Eszter Hargittai  <https://orcid.org/0000-0003-4199-4868>

### Note

1. <https://www.yworks.com/products/yed>

### References

- Bakshy, E., Hofman, J. M., Mason, W. A., & Watts, D. J. (2011). Everyone's an influencer: Quantifying influence on twitter. *Proceedings of the Fourth ACM International Conference on Web Search and Data Mining*, 65–74. <https://doi.org/10.1145/1935826.1935845>
- Bakshy, E., Messing, S., & Adamic, L. A. (2015). Exposure to ideologically diverse news and opinion on Facebook. *Science*, 348(6239), 1130–1132. <https://doi.org/10.1126/science.aal1160>
- Barry, J. (2009). Pandemics: Avoiding the mistakes of 1918. *Nature*, 459, 324–325. <https://doi.org/10.1038/459324a>
- Blank, G. (2017). The digital divide among Twitter users and its implications for social research. *Social Science Computer Review*, 35, 679–697. <https://doi.org/10.1177/0894439316671698>
- Blank, G., & Lutz, C. (2017). Representativeness of social media in Great Britain: Investigating Facebook, LinkedIn, Twitter, Pinterest, Google+, and Instagram. *American Behavioral Scientist*, 61(7), 741–756. <https://doi.org/10.1177/0002764217717559>
- boyd, danah. (2011). White flight in networked publics: How race and class shaped American teen engagement with MySpace and Facebook. In L. Nakamura & P. A. Chow-White (Eds.), *Race after the internet* (pp. 203–222). Routledge.
- Budak, C., & Watts, D. J. (2015). Dissecting the Spirit of Gezi: Influence vs. Selection in the Occupy Gezi Movement. *Sociological Science*, 2, 370–397. <https://doi.org/10.15195/v2.a18>
- Conover, M., Ratkiewicz, J., Francisco, M. R., Gonçalves, B., Menczer, F., & Flammini, A. (2011). Political polarization on Twitter. *Proceedings of the Fifth International Conference on Weblogs and Social Media*, 133, 89–96. <https://ojs.aaai.org/index.php/ICWSM/article/view/14126/13975>
- Costanza-Chock, S. (2020). *Design justice: Community-led practices to build the worlds we need*. The MIT Press. <https://mitpress.mit.edu/books/design-justice>
- Deshpande, M., & Karypis, G. (2004). Item-based top-N recommendation algorithms. *ACM Transactions on Information Systems*, 22(1), 143–177.
- Dvir-Gvirsman, S. (2020). Understanding news engagement on social media: A media repertoire approach. *New Media & Society*. <https://doi.org/10.1177/1461444820961349>
- Fosdick, B. K., Larremore, D. B., Nishimura, J., & Ugander, J. (2018). Configuring random graph models with fixed degree sequences. *SIAM Review*, 60(2), 315–355. <https://doi.org/10.1137/16M1087175>
- Gans, J. (2020). *The pandemic information gap*. The MIT Press. <https://mitpress.mit.edu/books/pandemic-information-gap>

- Gazit, T., Aharoni, N., & Amichai-Hamburger, Y. (2019). Tell me who you are and I will tell you which SNS you use: SNSs participation. *Online Information Review*, 44(1), 139–161. <https://doi.org/10.1108/OIR-03-2019-0076>
- González-Bailón, S., & Wang, N. (2016). Networked discontent: The anatomy of protest campaigns in social media. *Social Networks*, 44, 95–104. <https://doi.org/10.1016/j.socnet.2015.07.003>
- Gotelli, N. J., & Graves, G. R. (1996). *Null models in ecology*. Smithsonian Institution Press.
- Haight, M., Quan-Haase, A., & Corbett, B. A. (2014). Revisiting the digital divide in Canada: The impact of demographic factors on access to the internet, level of online activity, and social networking site usage. *Information, Communication & Society*, 17(4), 503–519. <https://doi.org/10.1080/1369118X.2014.891633>
- Hargittai, E. (2007). Whose space? Differences among users and non-users of social network sites. *Journal of Computer-Mediated Communication*, 13(1), 276–297. <https://doi.org/10.1111/j.1083-6101.2007.00396.x>
- Hargittai, E. (2015). Is bigger always better? Potential biases of big data derived from social network sites. *The ANNALS of the American Academy of Political and Social Science*, 659(1), 63–76. <https://doi.org/10.1177/0002716215570866>
- Hargittai, E. (2020). Potential biases in big data: Omitted voices on social media. *Social Science Computer Review*, 38(1), 10–24. <https://doi.org/10.1177/0894439318788322>
- Hargittai, E., & Hsieh, Y. P. (2010). From dabblers to omnivores: A typology of social network site usage. In Z. Papacharissi (Ed.), *A networked self: Identity, community, and culture on social network sites* (pp. 146–168). Routledge.
- Hargittai, E., & Hsieh, Y. P. (2012). Succinct survey measures of Web-use skills. *Social Science Computer Review*, 30(1), 95–107. <https://doi.org/10.1177/0894439310397146>
- Hargittai, E., & Litt, E. (2011). The tweet smell of celebrity success: Explaining variation in twitter adoption among a diverse group of young adults. *New Media & Society*, 13(5), 824–842. <https://doi.org/10.1177/1461444811405805>
- Haythornthwaite, C. (2001). Exploring multiplexity: Social network structures in a computer-supported distance learning class. *Information Society*, 17(3), 211–226. <https://doi.org/10.1080/01972240152493065>
- Heeter, C. (1985). Program selection with abundance of choice. *Human Communication Research*, 12(1), 126–152. <https://doi.org/10.1111/j.1468-2958.1985.tb00070.x>
- Helleman, J., Willems, K., & Brengman, M. (2020). Daily active users of social network sites: Facebook, Twitter, and Instagram-use compared to general social network site use. In F. J. Martínez-López & S. D'Alessandro (Eds.), *Advances in digital marketing and eCommerce* (pp. 194–202). Springer. [https://doi.org/10.1007/978-3-030-47595-6\\_24](https://doi.org/10.1007/978-3-030-47595-6_24)
- Hemphill, L., & Schöpke-Gonzalez, A. M. (2020). Two computational models for analyzing political attention in social media. *Proceedings of the International AAAI Conference on Web and Social Media*, 14, 260–271. <https://ojs.aaai.org/index.php/ICWSM/article/view/7297>
- Hsieh, Y. P. (2012). Online social networking skills: The social affordances approach to digital inequality. *First Monday*, 17(4). <https://doi.org/10.5210/fm.v17i4.3893>
- Huang, Y. L., Starbird, K., Orand, M., Stanek, S. A., & Pedersen, H. T. (2015). Connected through crisis: Emotional proximity and the spread of misinformation online. In *Proceedings of the 18Th ACM Conference on Computer Supported Cooperative Work & Social Computing* (pp. 969–980). <https://doi.org/10.1145/2675133.2675202>
- Jackson, S. J., Bailey, M., & Welles, B. F. (2020). *#HashtagActivism*. The MIT Press.
- Jiang, S., Metzger, M., Flanagin, A., & Wilson, C. (2020). Modeling and measuring expressed (dis)belief in (mis)information. *Proceedings of the International AAAI Conference on Web and Social Media*, 14, 315–326. <https://ojs.aaai.org/index.php/ICWSM/article/view/7302>
- Kim, S. J. (2016). A repertoire approach to cross-platform media use behavior. *New Media & Society*, 18(3), 353–372. <https://doi.org/10.1177/1461444814543162>
- Klinger, U., & Svensson, J. (2015). The emergence of network media logic in political communication: A theoretical approach. *New Media & Society*, 17(8), 1241–1257. <https://doi.org/10.1177/1461444814522952>
- Lazer, D., Hargittai, E., Freelon, D., Gonzalez-Bailon, S., Munger, K., Ognyanova, K., & Radford, J. (2021). Meaningful measures of human society in the twenty-first century. *Nature*, 595, 189–196. <https://doi.org/10.1038/s41586-021-03660-7>
- Leskovec, J., Rajaramand, A., & Ullman, J. D. (2014). *Mining of massive data sets: Second edition*. Cambridge University Press.
- Liang, T.-P., & Turban, E. (2011). Introduction to the special issue social commerce: A research framework for social commerce. *International Journal of Electronic Commerce*, 16(2), 5–14. <https://doi.org/10.2753/JEC1086-4415160201>
- Linden, G., Smith, B., & York, J. (2003). Amazon.com recommendations: Item-to-item collaborative filtering. *IEEE Internet Computing*, 7(1), 76–80. <https://doi.org/10.1109/MIC.2003.1167344>
- Lü, L., & Zhou, T. (2011). Link prediction in complex networks: A survey. *Physica A: Statistical Mechanics and Its Applications*, 390(6), 1150–1170. <https://doi.org/10.1016/j.physa.2010.11.027>
- Mellon, J., & Prosser, C. (2017). Twitter and Facebook are not representative of the general population: Political attitudes and demographics of British social media users. *Research & Politics*, 4(3). <https://doi.org/10.1177/2053168017720008>
- Mukerjee, S., Majó-Vázquez, S., & González-Bailón, S. (2018). Networks of audience overlap in the consumption of digital news. *Journal of Communication*, 68(1), 26–50. <https://doi.org/10.1093/joc/jqx007>
- Neiger, B. L., Thackeray, R., Wagenen, S. A. V., Hanson, C. L., West, J. H., Barnes, M. D., & Fagen, M. C. (2012). Use of social media in health promotion: purposes, key performance indicators, and evaluation metrics. *Health Promotion Practice*, 13(2), 159–164. <https://doi.org/10.1177/1524839911433467>
- Newman, M. E. J. (2010). *Networks: An introduction*. Oxford University Press.
- NORC. (n.d.). *AmeriSpeak: NORC's breakthrough panel-based research platform*. <https://perma.cc/2TBM-KLUJ>
- Olsen, R. K. (2020). Connecting people? Understanding media's role as democratic resources for people in digitally advanced local environments. *Digital Journalism*, 8(4), 506–525. <https://doi.org/10.1080/21670811.2019.1679029>
- Park, D. H., Kim, H. K., Choi, I. Y., & Kim, J. K. (2012). A literature review and classification of recommender systems



- research. *Expert Systems with Applications*, 39(11), 10059–10072. <https://doi.org/10.1016/j.eswa.2012.02.038>
- Paul, M. J., & Dredze, M. (2021). You are what you Tweet: Analyzing Twitter for public health. *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media*, 5(1), 265–272. <https://ojs.aaai.org/index.php/ICWSM/article/view/14137/13986>
- Pew Research Center. (2019). *Social media fact sheet*. <https://www.pewresearch.org/internet/fact-sheet/social-media/>
- Phalen, P. F., & Ducey, R. V. (2012). Audience behavior in the multi-screen “video-verse.” *International Journal on Media Management*, 14(2), 141–156. <https://doi.org/10.1080/14241277.2012.657811>
- Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P., & Riedl, J. (1994). GroupLens: An open architecture for collaborative filtering of netnews. In *Proceedings of the 1994 ACM Conference on Computer Supported Cooperative Work* (pp. 175–186). <https://doi.org/10.1145/192844.192905>
- Romero, D. M., Meeder, B., & Kleinberg, J. (2011). Differences in the mechanics of information diffusion across topics: Idioms, political hashtags, and complex contagion on Twitter. In *Proceedings of the 20th International Conference on World Wide Web* (pp. 695–704). <https://doi.org/10.1145/1963405.1963503>
- Romero, D. M., Tan, C., & Ugander, J. (2013). On the interplay between social and topical structure. In *Proceedings of ICWSM* (pp. 516–525). <https://arxiv.org/abs/1112.1115>
- Salton, G., & McGill, M. J. (1986). *Introduction to modern information retrieval*. McGraw Hill.
- Sarker, A., Lakamana, S., Hogg-Bremer, W., Xie, A., Al-Garadi, M. A., & Yang, Y.-C. (2020). Self-reported COVID-19 symptoms on Twitter: An analysis and a research resource. *Journal of the American Medical Informatics Association*, 27(8), 1310–1315. <https://doi.org/10.1093/jamia/ocaa116>
- Sarwar, B., Karypis, G., Konstan, J., & Riedl, J. (2001). Item-based collaborative filtering recommendation algorithms. *Proceedings of the International World Wide Web Conference*, Association for Computing Machinery, 285–295. <https://doi.org/10.1145/371920.372071>
- Schlauch, W. E., Horvát, E. Á., & Zweig, K. A. (2015). Different flavors of randomness: Comparing random graph models with fixed degree sequences. *Social Network Analysis and Mining*, 5(36). <https://doi.org/10.1007/s13278-015-0267-z>
- Sheldon, P., & Bryant, K. (2016). Instagram: Motives for its use and relationship to narcissism and contextual age. *Computers in Human Behavior*, 58(Suppl. C), 89–97. <https://doi.org/10.1016/j.chb.2015.12.059>
- Spitz, A., Gimmler, A., Stoeck, T., Zweig, K. A., & Horvát, E.-Á. (2016). Assessing low-intensity relationships in complex networks. *PLOS ONE*, 11(4), Article e0152536. <https://doi.org/10.1371/journal.pone.0152536>
- Starbird, K. (2017). Examining the alternative media ecosystem through the production of alternative narratives of mass shooting events on Twitter. *The 11th International AAAI Conference On Web And Social Media*, 11(1), 230–239. <https://ojs.aaai.org/index.php/ICWSM/article/view/14878>
- Sutton, J., Gibson, C. B., Phillips, N. E., Spiro, E. S., League, C., Johnson, B., Fitzhugh, S. M., & Butts, C. T. (2015). A cross-hazard analysis of terse message retransmission on Twitter. *Proceedings of the National Academy of Sciences*, 112(48), 14793–14798. <https://doi.org/10.1073/pnas.1508916112>
- Taneja, H. (2017). Mapping an audience-centric World Wide Web: A departure from hyperlink analysis. *New Media & Society*, 19(9), 1331–1348. <https://doi.org/10.1177/1461444816642172>
- Taneja, H., & Webster, J. G. (2016). How do global audiences take shape? The role of institutions and culture in patterns of web use. *Journal of Communication*, 66(1), 161–182. <https://doi.org/10.1111/jcom.12200>
- Taneja, H., Webster, J. G., Malthouse, E. C., & Ksiazek, T. B. (2012). Media consumption across platforms: Identifying user-defined repertoires. *New Media & Society*, 14(6), 951–968.
- Thackeray, R., Neiger, B. L., Smith, A. K., & Van Wagenen, S. B. (2012). Adoption and use of social media among public health departments. *BMC Public Health*, 12(1), Article 242. <https://doi.org/10.1186/1471-2458-12-242>
- Tufekci, Z. (2014). Big questions for social media big data: Representativeness, validity and other methodological pitfalls. In *ICWSM '14: Proceedings of the 8th International AAAI Conference on Weblogs and Social Media* (pp. 505–514). <http://arxiv.org/abs/1403.7400>
- Tufekci, Z. (2017). *Twitter and tear gas* —Yale University Press. Yale University Press. <https://yalebooks.yale.edu/book/9780300259292/twitter-and-tear-gas>
- Ugander, J., Backstrom, L., Marlow, C., & Kleinberg, J. (2012). Structural diversity in social contagion. *Proceedings of the National Academy of Sciences*, 109(16), 5962–5966. <https://doi.org/10.1073/pnas.1116502109>
- van Dijck, J. (2013). *The culture of connectivity: A critical history of social media*. Oxford University Press.
- Vieweg, S., Hughes, A. L., Starbird, K., & Palen, L. (2010). Microblogging during two natural hazards events: What twitter may contribute to situational awareness. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1079–1088). <https://doi.org/10.1145/1753326.1753486>
- Weng, L., Menczer, F., & Ahn, Y.-Y. (2013). Virality prediction and community structure in social networks. *Scientific Reports*, 3(1), Article 2522. <https://doi.org/10.1038/srep02522>
- World Health Organization. (2017). *WHO strategic communications framework for effective communication*. <https://www.who.int/mediacentre/communication-framework.pdf>
- Wu, A. X., Taneja, H., & Webster, J. G. (2020). Going with the flow: Nudging attention online. *New Media & Society*, 23(10), 2979–2998. <https://doi.org/10.1177/1461444820941183>
- Yuan, E. (2011). News consumption across multiple media platforms. *Information, Communication & Society*, 14(7), 998–1016. <https://doi.org/10.1080/1369118X.2010.549235>

## Author Biographies

Emőke-Ágnes Horvát is an Assistant Professor in the Communication Studies Department of the School of Communication and (by courtesy) the Computer Science Department in the McCormick School of Engineering at Northwestern University. She leads LINK (Lab on Innovation, Networks and Knowledge), a research group that develops network and data science approaches to understand and support collective intelligence and scholarly communication online. Her current research is funded by the US National Science Foundation through a CAREER award.

Eszter Hargittai is a Professor and holds the Chair in Internet Use & Society in the Department of Communication and Media Research at the University of Zurich where she heads the Web

Use Project research group. Her work focuses on the social and policy implications of digital media with a particular interest in how differences in people's Internet skills/digital literacy influence what they do online. Her work has received awards from the International Communication Association (ICA), the National Communication Association, the American Sociological

Association, the Eastern Sociological Society, and the Telecommunications Policy Research Conference. In 2010, ICA selected her to receive its Outstanding Young Scholar Award, and in 2019, ICA elected her as a Fellow. She is editor of *Research Exposed: How Empirical Social Science Gets Done in the Digital Age* (Columbia University Press, 2021).

**Appendix.** SNS Similarity Scores for the Full Sample Where User Background Is Not Considered.

SNS pair	Similarity score
Facebook–Pinterest	0.65
Instagram–Snapchat	0.63
Facebook–Instagram	0.61
Twitter–Instagram	0.61
Facebook–LinkedIn	0.58
Pinterest–Instagram	0.57
Facebook–GooglePlus	0.56
Facebook–Twitter	0.55
LinkedIn–GooglePlus	0.51
Pinterest–LinkedIn	0.49
Facebook–Snapchat	0.48
Twitter–LinkedIn	0.48
LinkedIn–Instagram	0.48
Twitter–Pinterest	0.47
GooglePlus–Instagram	0.47
Twitter–Tumblr	0.46
Twitter–Snapchat	0.46
Pinterest–GooglePlus	0.46
Tumblr–Reddit	0.44
Twitter–GooglePlus	0.43
Tumblr–Instagram	0.42
Pinterest–Snapchat	0.42
Twitter–Reddit	0.39
Tumblr–Pinterest	0.37
Tumblr–Snapchat	0.37
GooglePlus–Snapchat	0.36
Instagram–Reddit	0.36
Reddit–Snapchat	0.36
Facebook–Tumblr	0.34
Facebook–Reddit	0.34
LinkedIn–Snapchat	0.34
Tumblr–LinkedIn	0.33
LinkedIn–Reddit	0.33
Tumblr–GooglePlus	0.29
Pinterest–Reddit	0.28
GooglePlus–Reddit	0.28
MySpace–GooglePlus	0.22
MySpace–LinkedIn	0.21
MySpace–Twitter	0.2
MySpace–Pinterest	0.2
MySpace–Instagram	0.19
Facebook–MySpace	0.16
MySpace–Snapchat	0.15
MySpace–Tumblr	0.14
MySpace–Reddit	0.1

SNS: social network site.