ISSN: 2473-4901
v9 n5949

Social Media Only Has Two Clusters: A United States Analysis

Alan Peslak
arp14@psu.edu
Information Sciences and Technology
Penn State University
Dunmore, PA 18512 USA

Pratibha Menon
menon@pennwest.edu
Computing and Engineering Technology

Lisa Kovalchick
kovalchick@pennwest.edu
Computing and Engineering Technology

Pennsylvania Western University
California, PA 15419 USA

Abstract

The expansion of social media and networking has been remarkable. Since its inception in 1995 with Classmates.com, the landscape evolved to include Friendster in 2002, LinkedIn and MySpace in 2003, and Facebook in 2004. Today, social networking is a global phenomenon, with Facebook boasting nearly 2.95 billion active users worldwide (Statista, 2023a). The number of significant social media platforms has also increased, with the top sites in the United States accounting for most of the activity. This study explores a 2021 Pew Internet dataset through Two-Step Cluster Analysis to identify Social Networking User Groups. By combining usage data from top social media websites with pertinent demographic and sociographic information, we establish two distinct user clusters for social media in the US as of 2021. The implications for marketers, researchers, and society at large are also considered.

Keywords: Social networking, social media, cluster analysis, Facebook, YouTube, TikTok

1. INTRODUCTION

Social networking involves utilizing internet-based platforms to engage with other users and establish new connections with individuals who share similar interests. Since the mid-1990s, the number and popularity of social networking platforms have experienced significant growth. Figure 1 illustrates the increasing monthly usage of these applications. In the United States, at least 72% of adults utilize some of the social media platforms (Auxier & Anderson, 2021). Prominent social networking sites and applications include Facebook, Instagram, Twitter, Snapchat, TikTok, and YouTube.

Users of these platforms may exhibit comparable traits. Identifying clusters of similar social networking users can benefit various audiences. To pinpoint social media networking groups, we initiate a literature review that explores social media usage in the US across multiple categories,

ISSN: 2473-4901
v9 n5949

such as age and gender, education level, and income level. We also offer a summary of cluster analysis, the technique employed to identify the groupings. Following this, we outline the methodology applied in our study and the predictor importance. We identify two distinct clusters of social media users in the US. In the discussion and conclusions section, we examine the implications of our discoveries and propose ideas for future research.

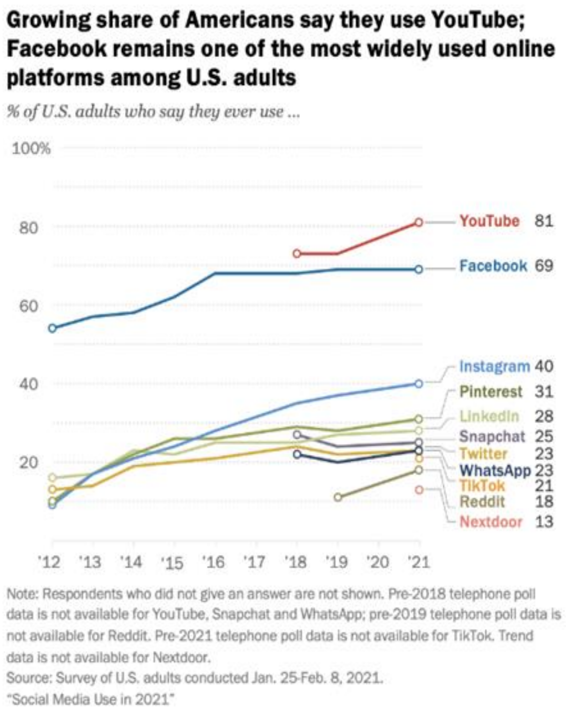

Figure 1: Increasing Monthly Use of Popular Social Networking Platforms

2. LITERATURE REVIEW

Many authors have studied social media usage and employed various techniques to categorize social media users. Java, Song, Finin, and Tseng (2007) analyzed Twitter users and their connections to understand the nature of microblogging communities and their communication patterns; they identified four types of user intentions in these communities: daily chatter, conversations, information sharing, and news reporting. In addition, they categorized Twitter users into three main categories: information sources, friends, and information seekers (Java et al., 2007). Gjoka, Kurant, Butts, and Markopoulou (2010) utilized sampling techniques to understand the structure and properties of online social networks, specifically using Facebook as a case study; they proposed a new sampling methodology that allowed them to identify and study unbiased samples of Facebook's user network. Riquelme and González-Cantergiani (2016) performed the first comprehensive study of measures used to identify the most influential Twitter users.

Researchers have utilized cluster analysis to develop new techniques, methods, and algorithms to study the vast number of social media users. Agarwal and Liu (2009) provide an overview of various research techniques for analyzing and mining the blogosphere; their book discusses topics such as blog data collection, preprocessing, analysis, and modeling, including social network analysis, to identify clusters and communities within the blogosphere. Catanese, Meo, Ferrara, Fiumara, and Provetti (2011) presented a methodology for crawling Facebook to perform social network analysis; they demonstrated how their methodology could be used to identify clusters and communities within Facebook, providing insights into the structure and dynamics of the network. McAuley and Leskovec (2012) developed a machine learning model to discover social circles in ego networks (i.e., networks centered around an individual user); their model was tested on various social media platforms, including Facebook, Google+, and Twitter, and demonstrated strong performance in identifying clusters of users with shared interests. Raghavan, Albert, and Kumara (2007) presented a near-linear time algorithm for detecting community structures in large-scale networks, including social media platforms; their proposed algorithm was tested on synthetic and real-world networks, showing its efficiency and scalability for analyzing social media clusters. Backstrom and Leskovec (2011) proposed a supervised random walk algorithm for predicting and recommending links in social networks; their algorithm was applied to various social media platforms, including Facebook, and showed strong performance in identifying potential connections between users based on their existing social media clusters. Zafarani and Liu (2009) utilized a user's behavior patterns to identify users across various social media platforms. Their technique could improve user experience, including verifying user identity across multiple social media platforms; researchers could also use it when studying user behavior across platforms.

Others have used cluster analysis to study user behavior on social media. Xu, Zhang, Wu, and Yang (2012) analyzed user posting behavior on Twitter; their work assumes user behavior is usually influenced by the following three factors:

ISSN: 2473-4901
v9 n5949

breaking news, friends' posts, and the user's interests. They proposed a mixture latent topic model to predict a user's motivation to create and share content on Twitter (Xu et al., 2012). Naveed, Gottron, Kunegis, and Alhadi (2011) studied tweets and retweets on Twitter and trained a prediction model to forecast the likelihood of a Tweet being retweeted; they discovered that Tweets on general topics are more likely to be retweeted than Tweets concerning very specific interests and content. Rizoiu, Xie, Sanner, Cebrian, Yu, and Van Hentenryck (2017, p. 735) studied videos on Twitter and developed a mathematical model using the Hawkes intensity process to "explain the complex popularity history of each video according to its type, content, network of diffusion, and sensitivity to promotion." The authors used this model to predict the likelihood of a video going viral and those with little likelihood of going viral, regardless of promotion.

3. METHODOLOGY

Data

Our analysis used data from Pew Research and was obtained from phone interviews conducted from January 25 to February 8, 2021, with a nationwide sample of 1,502 adults aged 18 or older residing in all 50 U.S. states and the District of Columbia. Abt Associates directed the interviewers who conducted the interviews with 300 respondents on landline phones and 1,202 on cellphones, including 845 without landlines. The survey employed a mix of landline and cellphone random-digit-dial samples provided by Dynata per Abt Associates' specifications. Interviews were in English and Spanish (Methodology, 2021). More details about the survey methodology can be found from the U. S. Survey published by Pew Research Center (U.S. Surveys, 2021).

For the landline sample, the youngest adult male or female present at home was randomly selected. In the cell sample, interviews were conducted with the adult (18 years or older) who answered the phone. The combined landline and cellphone samples were weighted using an iterative method, aligning gender, age, education, race, Hispanic origin, nativity, region, and population density with parameters from the U.S. Census Bureau's 2019 American Community Survey one-year estimates and the decennial census. The sample was also weighted to match current telephone usage patterns (landline only, cellphone only, or both) based on extrapolations from the 2019 National Health Interview Survey.

Cluster Analysis

Clustering refers to assembling similar data points into smaller subgroups within a broader dataset. Ideally, these clusters should consist of homogeneous elements that share more similarities with members within the same cluster than with those in different clusters. Clustering, or cluster analysis, is an unsupervised machine learning technique to detect inherent groupings in data (Wilson, 2020). It interprets the input data and identifies natural clusters or groups based on feature similarity.

In this study, we employed the silhouette method to create distinct clusters of social media users. The silhouette method, introduced by Kaufman and Rousseuw (1990), is a standard tool for validating data clusters and determining the optimal number of clusters. This method gauges a data point's similarity to its own cluster (cohesion) versus other clusters (separation), thereby assessing the quality of its placement within the cluster. Silhouette coefficients range between -1 and 1, with higher values denoting better clustering. A value close to 1 signifies that the data point is far from adjacent clusters, whereas a value near 0 means the data point is close to or between two clusters, without a clear preference for either. A negative silhouette value may suggest incorrect cluster assignment.

The number of clusters that yield the highest average silhouette value represents the optimal cluster number. To compute the silhouette score for each data point, i, the formula is $s(i) = (b(i)-- a(i)) / \max(b(i), a(i))$, where $a(i)$ represents the average distance between the data point and all other points in its cluster, and $b(i)$ represents the minimum average distance to points in any other cluster. A silhouette score of 1 implies highly dense and well-separated clusters. A score of 0 indicates an overlap between clusters, and a score below 0 suggests potential inaccuracies in data cluster assignment (Bhardwaj, 2020).

We aimed to obtain clusters with a minimum silhouette score of .3 or above. Cluster results are considered appropriate when the silhouette score is > 0.2. Though 0.2 is regarded as a fair score (Boos et al., 2021), we wished to provide a more robust clustering.

4. RESULTS

Social Media Use

We performed a two-step cluster analysis on the data provided by 1,502 users who responded to

 ISSN: 2473-4901
v9 n5949

the survey on social media usage. Social media usage was measured using predictor variables representing the use and non-use of 11 social media platforms, as tabulated for the WEB1 set of questions in the dataset for the questionnaire that can be obtained from the Social Media Use in 2021 report published by Pew Research Center (Auxier & Anderson, 2021). The two-step cluster analysis of social media usage displayed two clusters that we will first call Clusters 1 and 2.

The silhouette score was above 0.3 and was considered fair and meaningful. Cluster 1 comprised 61% of users in the dataset, and Cluster 2 comprised the remaining 39%. The predictor importance chart in Table 1 indicates the relative importance of each of the predictor variables in defining the cluster model.

Question: "Please tell me if you ever use any of the following. Do you ever use…	Predictor Importance
Pinterest?	0.2835
Nextdoor?	0.3631
TikTok?	0.4167
WhatsApp?	0.4199
Reddit?	0.4483
Facebook?	0.4991
Snapchat?	0.5183
Twitter?	0.6373
LinkedIn?	0.7181
YouTube?	0.8911
Instagram?	1

Table 1: Social Media Use Predictor Importance from the WEB1 Questions.

Table 2 shows the two clusters and the predictor variables arranged in the order of predictor importance. This chart also visually depicts, in the form of distinct bars, the use (and non-use) of each social media platform for users from each cluster.

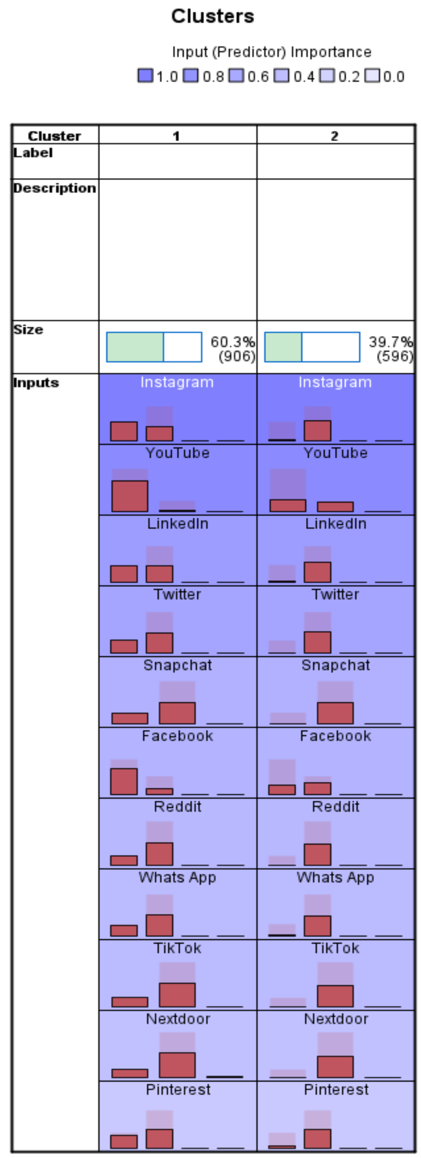

Table 2: Clusters with Input Predictors

Figure 2 presents another view of the predictor importance of the variables within each cluster.

ISSN: 2473-4901
v9 n5949

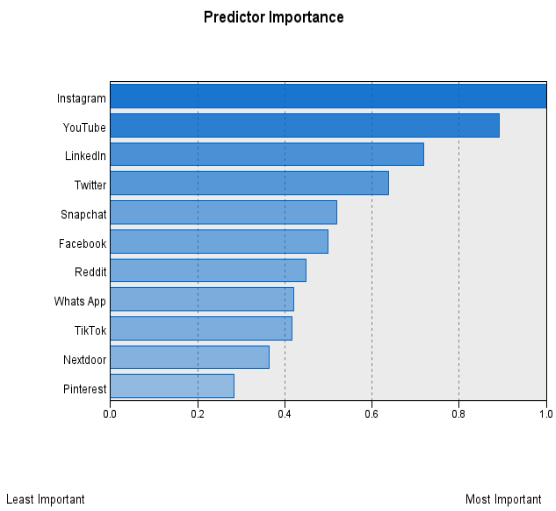

Figure 2: Predictor Importance of Cluster Variables

Table A.1 in Appendix A shows an example of crosstabulation results for the two-step clusters and Instagram usage. Similarly, results obtained for the remaining predictors are summarized in Appendix B. A correlation matrix was developed for each predictor, which is displayed in Appendix C.

Upon inspecting the usage of each social media platform, for each cluster, we determined that there are two clear groups of users: Multi-Platform (MP) social media users and Limited-Platform (LP) social media users. MP users are distinct from LP users. The MP users use all platforms surveyed and do so with a significant participation rate ranging from 24% to 97%, depending on the platform. The LP users primarily only use Facebook and YouTube, and even those platforms are 88% and 77%, respectively, more likely to be used by the MP user group. As shown in the last column in the table in Appendix B, we have calculated the percentage Cluster 1 is more likely to use each social media platform over Cluster 2 by diving the percentage of Cluster 1 users, who report using a specific platform, by the percentage of Cluster 2 users who report using that same platform, multiplying this result by 100 to obtain a percentage and subtracting 100. These results show that Twitter is 22324% more likely to be used by MP users than by LP users. Likewise, Reddit is 16812% more likely, and TikTok is 7894% more likely to be used by MP users than by LP users. The table shows a revealing picture of the current state of social media usage today. There are two distinct clusters of users, and though their participation rates vary by platform, there are significant differences in the usage of the platforms between the two groups. Some, such as Twitter, Instagram, Snapchat, Reddit, and TikTok, show tremendous differences; however, even Facebook and YouTube are nearly twice as likely to be used by MP than by LP users.

The values of the correlation coefficients (also known as the r values), tabulated in the correlation matrix in Appendix C, show many platform usages significantly correlated at $p < 0.001$. Only SnapChat and Nextdoor, YouTube and Nextdoor, and TikTok and Nextdoor are not correlated at $p < 0.05$. However, many correlations are not strong. Generally, an absolute r value less than or equal to 0.35 is viewed as showing a low or weak correlation. A value ranging from 0.36 to 0.67 represents a modest or moderate correlation, whereas a value from 0.68 to 1.0 indicates a strong or high correlation. An r coefficient equal to or greater than 0.90 symbolizes a very high correlation (Taylor, 1990). Many of the significant correlations in Appendix C can be considered low or weak. The ones that show modest or moderate correlation include TikTok and Snapchat, Twitter and Instagram, Twitter and TikTok, Twitter and Snapchat, Instagram and Facebook, and Instagram and TikTok. In addition, Facebook and YouTube show a moderate correlation as well. Twitter and Instagram use were more correlated with the usage of other platforms. Snapchat and TikTok usage showed relatively higher correlation coefficients with the use of other platforms.

Demographic Analyses

Demographic analyses of the two clusters are studied by considering predictors such as age, gender, educational attainment, and political affiliation that were obtained from social media users who responded to the survey.

Table 3 shows the mean age reported by users who are grouped under Clusters 1 (MP) and 2 (LP) by the two-step analysis. The active multi-platform users who characterize Cluster 1 are younger by almost two decades than the infrequent and limited platform users typical to Cluster 2.

 ISSN: 2473-4901
v9 n5949

Two-Step Cluster Number * AGE. What is your age?			
Cluster	Mean	N	Std. Dev
1	46.20	906	18.63
2	64.77	596	17.61
Total	53.57	1502	20.37

Table 3: Two-Step Cluster Number and Age

To analyze how gender responses from the social media users may be associated with the two clusters, we find the mean cluster number for each response value, as shown in Table 4. The mean cluster is the mean of the frequency of each of the responses for each cluster, weighted by the cluster number. Table A.2 in Appendix A shows the frequency of each of the gender related survey responses for each cluster number. As shown in Table 4, the majority of the respondents identified themselves along the traditional gender lines as a man, or a woman and the cluster numbers were somewhat the same for both of these responses. Table 4 indicates that although fewer in number, users who don't describe their gender as either male or female, and instead express their gender "in some other way" may have greater presence within Cluster 1 (MP). At the same time, users who refused to answer, or expressed that they "don't know" had higher mean values and therefore, could lean more towards the LP cluster (i.e., Cluster 2). While the survey did not ask the respondents to identify themselves as the LGBTQ+ group, prior studies have shown social media serves as informal learning environments for LGBTQ+ youth during their identity developmental processes (Fox & Ralston 2016; McInroy, Craig, & Leung, 2019). Therefore, there is a possibility that active use of social media may be prevalent with users who prefer to identify their gender "in some other way".

Two-Step Cluster Number * GENDER. Do you describe yourself as a man, a woman or in some other way?			
	Mean	N	Std. Dev
A man	1.40	854	.490
A woman	1.39	628	.489
In some other way	1.25	8	.463
Don't know	1.67	3	.577
Refused	1.78	9	.441
Total	1.40	1502	.489

Table 4: Two-Step Cluster Number and Gender

Table 5 shows the mean cluster number for various survey responses concerning marital status. Those who responded as "never being married" had a lower mean cluster number and, therefore, were more likely to fall under Cluster 1. This could also be indicative of the fact that Cluster 1 social media users tend to be younger (as evident from Table 3).

Two-Step Cluster Number * MARITAL. Are you currently married, living with a partner, divorced, separated, widowed, or have you never been married?			
	Mean	N	Std. Dev
Married	1.39	721	.488
Living with a partner	1.40	115	.492
Divorced	1.46	171	.500
Separated	1.36	36	.487
Widowed	1.71	107	.456
Never been married	1.26	325	.438
Don't know	1.67	3	.577
Refused	1.62	24	.495
Total	1.40	1502	.489

Table 5: Two-Step cluster number * marital status

Table 6 shows a relatively smaller mean cluster number associated with Democrats and Independents than with Republicans. Younger social-media users, who are more present in Cluster 1 might tend to affiliate themselves with the Democratic party (Statista, 2023b).

 ISSN: 2473-4901
v9 n5949

Two-Step Cluster Number * PARTY. In politics TODAY, do you consider yourself a Republican, Democrat, or Independent?			
	Mean	N	Std. Deviation
Republican	1.48	359	.500
Democrat	1.33	482	.469
Independent	1.36	470	.480
No preference	1.47	78	.503
Other party	1.39	18	.502
Don't know	1.45	20	.510
Refused	1.59	75	.496
Total	1.40	1502	.489

Table 6: Two-Step cluster number * Party

Educational attainment of respondents, as listed in Table 7, shows that social-media users with a higher level of education tend to have lower cluster numbers and, therefore, lean towards Cluster 1 and may tend to be MP. However, people with a bachelor's degree showed a lower mean cluster number than people with post graduate schooling and those with a post grad degree.

Two-Step Cluster Number *EDUC			
Education	Mean	N	Std. Dev
1 – Less than High School	1.65	17	.493
2 – High school incomplete	1.59	44	.497
3 -High school graduate	1.49	313	.501
4 – Some college, no degree	1.42	244	.494
5 – 2 yr. associate degree	1.49	156	.501
6 – 4 yr. bachelor's degree	1.29	389	.455
7 – Some Postgrad or Professional schooling	1.36	42	.485
8 – Post grad/professional degree	1.31	274	.462
98 – don't know	1.67	3	.577
99 – refuse to answer	1.75	20	.444
Total	1.40	1502	.489

Table 7: Two-Step cluster number * Educational Attainment

Income seems to influence social-media use. Table 8 shows that the mean cluster number is lower for users who reported a higher income. Therefore, there is a likelihood that users who have a higher income tend display the MP characteristics associated with Cluster 1 and people with a lower income tend to be LP and reside in Cluster 2.

Two-Step Cluster Number * INCOME			
INCOME. Last year, that is in 2020, what was your total family income from all sources, before taxes? Just stop me when I get to the right category.	Mean	N	Std. Dev
Less than $10,000	1.49	70	.503
10 to under $20,000	1.57	99	.498
20 to under $30,000	1.45	110	.500
30 to under $40,000	1.37	120	.484
40 to under $50,000	1.48	89	.503
50 to under $75,000	1.48	182	.501
75 to under $100,000	1.28	193	.453
100 to under $150,000	1.26	193	.439
$150,000 or more	1.27	217	.444
Don't know/Refused	1.52	229	.501
Total	1.40	1502	.489

Table 8: Two-Step cluster number * Income

5. DISCUSSION AND CONCLUSION

The result of this study, which is based on the survey data collected from the Pew Research, indicate that social media users form two clusters based on the number of social media platforms they use.

Based on the data collected from a nationwide survey, social-media users could be classified as multi-platform (MP) or limited-platform (LP) users. Our paper is an update of a prior study on the cluster analysis of social media user groups (Peslak, Ceccucci, & Hunsinger, 2022). That study reviewed a 2019 Pew data survey but did not include many newer platforms including TikTok and Nextdoor; therefore, this study expands upon the prior study. Cluster analysis of the 2019 social media usage data revealed three clusters but since then, as evidenced by the current study, the distinction has narrowed to only these two clusters, despite an increase in the number of social media platforms.

Based on the survey results, demographic analyses of users who fall under the two clusters indicates that MP users have a higher chance of reporting themselves as younger, single, and democrat, or independent. MP users also tend to report their gender non-traditionally, have a higher chance of attaining higher levels of education, and tend to report higher income levels. On the other hand, there is a greater chance that LP users are older, report lower income levels and educational attainment lower than a bachelor's degree. LP users also tend to report their political stance as leaning toward Republican and have a higher chance of refusing to report their gender identities, or of choosing the gender identity option of ‘do not know’. Survey responses also reveal that LP users have reported their marital status as widowed or as something that they 'do not know’.

The existence of two social media clusters that display divergent demographic characteristics may have several economic and social implications. One such implication may result from the spillover effects of incidental information exposure from one platform to another. Spillover effects have been observed in marketing of product brands that allocate their social media advertising across multiple platforms such as Facebook, Twitter, Instagram, and YouTube. Since consumers use multiple social media platforms, brand communications on one platform could potentially impact engagement with the brand on the other platforms; this phenomenon is known as spillover effect. By knowing the demographics that constitute multi-platform users, social media advertising could take advantage of the spillover of brand information from one platform into another. This spillover effect has been previously used to inform marketing resource allocation across platforms for a company’s brand (Unnava & Aravindakshan, 2021).

Similar spillover effects could also influence the way social media users consume news. Multi-platform social media news consumption affords diversified information and exposure to pro- and counter-attitudinal viewpoints (Lee, Choi, Kim, & Kim, 2014). At the same time, studies have shown how incidental, counter-attitudinal exposure enabled by multiple-platform social media use leads to a greater tendency of in-group support consisting of users from the same demographic profile and criticism against out-groups possibly consisting of users with different demographic characteristics (Guo & Chen, 2022). Therefore, awareness of the fact that social media clusters could display polarized demographic characteristics makes it critical to ensure that social media news content equitably serves a larger population.

The two social media usage clusters identified in this study vary based on two main factors that impact economic equality among the US population – education and income levels. MP users tend to report higher education levels and higher income and more LP users have reported lower education and income levels. The social benefits and opportunities afforded by networking via multiple social media platforms could go unrealized by people with lower incomes, who also typically tend to have lower educational attainment.

This study does not address the factors that could have led to the formation of the two social media clusters. More research is needed to investigate how factors such as unequal access to digital media, lack of digital skills, or the inability to leverage the affordances of social media could be reasons for the formation of demographically distinct social media clusters that are identified in this study. Nevertheless, the findings of this study indicate a possible digital divide among social media users based on their use of multiple platforms that could potentially confer more economical and social advantages to one group of demographics over the other. Future studies could systematically investigate why and how social media clusters are formed due to the demographic characteristics of users.

6. ACKNOWLEDGEMENTS

The article includes contents edited by ChatGPT, a language model developed by OpenAI in San Francisco, CA, USA. Additionally, ChatGPT provided assistance in sentence editing.

Papers and publications using our data should reference Pew Research Center’s Religion & Public life project as the source of the data and should acknowledge that the Center bears no responsibility for the interpretations presented or conclusions reached based on analysis of the data.

7. REFERENCES

Agarwal, N., & Liu, H. (2009). Modeling and Data Mining in Blogosphere. Morgan & Claypool Publishers.

ISSN: 2473-4901
v9 n5949

Auxier, B., & Anderson, M. (2021, April 7). Social Media Use in 2021. Retrieved June 15, 2023. https://www.pewresearch.org/internet/2021/04/07/social-media-use-in-2021/

Backstrom, L., & Leskovec, J. (2011). Supervised random walks: Predicting and recommending links in social networks. Proceedings of the Fourth ACM International Conference on Web Search and Data Mining, Hong Kong, China, 635-644.

Bhardwaj. A. (2020). Silhouette Coefficient. Retrieved June 15, 2023, from https://towardsdatascience.com/silhouette-coefficient-validating-clustering-techniques-e976bb81d10c

Boos, S. C., Wang, M., Karst, W. A., & Hymel, K. P. (2021). Traumatic Head Injury and the Diagnosis of Abuse: A Cluster Analysis. *Pediatrics,* 149(1). https://doi.org/10.1542/peds.2021-051742

Catanese, S., Meo, P.D., Ferrara, E., Fiumara, G., & Provetti, A. (2011). Crawling Facebook for social network analysis purposes. *Web Intelligence, Mining and Semantics*.

Fox J., & Ralston R. (2016). Queer identity online: Informal learning and teaching experiences of LGBTQ+ individuals on social media. *Computers in Human Behavior,* 65, 635–642. https://doi.org/10.1016/j.chb.2016.06.009

Gjoka, M., Kurant, M., Butts, C. T., & Markopoulou, A. (2010). Walking in Facebook: A case study of unbiased sampling of OSNs. Proceedings of the 2[9t]h Conference on Information Communications, San Diego, CA, 1-9.

Guo, J., & Chen, H.-T. (2022). How Does Multi-Platform Social Media Use Lead to Biased News Engagement? Examining the Role of Counter-Attitudinal Incidental Exposure, Cognitive Elaboration, and Network Homogeneity. *Social Media + Society*, 8(4). https://doi.org/10.1177/20563051221129140

Java, A., Song, X., Finin, T., & Tseng, B. (2007). Why we Twitter: Understanding microblogging usage and communities. Proceedings of the 9th WebKDD and [1s]t SNA-KDD 2007 Workshop on Web Mining and Social Network Analysis, San Jose, CA, 56-65.

Kaufman, L., & Rousseuw, P. J. (1990). Finding Groups in Data, New York: John Wiley and Sons.

Lee J. K., Choi J., Kim C., Kim Y. (2014). Social media, network heterogeneity, and opinion polarization. *Journal of Communication*, 64(4),702–722. https://doi.org/10.1111/jcom.12077

McAuley, J. J. & Leskovec, J. (2012). Learning to discover social circles in ego networks. Proceedings of the 2[5t]h International Conference on Neural Information Processing Systems, Lake Tahoe, NV, 1, 539-547.

McInroy L. B., Craig S. L., Leung V. W. Y. (2019). Platforms and patterns for practice: LGBTQ+ youths' use of information and communication technologies. *Child and Adolescent Social Work Journal*, 36, 507–520. https://doi.org/10.1007/s10560-018-0577-x

Naveed, N., Gottron, T., Kunegis, J., & Alhadi, A. C. (2011). Bad news travel fast: A content-based analysis of interestingness on Twitter. Proceedings of the 3[rd] International Web Science Conference, Koblenz, Germany, 1-7.

Methodology. (2021). https://www.pewresearch.org. Retrieved June 18, 2023, from https://www.pewresearch.org/internet/2021/04/07/social-media-use-methodology/

Peslak, Alan, Ceccucci, Wendy and Hunsinger. Scott (2022). Using unsupervised machine learning to determine social networking user groups. *Issues in Information Systems,* 23(2), 215-230.

Raghavan, U. N., Albert, R., & Kumara, S. (2007). Near-linear time algorithm to detect community structures in large-scale networks. *Physical Review E*, 76(3), 036106.

Riquelme, F., & González-Cantergiani, P. (2016). Measuring user influence on Twitter: A survey. *Information Processing & Management*, 52(5), 949-975.

Rizoiu, M. A., Xie, L., Sanner, S., Cebrian, M., Yu, H., & Van Hentenryck, P. (2017). Expecting to be HIP: Hawkes Intensity Processes for social media popularity. Proceedings of the 26th International Conference on World Wide Web, Republic and Canton of Geneva, Switzerland, 735-744.

Statista (2023a, February 14). https://www.statista.com. Most popular social networks worldwide as of January 2023, ranked by number of monthly active users. Retrieved June15,2023. https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/

ISSN: 2473-4901
v9 n5949

Statista. (2023b, April 6). https://www.statista.com. Party identification in the United States in 2022, by generation. Retrieved June 15,2023. https://www.statista.com/statistics/319068/party-identification-in-the-united-states-by-generation/

Taylor, R. (1990). Interpretation of the correlation coefficient: a basic review. *Journal of Diagnostic Medical Sonography*, *6*(1), 35-39.

Unnava, V., Aravindakshan, A. (2021) How does consumer engagement evolve when brands post across multiple social media? *Journal of the Academy of Marketing Science,* 49, 864–881. https://doi.org/10.1007/s11747-021-00785-z

U.S. Surveys. (2021). https://www.pewresearch.org. Retrieved June 18, 2023, from https://www.pewresearch.org/our-methods/u-s-surveys/

Wilson, J. (2020). How are clustering algorithms different from supervised learning? Technical-QA.COM. Retrieved June 15, 2022, from https://it-qa.com/how-are-clustering-algorithms-different-from-supervised-learning/¬.

Xu, Z., Zhang, J., Wu, Y., & Yang, Q. (2012). Modeling user posting behavior on social media. Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval, Portland, OR, 545-554.

Zafarani, R., & Liu, H. (2009). Connecting users across social media sites: A behavioral-modeling approach. Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Chicago, IL, 41-49.

 ISSN: 2473-4901
v9 n5949

Appendices and Annexures

APPENDIX A
Crosstabulation Results

Crosstabulation – Two-Step Cluster Number * WEB1B (Instagram use)

			WEB1B. Please tell me if you ever use any of the following. Do you ever use... **Instagram**?				Total
			Yes, do this	No, do not do this	Don't know	Refused	
Two-Step Cluster Number	1	Count	513	392	0	1	906
		% within Two-Step Cluster Number	56.6%	43.3%	0.0%	0.1%	100.0%
	2	Count	17	577	2	0	596
		% within Two-Step Cluster Number	2.9%	96.8%	0.3%	0.0%	100.0%
Total		Count	530	969	2	1	1502
		% within Two-Step Cluster Number	35.3%	64.5%	0.1%	0.1%	100.0%

Table A.1: Crosstabulation of Two-Step cluster number and Instagram use
Similar crosstabulations were generated for the Two-Step cluster number and the use of each of the social media types discussed in this paper. WEB1 is a question identifier that was used in the survey.

Two-Step Cluster Number * GENDER. Do you describe yourself as a man, a woman or in some other way? Crosstabulation

			A man	A woman	In some other way	Don't know	Refused	Total
Two-Step Cluster Number	1	Count	515	382	6	1	2	906
		% within Two-Step Cluster Number	56.8%	42.2%	0.7%	0.1%	0.2%	100.0%
	2	Count	339	246	2	2	7	596
		% within Two-Step Cluster Number	56.9%	41.3%	0.3%	0.3%	1.2%	100.0%
Total		Count	854	628	8	3	9	1502
		% within Two-Step Cluster Number	56.9%	41.8%	0.5%	0.2%	0.6%	100.0%

Table A.2: Crosstabulation of Two-Step cluster number and Gender responses
Similar crosstabulations were generated for the Two-Step cluster number and each of the demographic variables discussed in this paper.

ISSN: 2473-4901
v9 n5949

APPENDIX B
Participation of users from each cluster for each social media platform

Cluster->	**1**	**1**	**2**	**2**	1 vs 2
User Participation Response ->	**Yes**	**No**	**Yes**	**No**	more likely
Twitter	38.12%	61.88%	0.17%	99.83%	22324%
Instagram	56.69%	43.31%	2.86%	97.14%	1882%
Facebook	80.97%	19.03%	43.03%	56.97%	88%
Snapchat	33.37%	66.63%	0.84%	99.16%	3873%
YouTube	96.91%	3.09%	54.70%	45.30%	77%
WhatsApp	33.81%	66.19%	3.38%	96.62%	900%
Pinterest	39.98%	60.02%	12.58%	87.42%	218%
LinkedIn	49.83%	50.17%	5.37%	94.63%	828%
Reddit	28.75%	71.25%	0.17%	99.83%	16812%
TikTok	27.18%	72.82%	0.34%	99.66%	7894%
Nextdoor	24.13%	75.87%	0.67%	99.33%	3501%

ISSN: 2473-4901
v9 n5949

APPENDIX C
Social media usage survey responses correlation matrix. Sample size n = 1502. Moderate correlation coefficients (r value) are highlighted in gray.

	WEB1A **Twitter?**	WEB1B **Instagram?**	WEB1C **Facebook?**	WEB1D **Snapchat?**	WEB1E **YouTube?**	WEB1F **WhatsApp?**	WEB1G **Pinterest?**	WEB1H **LinkedIn?**	WEB1I **Reddit?**	WEB1J **TikTok?**	WEB1K **Nextdoor?**
Twitter	**1**	**.460**	**.289**	**.420**	**.351**	**.233**	**.238**	**.328**	**.350**	**.427**	**.099**
		<.001	*<.001*	*<.001*	*<.001*	*<.001*	*<.001*	*<.001*	*<.001*	*<.001*	*<.001*
Instagram	**.460**	**1**	**.325**	**.508**	**.392**	**.242**	**.310**	**.279**	**.296**	**.439**	**.102**
	<.001		*<.001*	*<.001*	*<.001*	*<.001*	*<.001*	*<.001*	*<.001*	*<.001*	*<.001*
Facebook	**.289**	**.325**	**1**	**.294**	**.387**	**.226**	**.318**	**.212**	**.165**	**.270**	**.066**
	<.001	*<.001*		*<.001*	*<.001*	*<.001*	*<.001*	*<.001*	*<.001*	*<.001*	*0.01*
Snapchat	**.420**	**.508**	**.294**	**1**	**.339**	**.192**	**.287**	**.196**	**.285**	**.545**	**0.041**
	<.001	*<.001*	*<.001*		*<.001*	*<.001*	*<.001*	*<.001*	*<.001*	*<.001*	*0.113*
YouTube	**.351**	**.392**	**.387**	**.339**	**1**	**.252**	**.305**	**.293**	**.265**	**.345**	**0.048**
	<.001	*<.001*	*<.001*	*<.001*		*<.001*	*<.001*	*<.001*	*<.001*	*<.001*	*0.063*
Whats App	**.233**	**.242**	**.226**	**.192**	**.252**	**1**	**.146**	**.303**	**.155**	**.169**	**.126**
	<.001	*<.001*	*<.001*	*<.001*	*<.001*		*<.001*	*<.001*	*<.001*	*<.001*	*<.001*
Pinterest	**.238**	**.310**	**.318**	**.287**	**.305**	**.146**	**1**	**.172**	**.155**	**.305**	**.070**
	<.001	*<.001*	*<.001*	*<.001*	*<.001*	*<.001*		*<.001*	*<.001*	*<.001*	*0.006*
Linked	**.328**	**.279**	**.212**	**.196**	**.293**	**.303**	**.172**	**1**	**.215**	**.162**	**.145**
	<.001	*<.001*	*<.001*	*<.001*	*<.001*	*<.001*	*<.001*		*<.001*	*<.001*	*<.001*
Redditt	**.350**	**.296**	**.165**	**.285**	**.265**	**.155**	**.155**	**.215**	**1**	**.272**	**.089**
	<.001	*<.001*	*<.001*	*<.001*	*<.001*	*<.001*	*<.001*	*<.001*		*<.001*	*<.001*
TikTok	**.427**	**.439**	**.270**	**.545**	**.345**	**.169**	**.305**	**.162**	**.272**	**1**	**0.041**
	<.001	*<.001*	*<.001*	*<.001*	*<.001*	*<.001*	*<.001*	*<.001*	*<.001*		*0.109*
Nextdoor	**.099**	**.102**	**.066**	**0.041**	**0.048**	**.126**	**.070**	**.145**	**.089**	**0.041**	**1**
	<.001	*<.001*	*0.01*	*0.113*	*0.063*	*<.001*	*0.006*	*<.001*	*<.001*	*0.109*	