8-2024

An Online Scientific Twitter World: Social Network Analysis of #ScienceTwitter, #SciComm, and #AcademicTwitter

Man Zhang
Utah State University

AN ONLINE SCIENTIFIC TWITTER WORLD: SOCIAL NETWORK ANALYSIS OF

#ScienceTwitter, #SciComm, AND #AcademicTwitter

by

Man Zhang

A thesis submitted in partial fulfillment
of the requirements for the degree

of

MASTER OF SCIENCE

in

Instructional Technology and Learning Sciences

Approved:

Lisa Lundgren, Ph.D. Major Professor	Ha Nguyen, Ph.D. Committee Member
Hillary Swanson, Ph.D. Committee Member	D. Richard Cutler, Ph.D. Vice Provost of Graduate Studies

UTAH STATE UNIVERSITY

Logan, Utah

2024

ABSTRACT

An Online Scientific Twitter World: Social Network Analysis of #ScienceTwitter, #SciComm, and #AcademicTwitter

by

Man Zhang, Master of Science

Utah State University, 2024

Major Professor: Dr. Lisa Lundgren
Department: Instructional Technology and Learning Science

Understanding who makes up online social worlds as well as how information flows within those communities is important as more people access news, research topics, collaborate with others, and entertain themselves. I analyzed a dataset that includes 53,311 users and 100,000 tweets with three hashtags #ScienceTwitter, #SciComm, and #AcademicTwitter. This study answers the following three questions: Who are the members of the online world of #ScienceTwitter, #SciComm, and #AcademicTwitter? How is scientific information shared and consumed in this online world? How do members influence and control the flow of information in the Twitter network? In this study, I identified and classified the people discussing scientific topics on Twitter into three categories by using multiclass classification, visualized and determined the social network structure through NodeXL, and described the member composition of this online world by identifying the locations of influential users. Scientists, the public, and educators formed this online world. They built connections by initiating activities and interacting with others, which created the Community Clusters social network structure. All three categories of people are in positions of influence in this network leading and controlling the conversations. The results show that scientists, the public, and educators share the space and contribute to communication in this online world.

(53 pages)

PUBLIC ABSTRACT

An Online Scientific Twitter World: Social Network Analysis of #ScienceTwitter, #SciComm, and #AcademicTwitter

Man Zhang

Understanding who makes up online social worlds as well as how information flows within those communities is important as more people access news, research topics, collaborate with others, and entertain themselves. This study identified and classified the people discussing scientific topics on Twitter, determined the type of social network, and described the member composition of this online world. Scientists, the public, and educators formed this online world. They built connections by initiating activities and interacting with others, which created the Community Clusters social network structure. All three categories of people are in positions of influence in this network leading and controlling the conversations. The results show that scientists, the public, and educators share the space and contribute to communication in this online world.

ACKNOWLEDGMENTS

I would like to express my sincere gratitude to my academic advisor, Dr. Lisa Lundgren, for her invaluable guidance and encouragement. I appreciate her helping me familiar with scientific concepts and research theories and tools as I had no previous experience in research. Whenever I was upset, excited, nervous, or curious about my research, she was always beside me. Thanks to Dr. Lisa Lundgren, my confidence has grown. I would like to extend my sincere thanks to my committee members, Dr. Ha Nguyen and Dr. Hillary Swanson, for their guidance and insightful feedback throughout my research. I benefited a lot from their professional suggestions and warm words.

I also would like to express my appreciation to the ITLS department. Everyone I met here was very friendly, we had plenty of activities, great service, and wonderful communication. I missed the time I was on campus, and I love all I experienced in this department. Furthermore, I would like to thank the members of our lab, for helping me improve my presentation skills, providing writing advice, and always praising and rooting for me.

I am deeply grateful to my parents, Honglun Zhang and Zhengli Wu, and my older brother, Bin Zhang, for their full support of my pursuit of a master's degree. I would like to thank my husband, Rui Gao, for always being supportive and trustworthy. I want to thank my lovely son, Yibo Gao, for his smile that relieved my stress. I would like to express my gratitude to all my friends for their kind encouragement. My family and my friends are my strongest support and give me the motivation to keep moving forward.

Man Zhang

CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I INTRODUCTION

Understanding who makes up online social worlds in microblogging spaces such as Twitter[1], as well as how information flows within those communities is important as more people access news, research topics, collaborate with others, and entertain themselves. In recent years, scholars and scientists are increasingly using the microblogging service Twitter as a communication platform to share their work and opinions. Some of the current research on the topic of science on Twitter focuses on specific user groups and their communication practices. For example, Lee et al. (2020) analyzed the science communication practices of a for-profit company. In addition, Bex and colleagues attempted to define the members of a paleontological community on Twitter (2019).

However, most of the current research is narrowly focused on very specific disciplines and rarely focuses on broad scientific topics. Additionally, researchers have also analyzed specific users, rather than entire groups in larger social networks or communities. For these reasons, additional research is needed to explore the composition of the scientific world on Twitter across a wide range of scientific topics, and how information is transmitted and controlled in this online world.

The benefits of this line of research include increased knowledge of who is part of the scientific community on Twitter, more awareness of scientific topics on social media, and how scientific information is transmitted and processed in these communities. There are many formal online spaces for education, like Canvas, MOOCs, etc. These represent instances of

[1] At the end of July 2023, Twitter was renamed “X”. But the majority of research was done when the platform was still known as Twitter (and, is more widely known as Twitter).

nontraditional/informal learning spaces. The results of this research can be used to provide the possibility of other informal online learning spaces. This research project identified and classified the people discussing scientific topics on Twitter, determined the type of social network, and described the member composition of this online community.

Research Questions

The following questions were used as a guide for this study:

1. Who are the members of the online world of #ScienceTwitter, #SciComm, and #AcademicTwitter?
2. How is scientific information shared and consumed in this online world?
3. How do members influence and control the flow of information in the Twitter network?

CHAPTER II LITERATURE REVIEW

Scientists and the public can make important contributions when interacting with others on open social media platforms on scientific topics (e.g., Ahmed et al., 2020; Bex et al., 2019). Research in this area is still needed to determine the social networks and dissemination of scientific information contained in a broad range of scientific fields. This literature review of the research is necessary to determine the next step in this line of research. This literature review had three objectives:

- To describe the current state of research on scientific topics on Twitter, including research using methods of social network analysis.
- To discuss the issues, strengths, and weaknesses in previous research.
- To draw conclusions based on this review and formulate my study's research questions and methods accordingly.

Keyword Search

Google Scholar and Utah State University Library online resources, including APA PsycArticles, Education Source and Education Resources Information Center (ERIC), were used to locate empirical studies on the analysis in a scientific field or topic on Twitter that have been published between 2012-2022. A variety of search terms were used both singularly and in combination, including, but not limited to: *Twitter, social media, social network analysis, science, and science communication.*

Inclusion and Exclusion Criteria

Articles included in this literature review met the following criteria:

- The study is a peer-reviewed primary source.
- It is an empirical study that was published between 2012-2022.
- The study examined scientific fields or topic analysis on Twitter.

Summary of the Literature

Several main findings emerged from the review, which produced eight articles.: Twitter social network structure can be better understood through analysis of users; a need for classifying tweets; and the influence of Twitter on the online world.

First, the Twitter social network is presented through the analysis of Twitter users. This was found in four of the eight studies reviewed. For example, Ahmed et al. (2020) found that an isolated group and a broadcast group constituted the two largest social network structures through an analysis of 5G and COVID-19 conspiracy theories on Twitter. The analysis also showed that many users retweeted fake news links without an authority actively fighting such misinformation. Additionally, Bex et al. (2019) defined the members of a paleontology social world from three levels (Structure, Category, and Type) and how members from different Categories such as scientist and public were disseminating messages. Further, Bhandoria et al. (2021) confirmed the network of the '#IGCS2020' on Twitter as a community network shape with elements of broadcast. They also identified the ten most influential Twitter users within this community, five of whom are individual users. Finally, Moukarzel et al. (2020) categorized influencers into three categories: scientific community (SC), interested citizens (IC), and for-profit companies. SC includes academics, researchers, health care practitioners, and/or non-governmental agencies, while IC represents the public. Thus, these studies that examined Twitter through Twitter users found that different categories of users all play a role in the dissemination of information. Categorizing and analyzing Twitter users can help understand who the people are in the Twitter online world.

Second, content analysis focuses on tweet categorization and its effect on message dissemination. For example, Bex et al. (2019) and Lundgren et al. (2022) classified five types of messages: Information, News, Opportunity, Research, and Off-Topic. They also found different post types were determined to be effective or ineffective at creating connections with different segments of the membership. The most successful, engaged with, and far-reaching expression of scientific practice were Information posts. Ineffective messaging was found within the post types of News and Research. Additionally, Bombaci et al. (2016) defined the type of session in a conservation science conference and found Twitter effectively conveyed conservation science to a diverse and somewhat unexpected audience beyond the conference. Finally, Su et al. (2017) divided tweets into three categories based on the communication function: Information, Participant, and Community, and determined what type of communication they were and what their purpose was. The classification of Twitter content could better define the kinds of information dissemination and thus determine what kind of information is highly effective.

Third, Twitter is a great place for the dissemination of information, where people connect and interact in this online world. For example, Moukarzel et al. (2020) found Twitter is an opportunity space for the scientific community, including researchers, to effectively communicate science to the public. The science community does face a wide range of identified challenges: influencers had less heterogeneous relations than companies; they engage in more activity but reach fewer unique individuals; and they primarily use networks for research, announcements, and commercial purposes (Moukarzel et al. (2020). The study by Anderson et al. (2017) highlighted how social media discussions have the potential to influence the way people engage with science. Finally, Lee et al. (2020) found that companies communicate about various topics related to health and the environment that can have substantial implications at the individual and society levels. Their study strongly suggests that 23andMe communicates about genetic science beyond explaining its own products but that these efforts have declined over time. It provided a basis for further research on some companies, their science communication, and the effects of that communication.

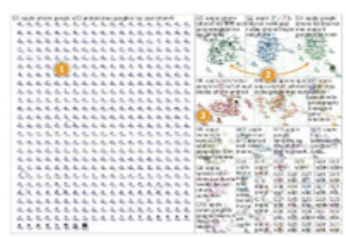
Social Network Analysis

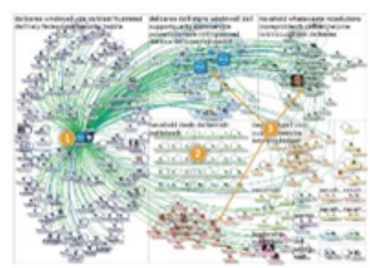
A network is a collection of things and their relationships to one another. Connections are made when people interact. Social networks are created whenever people interact directly or indirectly with other people, institutions, or artifacts (Hansen et al., 2020). Social network analysis can visualize complex relationships through graphs (Hansen et al., 2020). Social network analysis displays connected individuals and calculates the size, shape, and density of the entire network. As people retweet, reply to, and mention each other on Twitter, conversations on Twitter form networks. Researchers used social network analysis to examine topics around Twitter, identifying the centers of information dissemination and making recommendations for the promotion of scientific knowledge (Brajawidagda, 2012; Lee et al.,2017; Ahmed et al., 2020; Milani et al., 2020). Smith et al. (2014) suggested mapping social media networks can enable a better understanding of the variety of ways individuals form groups and organize online. They illustrated six different structures of connections around different kinds of topics on Twitter (Smith et al., 2014) (Table 1).

Previous research has been conducted on some specific scientific topics or users, and user categorization helps with knowing who people are within an online world, and the analysis of social networks can learn how the information is controlled to be transmitted. Thus, our research analyzes the broader scientific topic on Twitter, including who is in the social world (Twitter users) and how information spreads through it and influences user activity (social network analysis).

Table 1

Six main Twitter topic network structures

Name	Features	Example
Polarized Crowd	Polarized discussions feature two big and dense groups that have little connection between them. The topics being discussed are often highly divisive and heated political subjects. There is usually little conversation between these groups despite the fact that they are focused on the same topic. Polarized Crowds on Twitter are not arguing. They are ignoring one another while pointing to different web resources and using different hashtags.	
Tight Crowd	These discussions are characterized by highly interconnected people with few isolated participants. Many conferences, professional topics, hobby groups, and other subjects that attract communities take this Tight Crowd form.	
Brand Clusters	When well-known products or services or popular subjects like celebrities are discussed in Twitter, there is often commentary from many disconnected participants: These "isolates" participating in a conversation cluster are on the left side of the picture on the left). Well-known brands and other popular subjects can attract large, fragmented Twitter populations who tweet about it but not to each other. Brand-mentioning participants focus on a topic but tend not to connect to each other.	

Name	Features	Example
Community Clusters	Some popular topics may develop multiple smaller groups, which often form around a few hubs each with its own audience, influencers, and sources of information. These Community Clusters conversations look like bazaars with multiple centers of activity. Global news stories often attract coverage from many news outlets, each with its own following. That creates a collection of medium-sized groups—and a fair number of isolates.	
Broadcast Network	Twitter commentary around breaking news stories and the output of well-known media outlets and pundits has a distinctive hub and spoke structure in which many people repeat what prominent news and media organizations tweet. The members of the Broadcast Network audience are often connected only to the hub news source, without connecting to one another. In some cases, there are smaller subgroups of densely connected people— think of them as subject groupies—who do discuss the news with one another.	
Support Network	Customer complaints for a major business are often handled by a Twitter service account that attempts to resolve and manage customer issues around their products and services. This produces a hub and spoke structure that is different from the Broadcast Network pattern. In the Support Network structure, the hub account replies to many otherwise disconnected users, creating outward spokes. In contrast, in the Broadcast pattern, the hub gets replied to or retweeted by many disconnected people, creating inward spokes.	

CHAPTER III CONCEPTUAL FRAMEWORK

Affinity spaces are unique spaces that unite people with common interests or passions (Gee, 2004). Unlike communities of practice that emphasize membership or apprenticeship, Gee (2004) uses the term "space" to define another social configuration that focuses on the interaction when people engage and learn. This space is not just about people and the content in the space and how people interact through that content. A particular one is called affinity space (Gee, 2012), where people gather to interact with each other, to share practices and ideas based on a common interest. There are 11 features to identify the degree to which a space is an affinity space (Table 2). Affinity space is not an all-or-nothing thing. If a space has more of these features, it can be considered more of an affinity space or be closer to a paradigmatic affinity space.

Affinity spaces are common in today's world. Fans of all things (e.g., movies, comic books, TV shows, video games, various lifestyles) create and maintain affinity spaces (Min et al., 2018; Shafirova et al., 2020; Barany & Foster, 2021; Dynel & Ross, 2022). Many businesses organize such spaces (Gee, 1996). Social activists also often organize themselves and others in terms of affinity spaces (Pour-Khorshid, 2018). And scientists from many different disciplines connect with others around the globe in a variety of ways, gradually taking on more and more of the features of an affinity space (Sharma & Land 2018; Fontaine et al., 2019). These spaces are not just physical spaces, they can also be virtual spaces on the web at a distance.

Neely and Marone (2016) explored the participation of people with similar interests in informal social and learning activities by identifying 11 affinity space features in a specific physical space--jam band parking lots. Some researchers have explored affinity spaces on social

Table 2

The 11 features defining an affinity space from Gee (2012)

Affinity Space Feature (Gee, 2012)	Example of Twitter Fulfilling the Feature
Common endeavor, not race, class, gender, or disability, is primary.	People can connect based on shared interests, endeavors, goals, or practices. Users are not required to highlight race, gender, age, disability, or social class.
Newbies and masters and everyone else share common space.	Celebrities, politicians, scientists, students, and the public are sharing their thoughts on Twitter.
Some portals are strong generators.	Various portals exist on Twitter, such as hashtags, links to external websites, or links to other Twitter users. They are created or shared by users. Other users can browse or skip these portals.
Content organization is transformed by interactional organization.	Many hashtags on Twitter are not original but derived from other websites.
Both intensive and extensive knowledge are encouraged.	On Twitter, some scientists have specialized knowledge as well as educators with broad knowledge.
Both individual and distributed knowledge are encouraged.	Twitter provides a space to enable people to gain individual knowledge (stored in their heads) and learn to use and contribute to distributed knowledge by sharing personal ideas and retweeting other users' tweets.
Dispersed knowledge is encouraged.	Twitter encourages and enables people to use dispersed knowledge that is not actually at the site itself, but at other sites such as YouTube, academic conferences, etc., or in other spaces.
Tacit knowledge is encouraged and honored.	People share their personal lives and experiences. Examples include how to deal with the editors of journals and how to quickly familiarize themselves with a college.
There are many different forms and routes to participation.	People participate on Twitter by creating posts, following hashtags/users, and interacting with others, as well as finding content with the search function.

Affinity Space Feature (Gee, 2012)	Example of Twitter Fulfilling the Feature
There are lots of different routes to status.	Twitter allows people to achieve status in many ways, such as having many followers, positive replies and interactions, and being content creators.
Leadership is porous and leaders are resources.	On Twitter, there is no set leader; individuals can gain leadership by being a valuable resource, such as a reliable user of knowledge-sharing and information or an expert in solving all mechanical problems.

media, such as Twitter and Reddit, and other social networking sites. Greenhalgh et al. (2020) explored a teacher-focused Twitter hashtag, #mchiED, to determine whether different learning spaces exist in chat and non-chat tweets using the criteria of volume, content, interaction, and portal of the affinity space. Staudt Willet (2019) explored how and why educators use Twitter affinity space generated by #Edchat. The study categorized tweet types (including original posts, peer-retweets, self-retweets, extended posts, etc.) and tweet purposes (self, others, mutual, and miscellaneous). Staudt Willet and Carpenter (2020) also explored the change and continuity of two teaching-related subforums on Reddit and the contributions and interactions of four teaching-related subforums, respectively. Sharma and Land (2018), on the other hand, focused on knowledge sharing and interaction patterns in the discourse of a diabetic online affinity space. These studies suggest that affinity spaces can be identified in social media such as Twitter, and that affinity space features of specific topics will be more pronounced (e.g., groups of fans or groups of educators). Affinity spaces serve as the conceptual framework to determine whether this online science Twitter world fulfills the eleven affinity space features by categorizing people across a wide range of science hashtags, conducting social network analysis, and analyzing information dissemination.

CHAPTER IV METHODOLOGY

Research Design

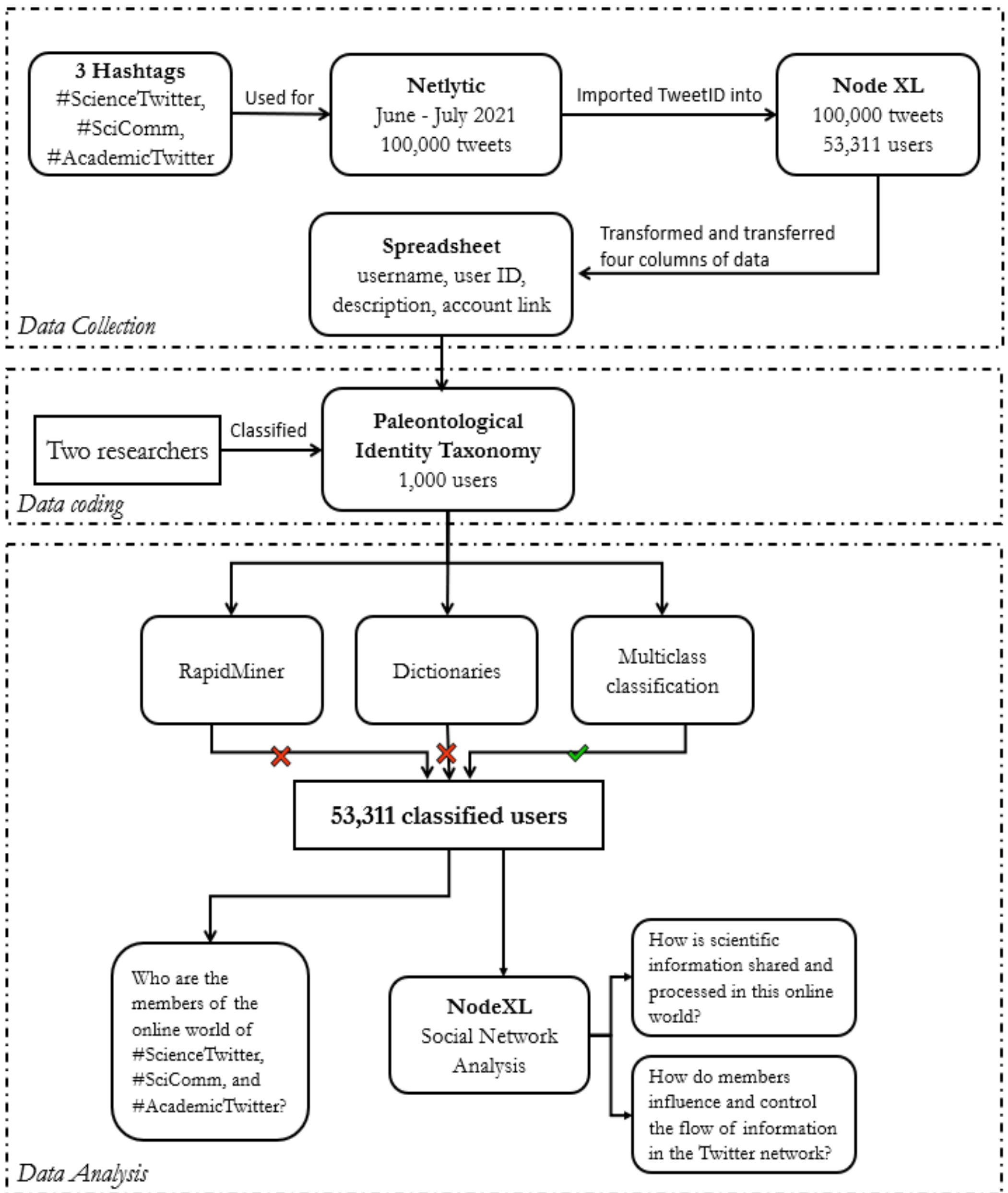
This study aims to identify members of the Twitter world on scientific topics and understand the wider conversations that occur in academic, scientific Twitter spaces to determine the social network structure. In this study, I used an existing Tweets dataset collected between June and July 2021 which I then processed and obtained a new database containing more attributes including Twitter users and their relationships. I classified users and visualized the social network based on the new dataset. This process consists of three stages: data collection, data coding, and data analysis (Figure 1).

Data Collection

The original dataset was collected from three hashtags: #ScienceTwitter, #SciComm, and #AcademicTwitter. These hashtags were chosen using hashtagify.me, a hashtag tracking tool, as well as knowledge from our researcher of using Twitter as an academic for the past ten years. Netlytic, a browser-based text and social network analysis service (Gruzd, 2016), was used to schedule a sampling of the Twitter public search API (application programming interface) every 15 minutes for a one-month period (June - July 2021). This process resulted in a dataset of 100,000 tweets with several attributes including TweetID. I then imported the TweetIDs from the dataset into NodeXL (Smith et al., 2010), a Microsoft Excel Add-In which allows for researchers to collect and analyze more attributes than Netlytic. Using NodeXL, the original 100,000 tweet dataset was associated with 53,311 Twitter users. This transformed dataset included more attributes such as username, description, and user relationship (i.e. mentions, retweets, replies to, etc.). The NodeXL dataset was used for social network visualization and analysis. Four specific columns of Twitter user data in this dataset including User ID, Vertex, Description, and Custom Menu Item Action, were transformed and transferred to a spreadsheet as User ID, Username, Bio, and Link to Profile.

Figure 1

Procedure



Data Coding

Two researchers manually classified 1,000 Twitter users to prepare the data machine learning classification of all users. Based on NodeXL's feature of randomly sorting usernames, we selected the first 1,000 users for classification. We used a taxonomy called Paleontological Identity Taxonomy (PIT) (Lundgren et al., 2018) for our manual coding process. PIT is a tool that can be used to classify members of digital social spaces. It uses members' self-descriptions to classify them into three different levels. This three-tiered taxonomy (i.e. structure, category, type) is based on self-identity with the domain (Lundgren et al., 2018). The main unit of analysis in this study is at the category level, where users are divided into Public, Scientists, or Education and Outreach (Table 3). We classified users based on the content of each Twitter user's bio (i.e., account description), discussing any discrepancies to consensus (Patton, 2002).

Table 3

Category of Paleontological Identity Taxonomy (PIT)

Category	Definition
Public	An entity that does not meet the definition of Scientist and Education and Outreach.
Scientist	Any entity that uses a scientific domain to classify themselves, use of "-ist"; students (graduate or undergraduate) using their major; centers, institutes, and research groups are included if they indicate their audience to be other scientists.
Education and Outreach	Any entity reference to working: in a k-12 setting; as a teacher, lecturer, or in a classroom; in/as a museum or the main focus of the account is education; reference to providing some kind of advocacy or promotion of diversity, equity, and inclusion efforts or providing services to populations that might not otherwise have access to those services (i.e. outreach)

Data Analysis

To answer research question 1: *who are the members of the online world of #ScienceTwitter, #SciComm, and #AcademicTwitter*, all 53,311 users were classified into one of three categories. Due to the large number of Twitter users in the whole dataset, the dataset was deemed appropriate for computational analyses. I have tried to use different methods for classification. RapidMiner is a software that can call different algorithms for data mining and processing. When I fed the sampled usernames into RapidMiner, it had a high accuracy rate (.96) in the Scientist category and most of the users who were actually in other categories were also classified into the category of Scientist. Emulating other researchers, I also tried to build dictionaries for classification (Côté et al., 2018; Li, 2019; Toupin et al., 2022; Walter et al., 2019), achieved high accuracy rates for Scientist (.95), Education and Outreach (.87), but not Public (0.46). However, since my study requires classifying users into one and only one category, the Scientist and Education and Outreach categories had many overlapping words thus reducing accuracy.

I decided to use multiclass classification for categorization (Grandini et al., 2020). Multiclass classification is a classification task with more than two classes. Each sample can only be labeled as one class. Several researchers have applied the multiclass classification to the analysis of Twitter user content (Balabantaray et al., 2012; Ceron et al., 2015; Bouazizi & Ohtsuki, 2019; Li et al., 2019; AlSomaikhi & Alzamil, 2020). I used Scikit-Learn (Pedregosa et al., 2011), a Python library, to implement machine learning models. Then I divided the 1000 manually coded users into a training set (800 users) and a test set (200 users) to evaluate seven models (80:20 ratio) with 5-fold cross-validation. The models included logistic regression, random forest, linear Support Vector Machine, Support Vector machine, multinomial Naive

Bayes, Stochastic Gradient Descent and Multilayer Perceptron. I extracted features from user bios text using the term frequency-inverse document frequency (TF-IDF, evaluating the importance of different words in a sentence). Due to good model performance, with high accuracy along with high precision, recall, and f1-score in all three categories, I selected Stochastic Gradient Descent as the best model for classification (Table 4). Performance metrics included accuracy as well as precision (ratio of true positives and total positives predicted), recall (ratio of true positives to all actual positives), and F1 score (harmonic mean of precision and recall) for each category and macro (arithmetic mean) and weighted values (mean while considering each class's support).

Table 4

Classification model performance evaluation

Model	Category	precision	recall	f1-score	support
Logistic Regression	Public	0.85	0.53	0.65	74
	Scientist	0.52	0.96	0.67	81
	Education and Outreach	1.00	0.07	0.12	45
	accuracy			0.60	200
	macro avg	0.79	0.52	0.48	200
	weighted avg	0.75	0.60	0.54	200
Random Forest	Public	1.00	0.19	0.32	74
	Scientist	0.44	1.00	0.61	81
	Education and Outreach	0.00	0.00	0.00	45
	accuracy			0.48	200
	macro avg	0.48	0.40	0.31	200
	weighted avg	0.55	0.47	0.36	200
Linear Support Vector Machine	Public	0.74	0.82	0.78	74
	Scientist	0.66	0.84	0.74	81
	Education and Outreach	0.73	0.24	0.37	45
	accuracy			0.70	200
	macro avg	0.71	0.64	0.63	200
	weighted avg	0.71	0.70	0.67	200

Model	Category	precision	recall	f1-score	support
Support Vector Machine (SVC)	Public	1.00	0.22	0.36	74
	Scientist	0.44	1.00	0.61	81
	Education and Outreach	1.00	0.02	0.04	45
	accuracy			0.49	200
	macro avg	0.81	0.41	0.34	200
	weighted avg	0.77	0.49	0.39	200
Multinomial Naive Bayes	Public	0.70	0.69	0.69	74
	Scientist	0.56	0.86	0.68	81
	Education and Outreach	1.00	0.02	0.04	45
	accuracy			0.61	200
	macro avg	0.75	0.53	0.47	200
	weighted avg	0.71	0.61	0.54	200
Stochastic Gradient Descent (SGD)	Public	0.78	0.64	0.70	74
	Scientist	0.71	0.80	0.76	81
	Education and Outreach	0.55	0.60	0.57	45
	accuracy			**0.69**	200
	macro avg	0.68	0.68	**0.68**	200
	weighted avg	0.70	0.69	**0.69**	200
Multilayer Perceptron (MLP)	Public	0.60	0.86	0.71	74
	Scientist	0.73	0.65	0.69	81
	Education and Outreach	0.76	0.36	0.48	45
	accuracy			0.64	200
	macro avg	0.70	0.62	0.63	200
	weighted avg	0.69	0.67	0.65	200

Following the classification of user biographies, I used NodeXL to conduct a social network analysis to visualize the network structure with the Harel-Koren Fast Multiscale algorithm. All users were grouped into clusters using the Clauset-Newman-Moore algorithm (Clauset et al., 2004), which enables the discovery of subgroups within the larger dataset. The nodes represent the users and the edges represent five connection types between the users (Table 5). Three different colors were used to distinguish the Public, Scientist, and Education and

Outreach categories. This network showed the finding of research question 2*: how is scientific information shared and processed in this online world*.

Table 5

Connection types in the Twitter network

Type	Definition	How it shows in a network
Mentions	A user creates a Tweet containing another user's name, indicated by the "@" character preceding the other user's name.	A line from the user to another mentioned user.
Retweet	A user reposts or forwards a Tweet written by someone else.	A line from the user to another retweeted user.
Mentions in Retweet	A user mentioned in the original post when the other user retweeted the Tweet.	A line from the user to the mentioned user.
Replies to	A user responds to another user's Tweet.	A line from the user to the user being replied to.
Tweet	A user posts an original message without any other users' information.	A self-loop.

To answer research question 3: *how do members influence and control the flow of information in the Twitter networ*k, I used social network graph metrics. Specifically, centrality measures of influence: Degree, Betweenness, Closeness, and Eigenvector were calculated by NodeXL (Smith et al., 2010). Degree centrality measures the number of connections a person has in the network. (Hansen et al., 2020, p. 83). Betweenness centrality helps identify individuals who play a “bridge spanning” role in a network (Hansen et al., 2020, p. 83). Closeness centrality shows that if information needs to flow through the network, a person may need a few or many steps to send messages to all other people (Hansen et al., 2020, p. 83). The eigenvector centrality network metric considers not just “how many people you know,” but also “who you know” (Hansen et al., 2020, p. 84). These measures can identify key people in influential locations in the discussion network, highlighting the people leading the conversation. Additionally, network graph metrics including the numbers of edges, vertices, density, and other features were collected. These metrics summarize the key properties of the entire network, which helps understand the network.

CHAPTER V RESULTS

This social world contains 53,311 members and 136,126 connections (i.e. original tweets, replies, and retweets). Regarding categories, 45% were Scientist (n = 24,125), 32% were Public (n = 16,803), and 23% were Education and Outreach (n = 12,383). While scientists remain the primary participants in scientific topics on Twitter, the public and educational outreach are also shared this space to communicate about science.

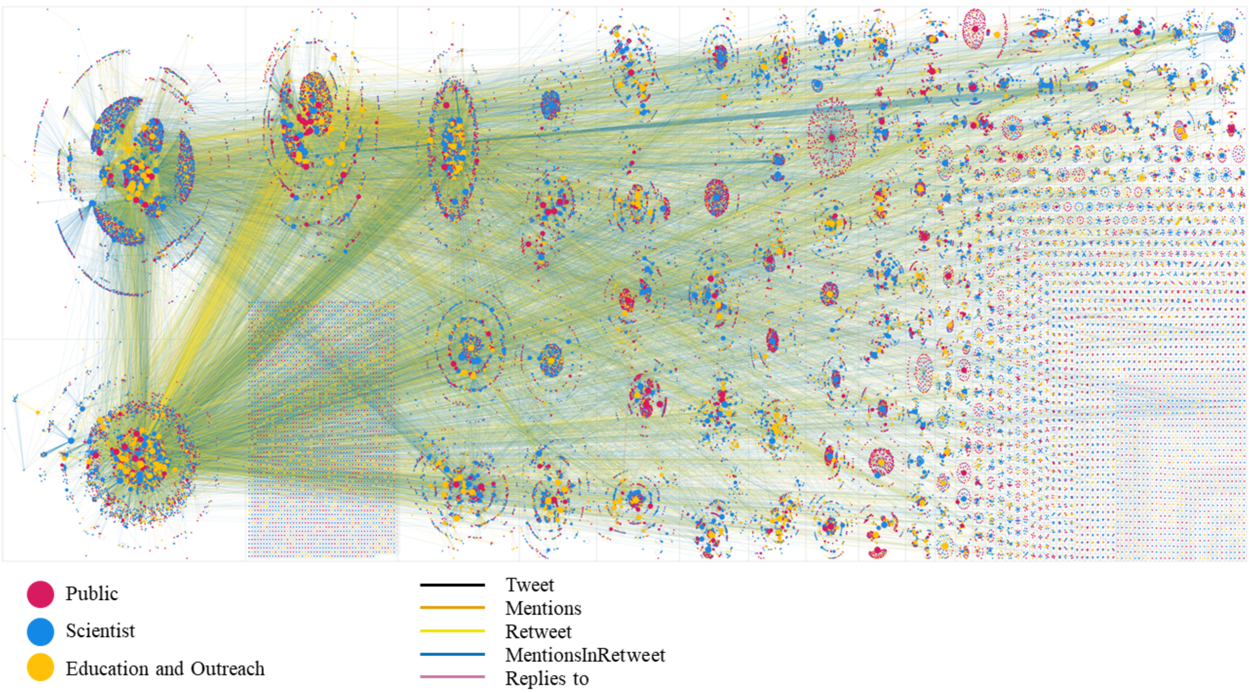
Social Network Analysis

The entire social network revealed that the members and their interactions in this online world form Community Clusters (Figure 2). According to Smith et al. (2014), the Community Clusters structure is usually formed by some popular topics and may develop multiple smaller groups, which often form around a few hubs each with its own audience, influencers, and sources of information. These Community Clusters conversations look like bazaars with multiple centers of activity (Smith et al., 2014). This structure creates a collection of medium-sized groups—and a fair number of isolates.

Conversations surrounding the three hashtags of #ScienceTwitter, #SciComm, and #AcademicTwitter consisted of a total of 2,240 groups, ranging from 2 to 6,247 people. Most groups were medium-sized groups formed around a few central entities. There is also one group containing 3,093 individuals, who were isolated, meaning that they did not communicate with anyone else. People built connections with others in groups in 4 main ways: Mentions, Retweet, Mentions in Retweet, and Replies to (Table 5). Of the 136,126 connections created by 53,311 users, there were 59,941 Mentions in retweets, 49,134 Retweets, 17,237 Mentions, 7,976 tweets, and 1,838 Replies (Table 6). The distinct differences in numbers (i.e. 7,976 tweets versus 59,941 mentions in retweet) means that most people in this network interact with others in different ways. Figure 3 shows two central users' self-descriptions and categories, and how their tweets spread across the social network.

Figure 2

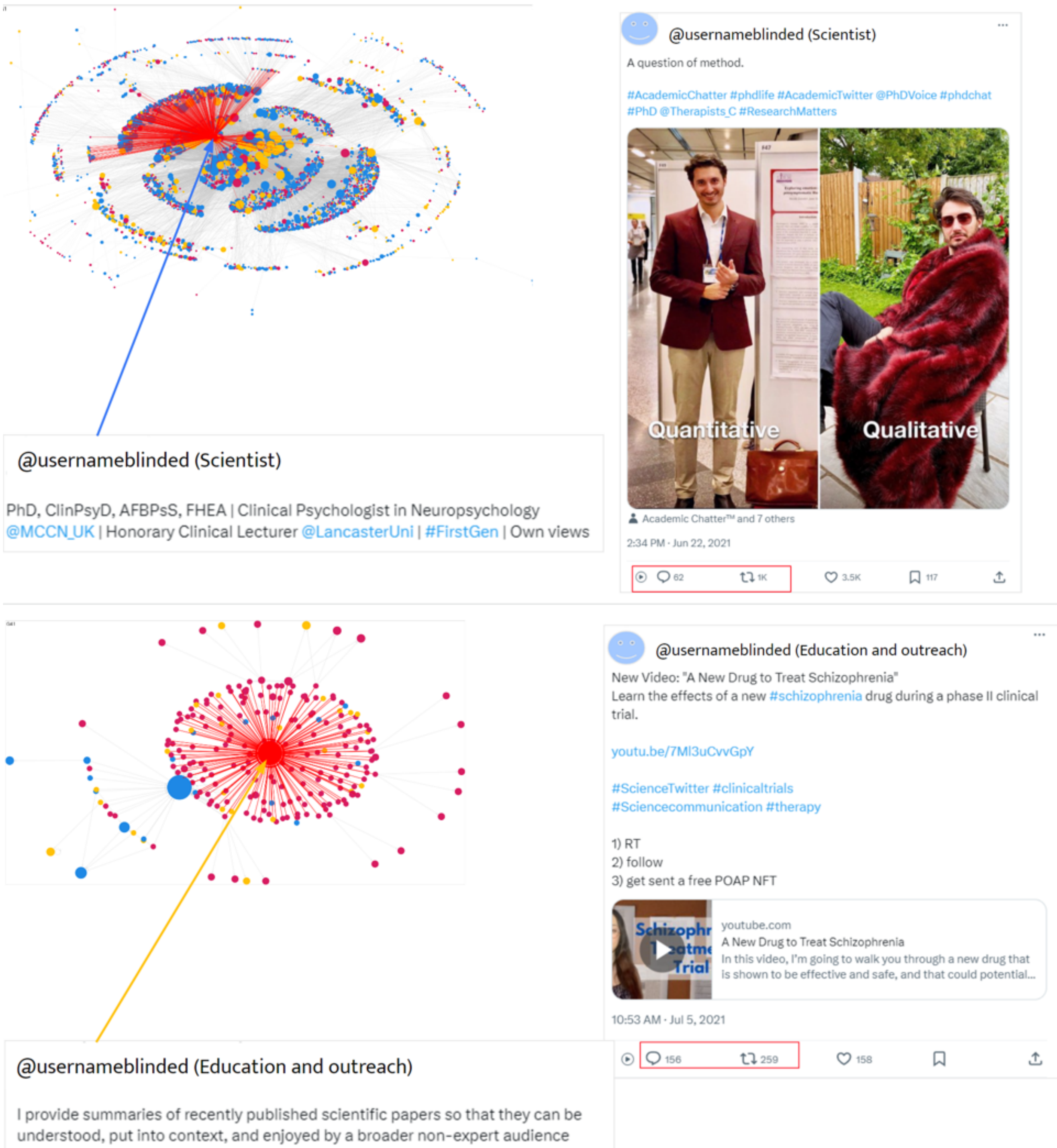
Entire Social Network graph



Note. Red disks represent Public, blue disks represent Scientist, and yellow disks represent Education and Outreach. The size of users corresponds to betweenness centrality. Lines of different colors represent connection events between two people, i.e. Mentions, Retweet, Mentions in retweet, and Replies to. Self-loops, indicated by a black circular loop, represent the same user linking back to themselves, i.e. tweet with no other interaction.

Figure 3

Social network of two users



Note. The blue and yellow arrows point to the two users' positions in the social network, and the red arrows indicate the connections they built.

Table 6

Overall network graph metrics

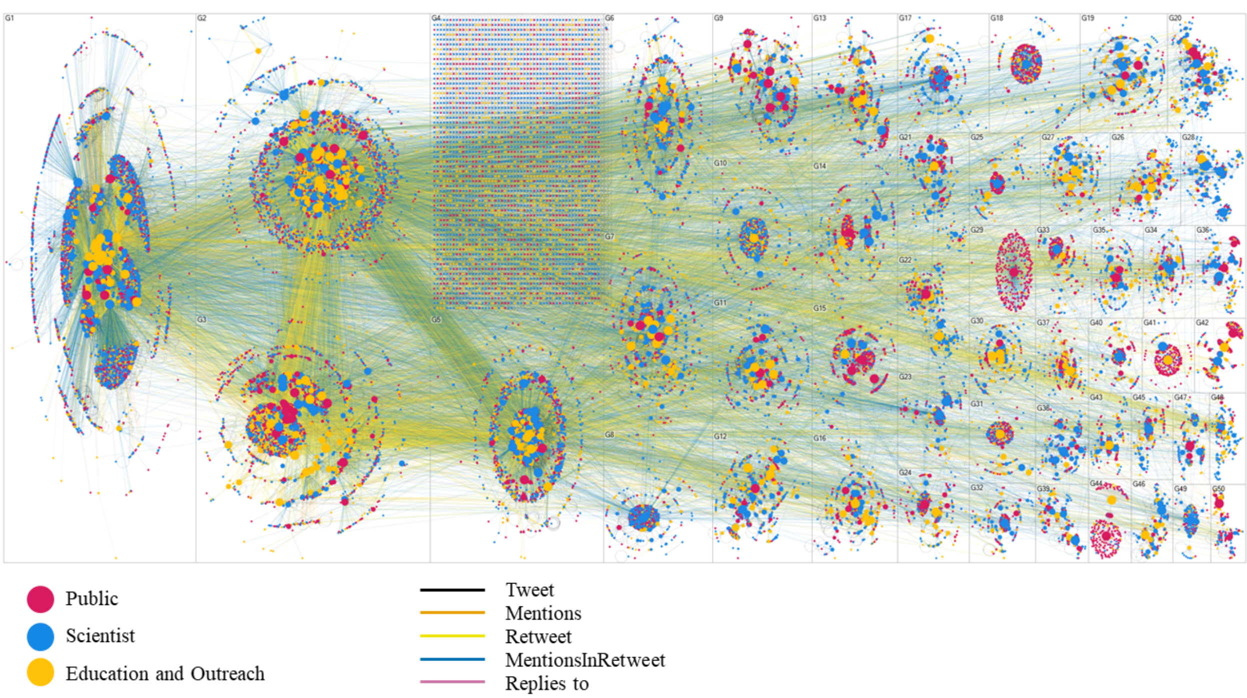
Graph Metric	Value
Graph Type	Directed
Vertices	53,311
Total Edges	136,136
Number of Edge Types	5
Mentions	17,237
Tweet	7,976
Retweet	49,134
MentionsInRetweet	59,941
Replies to	1,838
Self-Loops	8,210
Reciprocated Vertex Pair Ratio	0.03613
Reciprocated Edge Ratio	0.06975
Connected Components	4,700
Single-Vertex Connected Components	3,093
Maximum Vertices in a Connected Component	44,955
Maximum Edges in a Connected Component	127,042
Maximum Geodesic Distance (Diameter)	18
Average Geodesic Distance	4.42965
Graph Density	0.00005
Modularity	0.71383
Groups	2240

In this network, the maximum geodesic distance (diameter) is 18 and the average geodesic distance is approximately equal to 4. The geodesic distance is the length of the shortest path between two people in a network. It gives a sense of how “close” community members are to one another. This suggests that the farthest distance between two people from each other is not too far and the average length of the distance between all pairs of people is small. The graph density is 0.00005. The graph density is a number between 0 and 1 and is calculated from the number of actual connections in the network and the number of possible connections that are determined by the number of people in the network. It's a measure of how interconnected people are in the network. Hansen et al. (2020) suggest larger social networks tend to have lower graph density. In this large online world, people form a less tightly connected community structure despite interacting and connecting with other members of the same group or other groups.

We selected the first 50 groups to display the new social network graph (Figure 4). It is important to note that for this network, the vast majority of groups--2,190 groups--consisted of small groups, isolates or dyads. These isolates or dyadic groups indicate that the majority of people in this social network did not interact with others or interacted little and were limited to a few people they knew.

Figure 4

First 50 groups Social Network graph

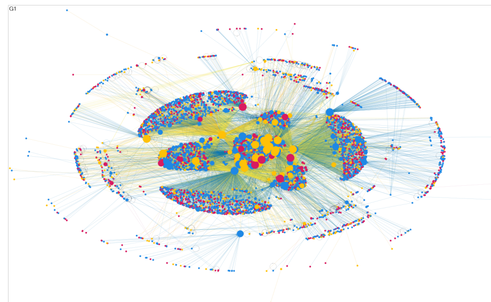
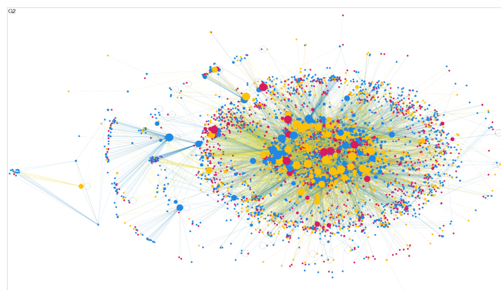
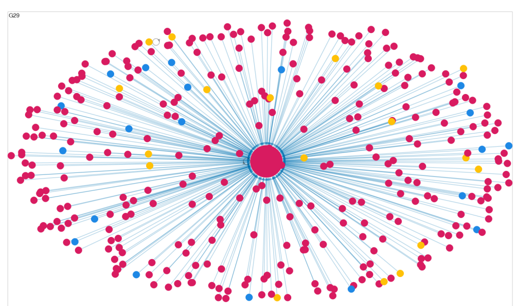
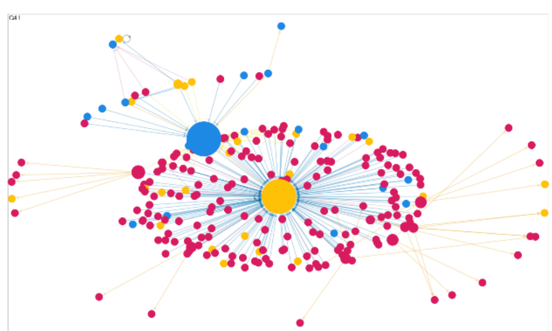


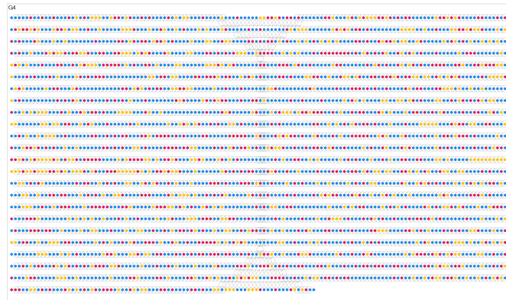
Note. Red disks represent Public, blue disks represent Scientist, and yellow disks represent Education and Outreach. The size of users corresponds to betweenness centrality. Lines of different colors represent connection events between two people, i.e. Mentions, Retweet, Mentions in retweet, and Replies to. Self-loops, indicated by a black circular loop, represent the same user linking back to themselves, i.e. tweet with no other interaction.

Among these 50 groups, most of them showed shapes of clusters, with a few broadcast networks and one isolated group (Table7). Group 1 (label G1) exhibited a shape centered on certain people, which indicates that the sources of information and the paths of dissemination are in the hands of Scientist and Education and Outreach users. Group 2 (label G2) showed a group formed by three categories of people in almost equal numbers, which suggests that the Public, Scientist, and Education and Outreach users are discussing scientific topics together. Some groups showed a broadcast network where members were usually connected only to the central user and not to each other. Groups 29 (label G29) and 41 (label G41) show such a network. This means that other members in these groups reply or retweet the central member, which often signifies that the central member has some influence. And these two groups also demonstrated the existence of groups where almost only one category of people exists, which indicates that within such groups, people only build connections with people who are close to them in terms of their identity. Group 4 (label G4), on the other hand, was an isolated group. It was distinguished from any of the other groups by the fact that it is made up of a certain number of independent members of different categories that do not connect with any other groups or persons. This suggests in this group people posted tweets discussing relevant topics but did not interact with anyone else.

Table 7

Social Network Structures

Network Name	Graph
Community Clusters	
Group 1	
Group 2	
Broadcast Network	
Group 29	
Group 41	

Network Name	Graph
Isolated Group	
Group 4	

People within groups interact, and those at the center of influence build bridges of communication externally, allowing different groups to connect and form this large social network. Of course, people influence in different ways.

User Analysis

Measuring how central users are reveals influential users and their connections to one another. Influential users were ranked by the score of degree centrality, betweenness centrality, closeness centrality, and eigenvector centrality. Category descriptions were also provided for all users.

Degree centrality measures the number of connections a user has in the network. In-degree centrality measures the connections others initiate with a user. Users with high in-degree centrality scores can be seen as the center of communication since others mention, reply to, or retweet their posts. Out-degree centrality counts the connections a user initiates with others. A high out-degree centrality score means a user tweets a lot about topics to gain attention by mentioning or replying to others. These two metrics demonstrate which members control the flow of information and the level of engagement among members. All top ten users in this network of in-degree centrality are Scientist and Education and Outreach users (Table 8). The highest in-degree centrality, 3,417, is an Education and Outreach user. This means that these users are the center of this online world and hold the source of information. Their posts receive attention and are shared. Six of the top ten users of out-degree centrality are Public, four of which are bots (Table 9). The user with the highest out-degree centrality (4,112) is a bot. These bot accounts automatically retweet the tweets with specific hashtags, resulting in high out-degree centrality. High in-degree users didn't have high out-degree at the same time. This demonstrates that they either focus only on others or only on themselves. There is only one Education and Outreach user who has both a high in-degree and out-degree score. This means that this user actively interacts with others while people pay attention to their posts.

Table 8

Top ten users by in-degree centrality

Rank	Category	In-Degree Centrality	Out-Degree Centrality	Network Group	Note
1	Education and Outreach	3417	73	G1	
2	Scientist	2984	177	G1	
3	Education and Outreach	2175	492	G1	
4	Scientist	893	1	G8	
5	Education and Outreach	783	1	G10	
6	Scientist	702	33	G1	
7	Scientist	610	4	G18	
8	Scientist	577	12	G1	
9	Education and Outreach	575	132	G1	
10	Scientist	565	4	G17	

Table 9

Top ten users by out-degree centrality

Rank	Category	In-Degree	Out-Degree	Network	Note
1	Public	2	4112	G5	bot
2	Education and Outreach	1	3631	G2	
3	Public	29	3132	G2	bot
4	Scientist	55	1658	G3	
5	Education and Outreach	35	1493	G3	
6	Public	0	508	G3	bot
7	Education and Outreach	2175	492	G1	
8	Public	14	480	G3	
9	Public	13	476	G3	bot
10	Public	105	253	G14	

Betweenness centrality helps identify individuals who play a “bridge-spanning” role in a network. In this social world, eight of the top ten users of betweenness centrality are composed of Scientist and Education and Outreach categories, and a few Public (Table 10). The highest betweenness centrality user is from the category of Public, which is also a bot account that has the highest out-degree centrality. All ten users have high in-degree or out-degree centrality scores. This shows that Scientist and Education and outreach users build connections throughout the social network by acting as a bridge to other users.

Table 10

Top ten users by betweenness centrality

Rank	Category	Betweenness Centrality	In-Degree Centrality	Out-Degree Centrality	Network Group	Note
1	Public	487,998,649.044	2	4,112	G5	bot
2	Education and Outreach	380,847,553.172	1	3,631	G2	
3	Education and Outreach	361,239,568.951	3,417	73	G1	
4	Public	357,300,566.203	29	3,132	G2	bot
5	Scientist	254,143,564.522	2,984	177	G1	
6	Education and Outreach	175,013,982.037	2,175	492	G1	
7	Education and Outreach	75,409,275.180	783	1	G10	
8	Scientist	72,572,046.716	893	1	G8	
9	Scientist	55,259,967.550	55	1,658	G3	
10	Scientist	50,486,043.866	610	4	G18	

Closeness centrality shows how closely connected people are in the network. Users with the highest closeness centrality scores are mainly in the Education and Outreach category (Table 11). This social world features a poor closeness centrality—0.336 was the highest observed—this indicates that users are relatively distant from each other throughout the entire Twitter network. Users with both a high closeness centrality and a high in-degree centrality are closer to others and can get information to others relatively quickly. Similarly, users with both a high closeness centrality and a high out-degree centrality can share others' posts relatively quickly. This means that these users are the first to be considered when information needs to be effectively communicated and spread to the majority of people in this online world.

Table 11

Top ten users by closeness centrality

Rank	Category	Closeness Centrality	In-Degree Centrality	Out-Degree Centrality	Network Group	Note
1	Public	0.336	2	4,112	G5	bot
2	Public	0.328	29	3,132	G2	bot
3	Education and Outreach	0.325	1	3,631	G2	
4	Education and Outreach	0.316	3417	73	G1	
5	Scientist	0.310	2984	177	G1	
6	Education and Outreach	0.304	2175	492	G1	
7	Scientist	0.293	55	1,658	G3	
8	Education and Outreach	0.289	35	1,493	G3	
9	Education and Outreach	0.287	90	88	G1	
10	Public	0.287	13	476	G3	bot

Eigenvector centrality helps identify who is most influential. The users with high eigenvector centrality are mostly those categorized as Public and Education and Outreach (Table 12). This indicates these users have many connections with others while being highly connected to some popular individuals. This might impact how information flows in the social world. Information disseminated among these users is more effective relative to others because they're influential and the users they are connected with are equally influential.

Table 12

Top ten users by eigenvector centrality

Rank	Category	Eigenvector Centrality	In-Degree Centrality	Out-Degree Centrality	Network Group	Note
1	Public	0.401	2	4,112	G5	bot
2	Public	0.331	29	3,132	G2	bot
3	Education and Outreach	0.309	1	3,631	G2	
4	Scientist	0.168	55	1,658	G3	
5	Education and Outreach	0.152	35	1,493	G3	
6	Education and Outreach	0.130	3,417	73	G1	
7	Scientist	0.122	2,984	177	G1	
8	Education and Outreach	0.106	2,175	492	G1	
9	Public	0.079	14	480	G3	
10	Public	0.072	13	476	G3	bot

Betweenness centrality, closeness centrality, and eigenvector centrality measures help identify who is important or central in this social world. They may be important bridge-builders to connect other different parts of the network, or they may be at the center of the network getting attention from or giving attention to other members. These key people are in influential positions and lead the conversations in this online world. Without them, the message would be difficult to send and share.

CHAPTER VI DISCUSSION

The results of this study showed that 53,311 members of the online world of #ScienceTwitter, #SciComm, and #AcademicTwitter consists of three categories: Scientist, Public, and Education and Outreach. All three categories are weighted rather than one category having a particularly large or small number of members. Scientists and the public remain an important part of the science-related conversation on Twitter (Bex et al., 2019; Moukarzel et al., 2020; Bhandoria et al., 2021). We found that users working in education and outreach also make up a significant portion (23%, n=12,383). This validates that educators use twitter as a tool for accessing information, participating in their respective communities of interest, and sharing their insights on specific topics (Malik et al., 2019).

We found that scientific information is shared among these members and forms a social network with a Community Clusters structure. This shows that the online world that under #ScienceTwitter, #SciComm, and #AcademicTwitter is shared regardless of expertise. This study also shows that social network analysis does not have to be limited to structural analysis (Smith et al., 2014); rather, social networks can be better understood by identifying users in the social world. This helps us understand the important and central members in the network to consider who will be the authoritative and quick members to spread and share information. This is important especially when specific information needs to be disseminated, and those members who are more connected in the network can help break down information barriers for the purpose of promoting or explaining science (Ahmed et al., 2020).

Within this online world, Scientist, Public, and Education and outreach members all have important information dissemination roles. Although the members with high centrality scores contain four bot users, according to their descriptions, they are all accounts that use the automated features instead of violating user privacy and posting spam messages. They forward useful information, automatically generate content or automatically reply to other users via direct messages. Their presence makes information dissemination faster, and their outward activities (retweets) build more connections in this world. Most research focuses on how to identify Twitter bots rather than discussing how to properly utilize bot accounts for science information campaigns (Lee et al., 2013; Minnich et al., 2017; Feng, 2021). In this study, bots users were found as having centrality scores in a network. In future research, it may be possible to explore how to use bot accounts to collect information, such as forwarding posts from specific accounts and extracting keywords to simplify information and then share the information to the public.

This study demonstrates that the science-based affinity space on Twitter is composed of diverse users attracted to certain hashtags. There are many different forms and routes to participation (Gee, 2012). People participate on Twitter by creating posts, following hashtags, and interacting with others (e.g. retweeting, replying, mentioning). In addition, this affinity space fulfills certain features evidenced by the centrality measure, including sharing the space regardless of expertise, providing multiple routes to status, and a porous leadership structure (Gee, 2012). Affinity space is not limited to physical space. With the current development of science and technology, digital space breaks geographical and time constraints, and more and more interactions and content sharing are taking place. Our research confirms this. In this online scientific Twitter world, people from diverse backgrounds come together to pursue common endeavors for scientific communication. From scientists to educators to the public, people can be novices or experts, express their opinions, and connect with people they know or do not know. Our research promotes the study of affinity spaces by identifying the degree of features. The online world formed by the three scientific hashtags does not fulfill all 11 features, which shows that the degree of this affinity space could still be improved. This network density is notably low (0.0005). By improving affinity space degree, users are more engaged, have access to better resources, and can interact more meaningfully with each other, they are likely to learn more and develop their skills and knowledge more effectively. We see similarities in our study to Staudt Willet's (2019) and Neely and Marone's (2016) studies in that certain feature of the affinity space need to be strengthened. This study contributes to understanding science learning within informal, online spaces.

Limitations

Since this study focused on determining who the users in the Twitter world are who discuss scientific topics and how they communicate, not much attention was paid to the content of the tweets themselves. In addition, the data was collected over a period of only one month, and it cannot be ruled out that a specific time or a specific event will affect the density of people's engagement in the topic. Although we include languages other than English which means analyzing Twitter worldwide, there are still limitations here since Twitter is banned or hardly used in some countries.

Conclusion

The purpose of this study was to identify the members of the Twitter online world formed by three scientific hashtags: #ScienceTwitter, #SciComm, and #AcademicTwitter, who are composed of Scientist, Public and Education and Outreach users. They connected and spread information with other members of this online world through initiated activities (tweets) and interactions with others (retweets, reply to, mentions, mentions in retweets). These connections created a Community Clusters social network structure. In this network, all three types of users were playing key roles. Some users actively connected with others, while some users focused only on themselves. Visualizing social networks helps to understand how information is disseminated while identifying the different categories of key users who can influence discussions helps to understand who should be considered as sources of information.

REFERENCES

Ahmed, W., Vidal-Alaball, J., Downing, J., & López Seguí, F. (2020). COVID-19 and the 5G conspiracy theory: Social network analysis of Twitter data. *Journal of Medical Internet Research*, *22*(5), e19458. https://doi.org/10.2196/19458

AlSomaikhi, N. A., & Alzamil, Z. A. (2020). Twitter users' classification based on interest. *International Journal of Information Retrieval Research, 10*(1), 1–12. https://doi.org/10.4018/IJIRR.2020010101

Anderson, A. A., & Huntington, H. E. (2017). Social media, science, and attack discourse: How Twitter discussions of climate change use sarcasm and incivility. *Science Communication, 39*(5), 598–620. https://doi.org/10.1177/1075547017735113

Balabantaray, R., Mohammad, M., & Sharma, N. (2012). Multi-Class Twitter Emotion Classification: A New Approach. *International Journal of Applied Information Systems, 4*(1), 48–53. https://doi.org/10.5120/ijais12-450651

Barany, A., & Foster, A. (2021). Context, community, and the individual: modeling identity in a game affinity space. *The Journal of Experimental Education*, *89*(3), 523–540. https://doi.org/10.1080/00220973.2020.1740966

Bex, R. T., Lundgren, L., & Crippen, K. J. (2019). Scientific Twitter: The flow of paleontological communication across a topic network. *Plos One*, *14*(7), e0219688. https://doi.org/10.1371/journal.pone.0219688

Bhandoria, G. P., Nair, N., Jones, S. E. F., Eriksson, A. G., Hsu, H.-C., Noll, F., Ahmed, W., & International Gynecological Cancer Society. (2021). International Gynaecological Cancer Society (IGCS) 2020 Annual Global Meeting: Twitter activity analysis. *International*

Journal of Gynecological Cancer, 31(11), 1453–1458. https://doi.org/10.1136/ijgc-2021-002781

Bombaci, S. P., Farr, C. M., Gallo, H. T., Mangan, A. M., Stinson, L. T., Kaushik, M., & Pejchar, L. (2016). Using Twitter to communicate conservation science from a professional conference. *Conservation Biology, 30*(1), 216–225. https://doi.org/10.1111/cobi.12570

Bouazizi, M., & Ohtsuki, T. (2019). Multi-class sentiment analysis on twitter: Classification performance and challenges. *Big Data Mining and Analytics, 2*(3), 181–194. https://doi.org/10.26599/BDMA.2019.9020002

Brajawidagda, U. (2012). Twitter tsunami early warning network: a social network analysis of Twitter information flows. *Faculty of Engineering and Information Sciences - Papers: Part A.*

Ceron, A., Curini, L., & Iacus, S. M. (2015). Using sentiment analysis to monitor electoral campaigns. *Social Science Computer Review, 33*(1), 3–20. https://doi.org/10.1177/0894439314521983

Clauset A., Newman M. E. J., Moore C. (2004). Finding community structure in very large networks. *Physical Review E*, 70(6), Article 066111. https://doi.org/10.1103/PhysRevE.70.066111

Côté, I. M., & Darling, E. S. (2018). Scientists on Twitter: Preaching to the choir or singing from the rooftops*? FACETS, 3*(1), 682–694. https://doi.org/10.1139/facets-2018-0002

Dynel, M., & Ross, A. S. (2022). Metarecipient parents' #Bluey tweets as a distributed fandom affinity space. *Poetics*, *92*, 101648. https://doi.org/10.1016/j.poetic.2022.101648

Feng, S., Wan, H., Wang, N., Li, J., & Luo, M. (2021). TwiBot-20: A Comprehensive Twitter Bot Detection Benchmark. *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*, 4485-4494. https://doi.org/10.1145/3459637.3482019

Fontaine, G., Maheu-Cadotte, M.-A., Lavallée, A., Mailhot, T., Rouleau, G., Bouix-Picasso, J., & Bourbonnais, A. (2019). Communicating science in the digital and social media ecosystem: scoping review and typology of strategies used by health scientists. *JMIR Public Health and Surveillance*, *5*(3), e14447. https://doi.org/10.2196/14447

Gee, J. P., G. Hull and C. Lankshear (1996) *The New Work Order: Behind the Language of the New Capitalism.* Boulder: Westview.

Gee, J. P. (2004). Affinity Spaces. In *Situated Language and Learning*. Routledge.

Gee, J. P. (2012). *Situated language and learning: A critique of traditional schooling.* Routledge. https://doi.org/10.4324/9780203594216

Grandini, M., Bagli, E., & Visani, G. (2020). Metrics for Multi-Class Classification: an Overview. *ArXiv.* https://doi.org/10.48550/arxiv.2008.05756

Greenhalgh, S. P., Rosenberg, J. M., Staudt Willet, K. B., Koehler, M. J., & Akcaoglu, M. (2020). Identifying multiple learning spaces within a single teacher-focused Twitter hashtag. *Computers & Education, 148*, 103809. https://doi.org/10.1016/j.compedu.2020.103809

Gruzd, A. (2016). *Netlytic: Software for Automated Text and Social Network Analysis*. https://netlytic.org/home/

Hansen, D. L., Shneiderman, B., Smith, M. A., & Himelboim, I. (2020). *Analyzing social media networks with NodeXL: Insights from a connected world* (2nd ed., pp. 83–84). Morgan Kaufmann. https://doi.org/10.1016/C2018-0-01348-1

Lee, M. K., Yoon, H. Y., Smith, M., Park, H. J., & Park, H. W. (2017). Mapping a Twitter scholarly communication network: a case of the association of internet researchers' conference. *Scientometrics,112*(2), 767–797. https://doi.org/10.1007/s11192-017-2413-z

Lee, N. M., Abitbol, A., & VanDyke, M. S. (2020). Science communication meets consumer relations: An analysis of Twitter use by 23andMe. *Science Communication, 42*(2), 244–264. https://doi.org/10.1177/1075547020914906

Lee, S., & Kim, J. (2013). WarningBird: A Near Real-Time Detection System for Suspicious URLs in Twitter Stream. *IEEE Transactions on Dependable and Secure Computing*, *10*(3), 183–195. https://doi.org/10.1109/TDSC.2013.3

Li, G., Zhou, H., Mao, J., & Chen, S. (2019). Classifying Social Media Users with Machine Learning. *Data Analysis and Knowledge Discovery*, *3*(8): 1-9. https://doi.org/10.11925/infotech.2096-3467.2018.1207

Lundgren, L., Crippen, K.J. & Bex, II, R.T. (2018). Digging into the PIT: A new tool for characterizing the social paleontological community. In *Proceedings of E-Learn: World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education* (pp. 121-128). Las Vegas, NV, United States: Association for the Advancement of Computing in Education (AACE). Retrieved August 10, 2022 from https://www.learntechlib.org/primary/p/184954/.

Lundgren, L., Crippen, K.J. & Bex, R.T. (2022). Lundgren, L., Crippen, K. J., & Bex, R. T. (2022). Social media interaction as informal science learning: a comparison of message design in two niches. *Research in Science Education*, *52*(1), 1–20. https://doi.org/10.1007/s11165-019-09911-y

Malik, A., Heyman-Schrum, C., & Johri, A. (2019). Use of Twitter across educational settings: a review of the literature. *International Journal of Educational Technology in Higher Education, 16*(1), 36. https://doi.org/10.1186/s41239-019-0166-x

Milani, E., Weitkamp, E., & Webb, P. (2020). The visual vaccine debate on twitter: A social network analysis. *Media and Communication*, *8*(2), 364–375. https://doi.org/10.17645/mac.v8i2.2847

Min, S., & Wohn, D. (2020). Underneath the filter bubble: The role of weak ties and network cultural diversity in cross-cutting exposure to disagreements on social media. *The Journal of Social Media in Society*, *9*(1), 22–38.

Min, W., Jin, D. Y., & Han, B. (2018). Transcultural fandom of the Korean Wave in Latin America: through the lens of cultural intimacy and affinity space. *Media, Culture & Society*, 016344371879940. https://doi.org/10.1177/0163443718799403

Minnich, A., Chavoshi, N., Koutra, D., & Mueen, A. (2017). Botwalk: efficient adaptive exploration of twitter bot networks. In J. Diesner, E. Ferrari, & G. Xu (Eds.), *Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2017* (pp. 467–474). ACM. https://doi.org/10.1145/3110025.3110163

Moukarzel, S., Rehm, M., Del Fresno, M., & Daly, A. J. (2020). Diffusing science through social networks: The case of breastfeeding communication on Twitter. *Plos One, 15*(8), e0237471. https://doi.org/10.1371/journal.pone.0237471

Neely, A. D., & Marone, V. (2016). Learning in parking lots: Affinity spaces as a framework for understanding knowledge construction in informal settings. *Learning, Culture and Social Interaction, 11*, 58–65. https://doi.org/10.1016/j.lcsi.2016.05.002

Patton, M. Q. (2002). *Qualitative research and evaluation methods* (3rd ed.). Sage.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*.

Pour-Khorshid, F. (2018). Cultivating sacred spaces: a racial affinity group approach to support critical educators of color. *Teaching Education*, *29*(4), 318–329. https://doi.org/10.1080/10476210.2018.1512092

Rosenberg, J. M., Greenhalgh, S. P., Koehler, M. J., Hamilton, E. R., & Akcaoglu, M. (2016). An investigation of State Educational Twitter Hashtags (SETHs) as affinity spaces. *E-Learning and Digital Media*, *13*(1–2), 24–44. https://doi.org/10.1177/2042753016672351

Shafirova, L., Cassany, D., & Bach, C. (2020). From “newbie” to professional: Identity building and literacies in an online affinity space. *Learning, Culture and Social Interaction*, *24*, 100370. https://doi.org/10.1016/j.lcsi.2019.100370

Sharma, P., & Land, S. (2018). Patterns of knowledge sharing in an online affinity space for diabetes. *Educational Technology Research and Development, 67*(2), 1–29. https://doi.org/10.1007/s11423-018-9609-7

Smith, M., Ceni, A., Milic-Frayling, N., Shneiderman, B., Mendes Rodrigues, E., Leskovec, J., & Dunne, C. (2010). *NodeXL: a free and open network overview, discovery and* *exploration add-in for Excel 2007/2010/2013/2016.* Social Media Research Foundation. https://www.smrfoundation.org/

Smith, M., Rainie, L., Shneiderman, B., & Himelboim, I. (2014): Mapping Twitter Topic Networks: From Polarized Crowds to Community Clusters. PEW Research Report: https://www.pewinternet.org/2014/02/20/mapping-twitter-topic-networks-from-polarized-crowds-to-community-clusters/

Staudt Willet, K. B. (2019). Revisiting how and why educators use twitter: tweet types and purposes in #edchat. *Journal of Research on Technology in Education, 51*(3), 273–289. https://doi.org/10.1080/15391523.2019.1611507

Staudt Willet, K. B., & Carpenter, J. P. (2020). A tale of two subreddits: Change and continuity in teaching-related online spaces. *British Journal of Educational Technology: Journal of the Council for Educational Technology.* https://doi.org/10.1111/bjet.13051

Staudt Willet, K. B., & Carpenter, J. P. (2020). Teachers on Reddit? Exploring contributions and interactions in four teaching-related subreddits. *Journal of Research on Technology in Education, 52*(2), 216–233. https://doi.org/10.1080/15391523.2020.1722978

Su, L. Y., Scheufele, D. A., Bell, L., Brossard, D., & Xenos, M. A. (2017). Information-sharing and community-building: Exploring the use of Twitter in science public relations. *Science Communication, 39*(5), 569–597. https://doi.org/10.1177/1075547017734226

Toupin, R., Millerand, F., & Larivière, V. (2022). Who tweets climate change papers? investigating publics of research through users' descriptions*. Plos One, 17*(6), e0268999. https://doi.org/10.1371/journal.pone.0268999

Walter, S., Lörcher, I., & Brüggemann, M. (2019). Scientific networks on Twitter: Analyzing scientists' interactions in the climate change debate. *Public Understanding of Science (Bristol, England), 28*(6), 696–712. https://doi.org/10.1177/0963662519844131