SCHOLARS' INTERACTIONS ON TWITTER: A CASE STUDY ON OCEAN RESEARCHERS' CONFERENCE PARTICIPATION

by

Keshava Pallavi Gone

Submitted in partial fulfillment of the requirements
for the degree of Master of Information

at

Dalhousie University
Halifax, Nova Scotia
March 2021

Table of Contents

List of Tables

List of Figures

Abstract

Twitter is a robust tool with which to analyze conference participation for increasing visibility, measuring public engagement, and other scholarly activities. This paper offers an overview of conference participation by implementing natural language processing techniques to generate sentiment analysis and social network analysis. The study assists conference organizers and researchers by providing helpful indicators that measure user engagement and social relationships on Twitter. To accomplish this, we examined a sample of 8,715 tweets from eight prestigious ocean conferences and participating researchers who attended them in the year 2018 to 2020. We used VADER to perform sentiment analysis and Gephi network tool to generate social interconnectivity. The results revealed the amount of user engagement as well as the sentiment of each conference in addition to researcher and generated networks that represented their social relationships. VADER further exposed the most frequent topics referenced and hashtags used in Twitter interactions for a conference.

List of Abbreviations and Symbols Used

AFINN	Affective Norms for English Words sentiment lexicon
AERA	American Educational Research Association
API	Application Programming Interface
CMOS	Canada Meteorological and Oceanographic Society
CZC	Coastal Zone Canada
ICEM	International conference of emergency medicine
NAIA	Newfoundland Aquaculture Industry Association
NLTK	Natural Language Tool Kit
ORG	Organization
POS	Part-of-speech
SCMO	Société canadienne de météorologie et d'océanographie
UNESCO	United Nations Educational, Scientific and Cultural Organization
VADER	Valence Aware Dictionary and sEntiment Reasoner
#	Hashtag (reference a quote/topic)
@	At (address mention/person)

Acknowledgements

I would like to thank Pisces Research Project Management Inc. for providing me with the inspiration for my thesis on ocean research conferences. I am thankful to my supervisors, Dr. Mike Smit and Dr. Philippe Mongeon, who have given their unfailing support for this thesis. I am extremely grateful for their helpful comments, suggestions, and continuous guidance throughout the project. I would like to extend my sincere gratitude to Dr. Mike Smit for providing the research funding for this thesis. Lastly, I thank my family for their constant love and support. Especially, my husband Raghav for understanding and giving me hisemotional support.

Chapter 1

Introduction

1.1 Background

International conferences provide a forum for education, research, collaboration, and development of the specialty concerned. Participation at such conferences is encouraged and credited as part of continuing professional development for researchers (Neill et al., 2014). The role of social media technologies, like Twitter, are playing an important role in this type of scholarly communication (Desai et al., 2012; Shiffman, 2012). Academics and conference organizers have another venue by which to increase their visibility and to improve their research and other scholarly activities (Al-Daihani et al., 2018; Sugimoto et al., 2017). In addition to the academic content provided there are multiple indicators by which funding agencies and policymakers can measure public outreach (Mastely, 2017).

As stated by McNamara (2013), nowadays conference-related conversations are no longer restricted to the conference venue, they are available on public domains. Twitter makes it convenient to set up this interaction up as most conference organizers use an unique hashtag in which researcher, conference organizers, and audiences can discuss a particular conference (Webb, 2016). The scholarly content generated through this conversation can lead to the conference participant's use of Twitter (Salzmann-Erikson & Martin, 2018). Despite a broad focus on conference participation in the literature, a Twitter users' social network use remains largely unexplored.

Traditional methods of measuring conference participation were mostly implemented either via content analysis or network structure analysis. Few studies analyzed the sentiment analysis approach. Moreover, previous research studies on conference network analysis focused on analyzing levels of participation and communication networks, but few empirical studies focussed on social relationships. Furthermore, there is no evidence of studies that have sought to explore ocean conferences for analyzing Twitter activity. To address this gap, the present research will develop an overview of conference participation by implementing sentiment analysis and social network analysis to examine public engagement and social networks using ocean conference related tweets.

The study aimed to provide an overview on analyzing conference participation based on scholarly content on Twitter. We implemented NLP techniques on the tweets related to eight prestigious ocean conferences and their participants from 2018 to 2020. This work generated multiple insights into conference participation for organizers and researchers.

1.2 Objective and Purpose of the Study

The overall objective of the research is to investigate how ocean scholars disseminate helpful content via Twitter in the conference context. In particular, the study has three main objectives:

1. To determine the amount of user engagement of a conference.
2. To determine the sentiment of conference tweets.
3. To determine the interactions between Twitter users.

The result of this study will be valuable to conference organizers, researchers, funding agencies, and relevant practitioners for developing tools for analyzing conference tweets and scholars' research contributions.

1.3 Organization of Thesis

This chapter has presented a context for ocean researchers' conference participation on Twitter. The rationale for the research, objective, purpose of the study has also been presented. Chapter two is a review of the relevant literature on scholars' interactions on Twitter, with an examination of various analyses using scholar's tweets and metrics for measuring conference participation on Twitter. Chapter three will discuss the proposed methodology used to conduct the research. Chapter four will present the findings and results. Chapter five will detail the discussion and reflections of the researcher. Finally, chapter six will summarize the finding's conclusions and will suggest implications for the study's results. This paper will end with recommendations for further study.

Chapter 2

Literature Review

The main aim of the study is to provide evidence of an added dimension to scholar's conference experience originating from the scholarly content found on Twitter's platform. I first describe the role of conferences and, specifically, ocean conferences. Next, I review the literature on social media, scholarly communication, and outreach. Lastly, I outline literature related to metrics for conference participation.

2.1 Research Conferences

2.1.1 Role of Conferences in Research

An academic conference allows scientists to analyze and solve various issues regarding their research and to exchange knowledge and opinions relevant to their scholarly pursuits (Dimitrios et al., 2014). It is an important venue for the presenters for brainstorming, networking, and making vital connections that can lead to new initiatives, papers, and funding (Oester et al., 2017). One of the key important factors for participating in conferences is publication. The conference proceedings are a great way to have researchers' work published and indexed (Franceschet, 2010). There are several other benefits for researchers to present at a conference such as: meeting individuals with the same interests and similar goals; exposure to the latest research studies in the field; providing an opportunity to exchange ideas on the similar fields of interests; and setting the foundation for future collaborations globally (Franceschet, 2010, Oester et al., 2017). Additionally, participation at conferences is encouraged and credited as part of continuing professional development for researchers (Mastley, 2017).

2.1.2 Ocean Research Conferences

International conferences have a key role in ocean research. Researchers and organizers put a lot of effort to use these conferences to inform about issues of concern and to promote ocean protection (Lubchenco & Grorud-Colvert, 2015). These conferences mostly focus on core ocean issues: ocean sustainability; climate change; marine science; fisheries; and maritime security (; Lubchenco, & Grorud-Colvert, 2015; Neumann & Unger, 2019;). A conference called Save Our Ocean, in 2017, was organized by the United Nations to raise awareness and elicit commitments to implement Sustainable Development Goal 14 (Hongbo 2017; Neumann & Unger, 2019). Along the same line, another conference, hosted by the European Union in 2017, made contributions to maritime security and the sustainable development of the blue economy worldwide (Our Ocean Conference 2017: Final report, n.d.). This conference generated over EUR $7 billion in commitments on oceans (Our Ocean Conference 2017: Final report, n.d). The ocean conferences are always a major platform to promote, protect, and address ocean-related issues and developments (Lubchenco & Grorud-Colvert, 2015; Hongbo 2017; Neumann & Unger, 2019).

2.2 Social Media, Scholarly Communication, and Outreach

Social media are widely used in the current educational context of emerging technologies, and they significantly influence academics' scholarly communications (Al-Daihani et al., 2018). Research in this area has accelerated due to several opportunities for promoting new perfomance indicators based on social media activity (Sugimoto et al., 2017). Most of the past literature on scholarly communication and science outreach on social media platforms has focussed on four types of activity: visibility of the research and configuring a bigger network; spreading information; social media as a backchannel for conferences; and measuring research impact.

2.2.1 Visibility and Impact of Research

Social media technology can be a powerful tool for education and outreach for scientific conferences (Shiffman, 2012). Al-Daihani et al. (2018) investigated improving academics' visibility by implementing quantitative and qualitative methods as well as conducting questionnaire surveys of academics and interviews to provide baseline data and detailed insights into academics' use of social media. The results suggested social media websites have a significant potential for enhancing academic visibility and access to information. Another study by López-Goñi and Sánchez-Angulo (2017) reviewed the use of Twitter in social networks, they analyzed a microbiology course that was delivered from different research centres in Spain and Latin America and involved highly dynamic, interactive, and accessible social learning. They noted that social media can be a dynamic display for research and publications to enhance the quality or visibility of science. A paper discussed by Shiffman (2012) used Twitter to share important conversation information with an interested public audience by promoting live tweeting of the scientific conference i.e., the 2011 International Congress for Conservation Biology. The online outreach conducted at the conference was successful in reaching an interested audience on Twitter. Further, Twitter has gained popularity among scholars and researchers for academic communication and has the potential to provide a new method for measuring research impact (Holmberg & Thelwall, 2014).

Altmetrics are popular indicators used by researchers to track their work, thereby facilitating the advancement in evaluating the social impact of their studies (Pulido et al., 2018; Sugimoto et al., 2017). The indicators allow quantitative measuring of the impact of research based on Twitter data, such as counts of tweets and retweets, Twitter mentions, including direct or indirect links from a tweet (Bornmann et al., 2016). Article-level Web activity captured by altimetric provides data to measure the correlation between citations, sentiment analysis, citation metrics, and other research impact measurements of the articles (Bornmann et al., 2016; Hassan et al., 2020). A study by Tattersall and Carroll (2018) used altimetrics for identifying and categorizing the policy impact of research articles from a single centre. The study only included published research articles from authors and used altimetrics to count citations. The results offered highly accessible data on the policy impact of an organization's published research articles. Another study by Hassan et al. (2020) applied altimetric techniques to measure the early impact of research articles through the sentiments expressed in tweets, which revealed a strong correlation between positive sentiments and citations.

2.2.2 Information Sharing

Social media has changed the way scholars communicate and share information. Social networks have already been exploited for learning, sharing, and demonstrating efficient vehicles for social learning (López-Goñi & Sánchez-Angulo, 2017). Luzón et al. (2020) reviewed an academic conference that used social media to: provide information about the conference and engage discussion by attendees about the conference topics; share information; and create social links and networks within the community. Their study investigated the conversations originating from two conferences in applied linguistics to analyze the network resources of scholars using social media during conferences. A conference conducted by healthcare providers used social media as a backchannel to inform and educate the public about kidney diseases. During Kidney Week 2011, Twitter is used as a communication channel to share information about the diseases and to enhance awareness. They reviewed the tweets acumulated in the official timeline during the conference week used to disseminate information to the general public (Desai et al., 2012).

Moreover, a survey study by Mastley (2017) presented a correlation between research information seeking and sharing on social media. They used the Web of Science, Scopus, and Google scholar to examine social media and information behaviour from 2008 to 2015. The study implemented citation analysis to analyze the impact of the selected papers and resulted in high rates of impact concerning information seeking and information sharing studies in the field. In this regard, Twitter can not only be used for sharing information but it can also be a communication channel to provide educational information to a broader audience.

2.2.3 Twitter as a Backchannel for Conference

Twitter is one of the more convenient social media platforms to spread information and to communicate among conference participants (Lee et al., 2017; Li & Greenhow, 2015; Xie & Luo, 2019). Wen et al. (2014) investigated the use of Twitter as a backchannel within the computer science community by studying sixteen conferences over a timespan of five years, they found an increasing trend of informational usage and more conversation than the information network. A similar study by Li & Greenhow (2015) used Twitter as a conference backchannel to explore how educational scholars incorporated social media.

Data generated by the members of AERA (American Educational Research Association) were examined from their previous conference tweet —stream. It revealed that the scholars not only seek social media to disseminate their work, but to also engage and mentor within the conference community. Another study by Xie & Luo (2019) examined the conference learning community on Twitter by reviewing user participating patterns and network structure in Twitter-enabled backchannels and communication network. The results revealed users' varying levels of participation and a relatively low network density.

Use of Twitter affordances in conference settings

Twitter has become the most popular tool in conference settings (Desai et al., 2012; Lee et al., 2017; Li & Greenhow, 2015). It is used in a varying manner in different stages of the conferences as organizers use Twitter for communication, promotion, and as a backchannel for conferences (Li & Greenhow, 2015). Researchers use Twitter to interact, connect, and share their research to broader audiences (Desai et al., 2012; Lee et al., 2017;). Attendees use Twitter to give feedback on presentations and to share within their networks (Shiffman, 2012). Twitter incorporates certain affordances such as hashtag (#), mentions (@) and retweets to facilitate these services to its users.

Hashtags are indicated by # and are used to facilitate conversation based on a particular content (Li & Greenhow, 2015). Most of the conferences incorporate a dedicated hashtag (#), that labels and categorizes tweets and enables social networks (Zappavigna, 2012). The unique conference hashtag will be used by different users to analyze several studies. Neill et al. (2014) proposed a study examining the demographic analysis of tweets produced regarding a major international conference emergency medicine (ICEM) using #ICEM

2012. The content in every tweet related to #ICEM2012 was reviewed and classified into multiple categories by participants and was analyzed by Neill et al. (2014). Another study by Bert et al. (2016) investigated a sample of 1,066 tweets with the hashtag epglasgow (#epglasgow) during the 7th European Public Health (EPH) conference to promote online discussion and knowledge dissemination on public health. Additionally, conference live streaming enables participants to share information. An experiment by Shiffman (2012) in online outreach conducted in 2011 analyzed #ICCB (International Congress for Conservation Biology) tweets. They were judged to be a huge success and Shiffman (2012) noted that Twitter is a robust tool for public education and outreach from scientific conferences.

Mentions allow tweet posters to address specific users in the interaction (Zappavigna, 2012). This feature is used by including @username in a tweet to communicate or integrate another user (Luzón et al., 2020). As Luzón et al. (2020) noted, the tweet on the conference often included the name of the presenter in the form of @username attached to the topic with a sequencing expression (eg. 'kicking off', 'just started','in the final session'). A study by Lee et al. (2017) on the conference Twitter network included mention and retweet interactions, where mentions exposed how scholars converse with each other on Twitter.

Retweets are used by users to share someone else's tweet wherein a user's followers can view regardless whether or not they follow (Shiffman, 2012). Retweeting is a common practice in the conference community and provides a plethora of information (Zappavigna, 2012). A study by Wen et al. (2014) derived the user-user network of retweets for a conference for many years with a directed link from one node to another node. The study also noted that retweets played an important role in disseminating information within and outside the conference. They calculated the retweet ratio to measure the proportion of tweets being shared in the conference.

All of these attributes contribute to the conversational purpose of the tweets. They will be analyzed to understand the interactions between conference participants, contents, and organizers.

2.3 Twitter Metrics for Conference Participation

Many studies have used several approaches to analyze conference interaction among scholars and organizers. This section focuses on an overview of currently used Twitter metrics for conference participation.

2.3.1 Content Analysis

A study by Neill et al. (2014) performed content analysis and categorized the tweets into five definitions: session related; social; logistic; advertising; and other analyzing the number of tweets, replies, and country of origin of the individual. They identified 212 speakers through the organizing committee and examined over 4,500 tweets about ICEM 2012. They further examined twitter use by attendees and non-attendees using an online archiving system. The result exposed significant use of the Twitter platform at ICEM 2012 where 34% of attendees were physically present at the conference and 74.4% of the tweets directly related to research findings and teaching at the conference.

A study by Desai et al. (2012) on the Kidney Week 2011 conference performed a content analysis using tweets. The researchers classified each tweet as informative or uninformative, using an industry-standard classification system. Results showed a highly correlated number of sessions and posters with the number of informative tweets. Another study, by Bert et al. (2016), reviewed Twitter content during the 7th European Health Conference. The study retrieved tweets including before and after the conference period that were analyzed using StataMp11 (a statistical software package). A total of 1,067 tweets were retrieved where 86.3% of the tweets were tweeted during the conference. Of those tweets 29.7% were images, 13.8% were external links, and conference speakers were mentioned in 30% of the tweets. The number of conference session's related tweets demonstrated that Twitter's use during conferences is a growing phenomenon for public health professionals.

2.3.2 Network Analysis

Network analysis is the study of social entities, their interactions, and relationships (Varlamis et al., 2010). It can be utilized to understanding the significance of each node in an information flow of networks (Hajian & White, 2011). There are several metrics that can be used to measure the role of each node such as its density, degree of centrality, closeness and betweenness (Hajian & White, 2011; Lemay et al., 2019; Varlamis et al., 2010). A study by Lemay et al. (2019) investigated social network analysis during the American Educational Research Association's 2017 annual conference to quantify the degree of network centrality and closeness of top users, and the density of the network. They calculated the statistics of the social graph of top 100 users retrieved through NetworkX, a Python package. The degree of centrality, closeness, and density are calculated using the number of nodes and the distance between the nodes. Degree of network centrality address different aspects to quantify the strength of specific connections within the network (Lemay et al., 2019). The results of Lemay et al. (2019) study revealed that the social network is sparse characterized by low density, with fewer active nodes and many unconnected users, and the degree of network centrality and closeness of top 10 users is high, relative to the top 100 users.

Xie and Luo (2019) performed in-depth social network statistical path analysis to examine the primary characteristics of the central users. They implemented a NodeXL approach – NodeXL facilitate social network analysis with network visualizations (Hansen et al., 2011). The approach divided the networks into several densely interconnected, but separate subgroups. The network revealed that users' have three varying levels of participation: interaction initiator; opinion leader; and conversation bridge among the variables. Another study by Wen et al. (2014) examined Twitter use in 16 Computer Science conferences over a span of five years by modelling conversation networks and retweet networks from interactions (hashtags, replies, mentions, and retweets) and use network properties to measure how people interacted over those years on Twitter. The results showed more participation over time, with the information flow (retweets) as the main component in their conversations.

Calma and Davies (2014) used citation network analysis of publications in *Studies in Higher Education* from 1976 to 2013 inclusive. The Gephi tool, an open-source network visualization software used for generating graph and network analysis (Bastian et al., 2009), was used to visualize associated citation networks. The results revealed the five most published authors, the most cited authors, and the most discussed topics with a number of subordinate topic clusters amongst them.One of Gephi's key features is the ability to display the spatialization process and to transform the network into a graph (Jacomy et al., 2014). Creating network graphs on the Gephi platform provides real-time visualization of the network and helps to customize the look of the network graph (Cherven, 2013). Additionally, it provides easy and broad access to network data and it allows for spatializing, filtering, navigating, manipulating, and clustering of data (Bastian et al., 2009).

2.3.3 Sentiment Analysis

Sentiment analysis is an opinion-mining technique that mainly focuses on the classification of sentiment polarity, i.e., positive or negative, and is typically detected by a machine-learning approach (Kim, 2016). It aims to analyze users' sentiments, attitudes, opinions, and emotions towards elements such as products, individuals, topics, organizations, and services (Kharde & Sonawane, 2016). Multiple machine learning algorithms are introduced to analyze sentiments such as: naive Bayes; Maximum Entropy Classification; Support Vector Machines; Artificial Neural Network; Decision Tree; K-Nearest Neighbor; and Ensemble Learning as sentiment polarity detection can be seen as a classification task (Han et al., 2018).

Desai et al. (2012) calculated sentiment scores for kidney disease outreach by performing conventional linguistic analyses of each tweet using modified AFINN (Affective Norms for English Words sentiment lexicon). Each word in the lexicon is given a value from -5 (highly negative) to +5 (highly positive). The results revealed that the mean sentiment score for all tweets was 0.094 (SD 0.476; range 21.70, 2.67) with the opinion tweets scoring highest mean score (0.454) amongst all types of tweets.

Alternately, the VADER lexicon performs exceptionally well in calculating sentiment in the social media domain (Hutto & Gilbert, 2014). It is based on a dictionary of a set of words with positive, negative, or neutral sentiment scores. The study by Hutto and Gilbert (2014) showed that the VADER ($r = 0.881$) correlation coefficient performed well with individual human raters ($r = 0.888$) at matching a ground truth (information obtained by direct observation). Further, the study showed the VADER sentiment performed better or equally well when compared against seven other sentiment analysis lexicons.

The brief overview of the literature concerning conference participation of researchers reveatd that many studies analyzed Twitter use by organizers, attendees, and researchers using conference-related tweets. Despite these previous reports, to date, no studies have described the use of Twitter in ocean-related conferences. For this reason, I have decided to perform an analysis of Twitter use over eight prestigious ocean conferences. Traditional methods of measuring conference participation of researchers were conducted through content analysis and network structure. Only one study, implemented by Desai et al. 2012, calculated the sentiment score for kidney disease outreach on Twitter during the Kidney Week 2011 conference by performing AFINN.

The present study seeks to help to fill in those knowledge gaps. This paper explores how Twitter can be helpful to scholars, organizers, and funding agencies. First, we determine the amount of user engagement using statistical analysis. Later on, in the process sentiment analysis is implemented. Based on a study by Hutto and Gilbert (2014), the sentiment of conference tweets in this paper is calculated using VADER. The metric used to calculate the compound polarity score is the sum of sentiment scores of each word in the lexicon, adjusted according to VADER's rules, often between -1 (most extreme negative) and +1 (most extreme positive) (Borg & Boldt, 2020). Last, we perform network analysis using the Gephi tool. As presented by Bastian et al. (2009) on several key features of Gephi, in the context of interactive exploration and interpretation of networks, this study generates network graphs using the Gephi tool to measure ocean conference participation on Twitter.

Chapter 3

Methodology

This section addresses applied technologies regarding the research question. Firstly, I focused on the process of data collection, and secondly, the applied methodologies are addressed.

3.1 Data Collection

The first step in data collection was to retrieve the tweets related to eight conferences from 2018 to 2020. The tweets on each conference promoters' timeline made by conference organizers were extracted. The twitter handles of the eight conferences committees assessed were: @CZCAssociation for Coastal Zone Canada; @ArcticNet for ArcticNet; @CMOS_SCMO for CMOS; @OceanTechNS for H2O; @OceanObs19 for Ocean Obs; @NAIA_NL for Cold Harvest; @OCEANS_Conf for Oceans; and @OceanologyIntl for Oceanology. I used the Python programming language to access Twitter API (Application Programming Interface) and collected 8,715 tweets and stored them in a PostgreSQL database.

The dataset also contained the tweets that used hashtags (#). Most of the conference organizers have registered a unique hashtag (#) for each year conference such as (#Oceans2020, #CMOS2018) to encourage people to include the hashtag in their tweet related to the specific conference (Abascal-Mena et al., 2015). Although I was unable to search for older hashtag tweets via Twitter API (Twitter API limits hashtag data access to last seven days), I acquired them from the users' tweets. Tweets extracted from the users' accounts only discussed: a conference; specific session/talk; or a speaker. The steps I followed to fetch these tweets were:

a) Reviewing the user's account whose tweets included hashtags manually from Twitter website and collected unique Twitter account holders.

b) Using Twitter API to extract the tweets from each user accounts.

c) Filtering the tweets that contained conference hashtags and stored in the database.

The developed dataset contained all the tweets related to eight conferences including date, time, number of retweets, and likes that further assisted me to analyze the dissemination of the conferences. Table 3.1 presents the total number of tweets collected for each conference from January 1, 2018 to December 31, 2020.

Table 3.1: Total Number of Tweets Collected for Each Conference's Tweets

Conference	**Twitter Source**	**Number of Tweets**
OCEANS	@Oceans_Conf #OceansConference #Oceans2020SG #Oceans2019 #Oceans2018 #OCEANS20GulfCoast #OCEANS20Singapore	1,192
ArcticNet	@ArcticNet #AC2020 #AC2019 #AC2018	1,773
CMOS_SCMO	@CMOS-SCMO #CMOS2020 #CMOS2019 #CMOS2018	303
Coastal zone Canada	@CZCAssociation #CZC2020 #CZC2019 #CZC2018	756

Conference	Twitter Source	Number of Tweets
Oceanology	@OceanologyIntl #oi2020 #oi2019 #oi2018	2,368
Cold Harvest	@NAIA_NL #coldharvest #coldharvest2019 #coldharvest2018	1,650
H2O	@OceanTechNS #H2020 #H2019 #H2oConference	177
Oceans Obs	@OceanObs19 #OceanObs19 #OSM2020 #OSM19 #OSM18	495
Total number of Tweets		8,715

The next step of the data collection was to obtain a list of speakers' names and Twitter accounts (@) from the tweets. First, I extracted researchers' names from the tweets by implementing NLP techniques such as tokenization, POS (Part-of-speech) tagging, and chunking. A Python module, human name parser, was also used for parsing a list of human names from the tweets. The parser reviewed each tweet for the researcher's name (First Name and Last Name). Then the tweets containing the researcher's name were extracted and were stored in the new column of the table. Table 3.2 shows an instance of the researcher data table. A total of 410 researcher names were extracted from eight conference tweets. The researcher names retrieved from these tweets are used for the analysis to determine the user engagement and measure conference participation.

The steps I followed to retrieve the researcher's name were:

1) Each tweet's sentences were extracted from the database.

2) Tokenization was implemented to break the sentences into meaningful tokens.

3) Chunking was used to parse the grammar of each token to find a proper noun.

4) A Python list called NLTK (Natural Language Tool Kit) subtree (Bird et al., 2009) was used to list the human names.

5) Extracted lists of human names were stored in a new column of the data table.

Secondly, I scraped user accounts from the tweets using a built-in package regex (regular expression). Regex helped to create a string pattern to locate a specific text, the tweet contained user accounts were extracted and stored in the new column of the data table as shown in Table 3.2. A total 701 user accounts were scraped from conference tweets.

The steps I followed to scrape user accounts in the tweets were:

1) Imported re built-in package.

2) Implemented a function to create a text pattern (a string starting with @).

3) Used the re.findall operation to scan each tweet from left to right to find and return the text pattern.

4) The extracted string from each tweet was stored in a new column of the data table.

The data collected is stored in a PostgreSQL database, an open source database, for further analysis. The database-schema of the study is presented in the Appendix. The database structure shows seven database tables and relationship between them. The schema diagram presents datatype for each table attribute, along with primary keys and foreign keys.

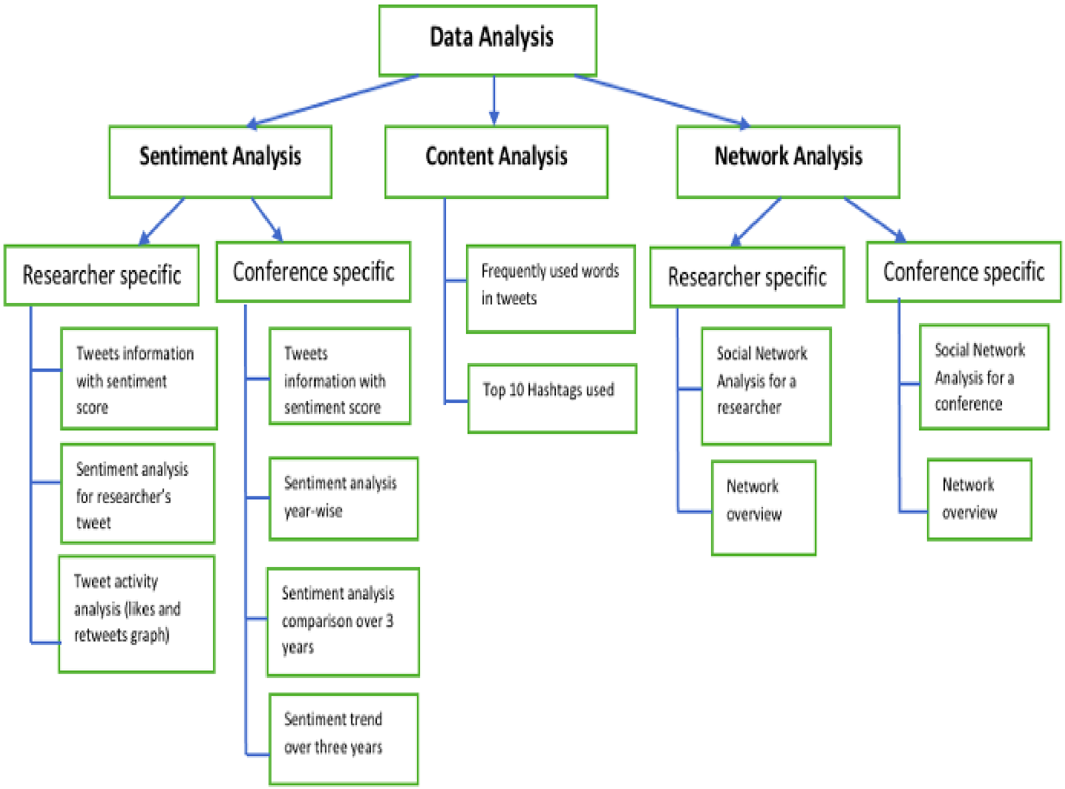
Table 3.2: An Instance of Researcher Data Table

Conference	Post	Researcher Name	Twitter Account
ArcticNet	Chris Derksen Research Scientist introducing Climate Change Special Report https XCcJSSoR	Chris Derksen	
ArcticNet	@aliciadobson She’s been posting lots of great stuff today! #ArcticChange2017		@aliciadobson
ArcticNet	@Nforsk @Norduniversitet @thoralffagertun Thank you for presenting! #ArcticChange2017		@Nforsk @Norduniversitet @thoralffagertun
ArcticNet	Muhammad Arshad Khalafzai explains poster Spring Flooding Recurring Evacuations Saskatchewan	Arshad Khalafzai	

3.2 Data Analysis

I conducted multiple analyses of the tweets to determine user engagement with the speakers and the conference. First, I implemented a sentiment analysis on researcher-specific tweets and conference-specific tweets. Second, I applied frequency distribution to retrieve frequently used words and hashtags to analyze user engagement for each conference. Finally, I performed a social network analysis to evaluate and monitor conferences’ and each researcher’s conference connectivity network. Figure 3.1 shows the hierarchical flow of methods implemented for the research project.

Figure 3.1: Hierarchical map for applied methodologies



3.2.1 Sentiment Analysis

I implemented sentiment analysis on the dataset, using VADER, which is based on a dictionary of a set of words with positive or negative sentiment scores. VADER calculates the sentiment score of a comment by summing up the sentiment scores of each word in the dictionary (Andy, 2018). The sentiment score of the tweet is counted as positive if the words in the tweet match with positive emotion in the dictionary. It is counted as negative when they match with negative emotion in the dictionary.

To compute sentiment analysis with VADER, it computes the compound score by summing the valence scores of each word in the lexicon, adjusted according to the rules, and then normalized to be between -1 (most extreme negative), +1 (most extreme positive) and 0 (neutral) (Hutto & Gilbert, 2014). A tweet with a compound score greater than 0.05 was classified as positive, a tweet with a score between -0.05 and 0.05 was classified as neutral, and a tweet with a score less than -0.05 was classified as negative (Chandrasekaran et al., 2020).

Researcher-specific sentiment analysis

Automated sentiment analysis is performed for the researcher by calculating the compound score of their tweets. Positive, negative, and neutral sentiment scores provide multidimensional measures of the sentiment and public opinion on the speaker's research and conference presentation. Additionally, the number of likes and retweets of each tweet is analyzed to provide more insights into the researcher's conference contribution. The information gathered from the analysis assist researchers and funding agencies to understand public outreach of scientific research on Twitter.

Using Python, I have developed a program that is able to:

1. Retrieve the tweets of a given researcher from the database.
2. Clean tweets using RegEx to eliminate uninformative text such as symbols, emojis, and hyperlinks.
3. Calculate sentiment compound score of each tweet.
4. Categorize the type of sentiment (positive, negative, or neutral).
5. Provide a number of positive, negative, and neutral sentiment tweets of a researcher.
6. Visualize a sentiment graph.

We retrieved and analyzed the researcher's tweets assuming that the Twitter user is aware that their public tweets could be used by other researchers without their consent. Table 3.3 shows the sentiment polarity score and sentiment type of each tweet of the researcher Steve

Gain. Each tweet is given a compound polarity score that is either positive, negative, and neutral sentiment. The count of each sentiment type of a researcher tweet is presented in Table 3.4.

Table 3.3: Data Representing Sentiment Polarity of Steve Gain's Tweets

Post	**Sentiment_compound_polarity**	**Sentiment_neutral**	**Sentiment_negative**	**Sentiment_positive**	**Sentiment_type**
Making the case for aquaculture — Dr. Steve Gaines, University of California	0	1	0	0	Neutral
Thank you for inviting Dr. Steve Gaines to outline the Future of Food and why farming our oc…	0.5859	0.625	0	0.375	Positive
Fish farming has the least average impact per pound of protein compared to all other species — Dr. Steve Gaines…	0	1	0	0	Neutral
We are so pleased to have Dr. Steve Gaines, Dean of Marine Science, Sustainable Fisheries, University of Californi… https://t.co/wtuYiv7onJ	0.4404	0.805	0	0.195	Positive

Table 3.4: Sentiment Count of Tweets

Sentiment Type	**Number of Tweets**
Neutral	14
Positive	7
Negative	0
Total number of tweets	21

Conference-specific sentiment analysis

Similarly, sentiment analysis is conducted on eight different conference tweets separately to analyze public engagement for each conference. Every conference sentiment is measured per year and measured three years 2018, 2019, and 2020 together to compare and analyze public outreach of a conference over three years.

Additionally, each conference tweets stored in the database are mined for frequent usage of words. Using NLP techniques such as tokenization and frequency distribution, I generated a word cloud for visual analysis of the most frequently used words by people in conference tweets.

To generate a word cloud for specific tweets, I have developed a program that is able to:

1. Split the tweets into meaningful tokens using a split() method. A split method seperates a string by a separator into a list of strings.
2. Eliminate non-informative and stop words via NLTK corpus library.
3. Get the frequency of each word.
4. Generate a word cloud with most frequently used words in the tweet.

Furthermore, the paper provided insights on the hashtags used by the people in the conference tweet conversation. I used frequency distribution function to record frequency of each hashtag and provided frequently used hashtags in the tweets.

To get most common hashtags, I developed a Python program that is able to:

1. Extract the words containing hashtags from the tweet.
2. Generate frequency distribution of each hashtag.
3. Get the most frequently used hashtags.

4. Visualize top 10 most commonly used hashtags of a given conference.

3.2.2 Social Network Analysis

In the proposed method, the social network is created using the Gephi visualization tool. To create researchers' social network, I captured @user accounts from the conference tweets, and to whom they directed. I created two csv files, a file with the list of the mentions called nodes, another file containing the relationship between each couple of mentions called edges. Tables 3.5, 3.6, and 3.7 presents details on nodes and edges data tables of a researcher.

Table 3.5: Nodes Data Table of a Researcher (Kes Morton)

ID	Label	Conference Name	Type
1	@kesmorton	CMOS_SCMO	Person
2	@SLGO	CZC Association	Organization
3	@coinatlantic	CZC Association	Organization
4	@CZCAssociation	CZC Association	Organization
5	@NAIA_NL	Cold Harvest	Organization
6	@piscesrpm	OceansObs	Organization
7	@OceanObs19	OceansObs	Organization
8	@CMOS_SCMO	CMOS_SCMO	Organization
9	@MEOPAR_NCE	CMOS_SCMO	Organization
10	@CMOS2018	CMOS_SCMO	Organization
11	@j_pye	CMOS_SCMO	Person
12	@AnneStMarie	CMOS_SCMO	Person
13	@MemorialU	CMOS_SCMO	Organization
14	@FACT_Network	CMOS_SCMO	Organization

Table 3.6: An Instance of Edges of a Researcher (Kes Morton)

Source	**Target**	**Weight**	**Type**	**Conference Name**
1	2	1	directed	CZC Association
1	3	1	directed	CZC Association
2	4	1	directed	CZC Association
3	4	1	directed	CZC Association
6	7	1	directed	CZC Association
1	5	1	directed	Cold Harvest
1	6	1	directed	OceansObs
1	7	3	directed	OceansObs
1	8	3	directed	CMOS_SCMO
1	9	1	directed	CMOS_SCMO

Table 3.7: More Details of Nodes and Edge's Table

ID	Unique Id is provided to each user account.
Label	User account (@) obtained from the tweets.
Conference name	Name of conference, from where the nodes are mentioned and interacted with other nodes.
User type	It specifies whether the user account (@) is of a person or an organization (I reviewed each twitter account and manually added user type information).
Source	Id of the user who tweeted or started interaction with others.
Target	Id of the user who got tweeted from others.
Type	Edge type that is either directed or undirected. In this case, it is directed because the @ mention by one user to another may not receive @ reply back.
Weight	Number of times the source mentioned the target in tweets.

The next part of the analysis is to calculate statistics of the graph. Statistical measures of the network are important as they help us to understand the meaning of the network graph (Cherven, 2013). I evaluated the following statistical measures of the graph.

Degree distribution

The term degree refers to the number of connections extending to and from a node. The term in-degree refers to the counts of the number of head endpoints adjacent to a node and the term out-degree refers to the counts of the number of tail endpoints coming from a node (Cherven, 2013).

Average Degree: It is defined as the average number of unweighted connections across a network (Cherven, 2013):

$$Avg\ Degree = \frac{sum\ of\ the\ total\ weight\ of\ edges}{number\ of\ nodes}$$

Average Weighted Degree: Is defined as the average number of weighted connections (Cherven, 2013):

$$Avg\ Weighted\ Degree = sum\ of\ weights\ of\ the\ edges\ of\ nodes$$

All the nodes in the network are associated with directed edges. Using the degree distribution, we calculated the number of connections to a respective node. The degree (unweighted and weighted) of each node is shown in Table 3.8 along with in-degree (number of nodes designated to a respective node) and out-degree values.

Graph density

It is defined as the interconnectivity of individuals in a network. The graph density may vary from low density, where a group of individuals are loosely connected, to high density, where users are highly interlinked (Hansen et al., 2011):

$$Graph\ density\ (directed\ graph) = \frac{total\ edges}{total\ possible\ edges}$$

All the nodes in the graph are not highly interlinked. Few groups of nodes have low density. Table 3.8 shows the graph density of each node of the conference.

Average shortest-path length

It is defined as the average number of steps along the shortest paths for all possible pairs of network nodes. It is a measure of the efficiency of information or mass transport on a network (Guoyong & Ning, 2013). The betweenness centrality is measured as the fraction of shortest paths going through a given node (Barthelemy, 2004):

$$Betweenness\ centrality\ of\ node\ N = \frac{\begin{array}{c} number\ of\ shortest\ paths\ from\ node\ A\ to \\ node\ B\ going\ through\ a\ given\ node \end{array}}{\begin{array}{c} total\ number\ of\ shortest\ paths\ from \\ node\ A\ to\ node\ B \end{array}}$$

The network with this centrality provided number of times a respective node is communicated between two other nodes. Table 3.8 presents an instance of values of betweenness centrality of each node. The concept is crucial to identify important nodes. This helps to analyze the frequently mentioned researcher with significant centrality.

Table 3.8: Statistical Measures of Each Node of Kes Morton's Network Connectivity

ID	Label	Organization Name	User Type	In-degree	Out-degree	Degree	Weighted In-degree	Weighted Out-degree	Weighted Degree	Betweenness Centrality
1	@kesmorton	CMOS_SCMO	person	2	10	12	2	19	21	16
2	@SLGO	CZC Association	org	2	1	3	2	1	3	0
3	@coinatlantic	CZC Association	org	1	1	2	2	1	3	0
4	@CZCAssociation	CZC Association	org	3	0	3	3	0	3	0
5	@NAIA_NL	Cold Harvest	org	2	0	2	2	0	2	0
6	@piscesrpm	OceansObs	org	1	1	2	1	1	2	0
7	@OceanObs19	OceansObs	org	2	0	2	4	0	4	0
8	@CMOS_SCMO	CMOS_SCMO	org	2	0	2	4	0	4	0
9	@MEOPAR_NCE	CMOS_SCMO	org	1	1	2	2	1	3	0
10	@CMOS2018	CMOS_SCMO	org	3	0	3	6	0	6	0
11	@j_pye	CMOS_SCMO	person	0	5	5	0	5	5	2
12	@AnneStMarie	CMOS_SCMO	person	0	2	2	0	2	2	2
13	@MemorialU	CMOS_SCMO	org	1	1	2	1	1	2	1
14	@FACT_Network	CMOS_SCMO	org	1	0	1	1	0	1	0

The network overview generated through the statistical measures provided valuable information on the researcher and conference network. It helped to analyze and understand the graph connectivity. Each value in the statistical table provided information on how actively a researcher is connected to conferences and the general public on Twitter.

A detailed theory and methods concerning the implementation of sentiment analysis and social network analysis are presented. The constructed process of the conference data collection using the two analysis assists conference organizers and researchers to measure user engagement and explore social interactions and relationships. The developed data collection model further acts as a foundation data design framework to build a software tool that measures conference participation at ocean conferences.

Chapter 4

Results

In this section, the results of the classification tasks are presented. Firstly, the amount of user engagement is presented. Secondly, the results of sentiment analysis for conference organizers and researchers are demonstrated. Lastly, results related to social interactions between Twitter users are revealed.

4.1 Amount of User Engagement

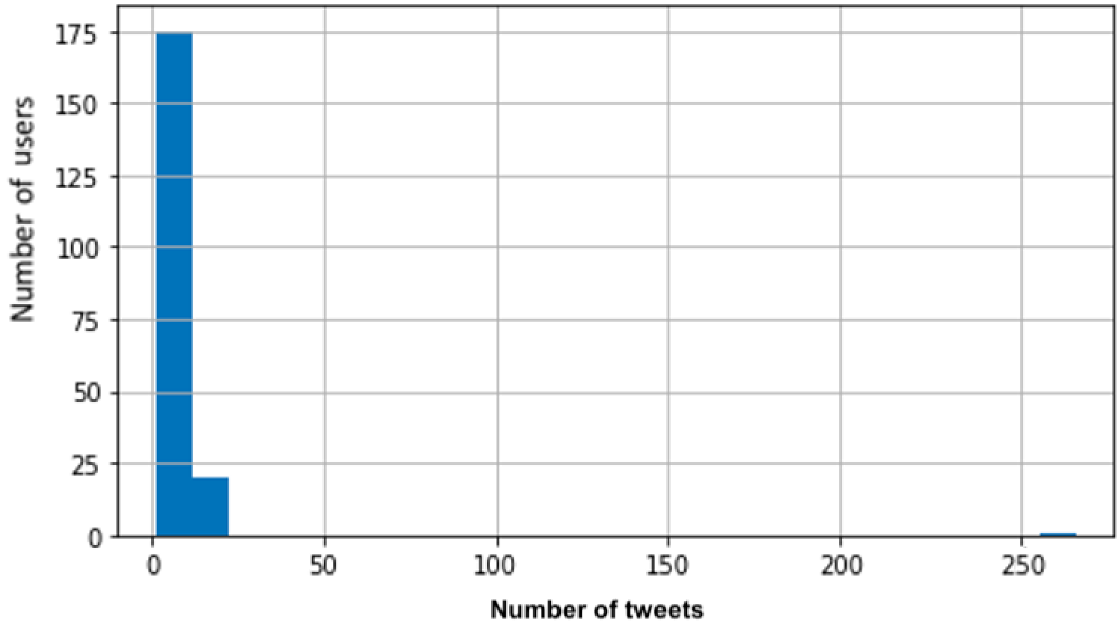
The amount of user engagement for each conference is shown in Table 4.1. It presents the total number of tweets with descriptive statistical data from 2018, January1, to 2020, December 31. The table reveals that a total number of 4,935 users are involved in eight conferences' Twitter conversations. Oceanology had high interaction with 542 users whereas H2O had low interaction with 72 users. In addition, the maximum and minimum values shown the greatest and least number of tweets produced by a user. However, the maximum number of tweets for all the conferences were from their Twitter timeline and minimum number was always one tweet. Furthermore, the results of the mean revealed that Coastal Zone Canada has highest amount of activity per person and H2O with low activity. Oceanology has high deviation within the tweets and H2O with low deviation.

Furthermore, Figure 4.1 is a histogram that visualizes the distribution of Ocean conference tweets. The graph reveals the user engagement in the tweet conversation. About 175 users had posted only 1 tweet and only one user posted more than 250 tweets in the conference conversation. The graph clearly conveyed that the most active participant in ocean Twitter conversation was the conference organizer.

Table 4.1: Descriptive Statistics Table for Each Conferences' Tweets

Conference	Twitter Source	No. of Tweets	No. of Users	Max.	Min.	Mean	Median	Standard Deviation
OCEANS	@Oceans_Conf #OceansConference #Oceans2020SG #Oceans2019 #Oceans2018 #OCEANS20GulfCoast #OCEANS20Singapore	1,192	219	260	1	5.44	1	21.1
ArcticNet	@ArcticNet #AC2020 #AC2019 #AC2018	1,773	414	539	1	4.28	1	26.91
CMOS_SCMO	@CMOS-SCMO #CMOS2020 #CMOS2019 #CMOS2018	303	109	122	1	2.77	1	9.5
Coastal zone Canada	@CZCAssociation #CZC2020 #CZC2019 #CZC2018	756	91	426	1	8.3	1	44.64
Oceanology	@OceanologyIntl #oi2020 #oi2019 #oi2018	2,368	542	1193	1	4.36	1	51.19
Cold Harvest	@NAIA_NL #coldharvest #coldharvest2019 #coldharvest2018	1,650	368	314	1	4.48	1	19.45
H2O	@OceanTechNS #H2020 #H2019 #H2oConference	177	72	25	1	2.45	1	3.45
Oceans Obs	@OceanObs19 #OceanObs19 #OSM2020 #OSM19 #OSM18	495	155	154	1	3.19	1	12.48

Figure 4.1: Number of users vs tweets for ocean conference

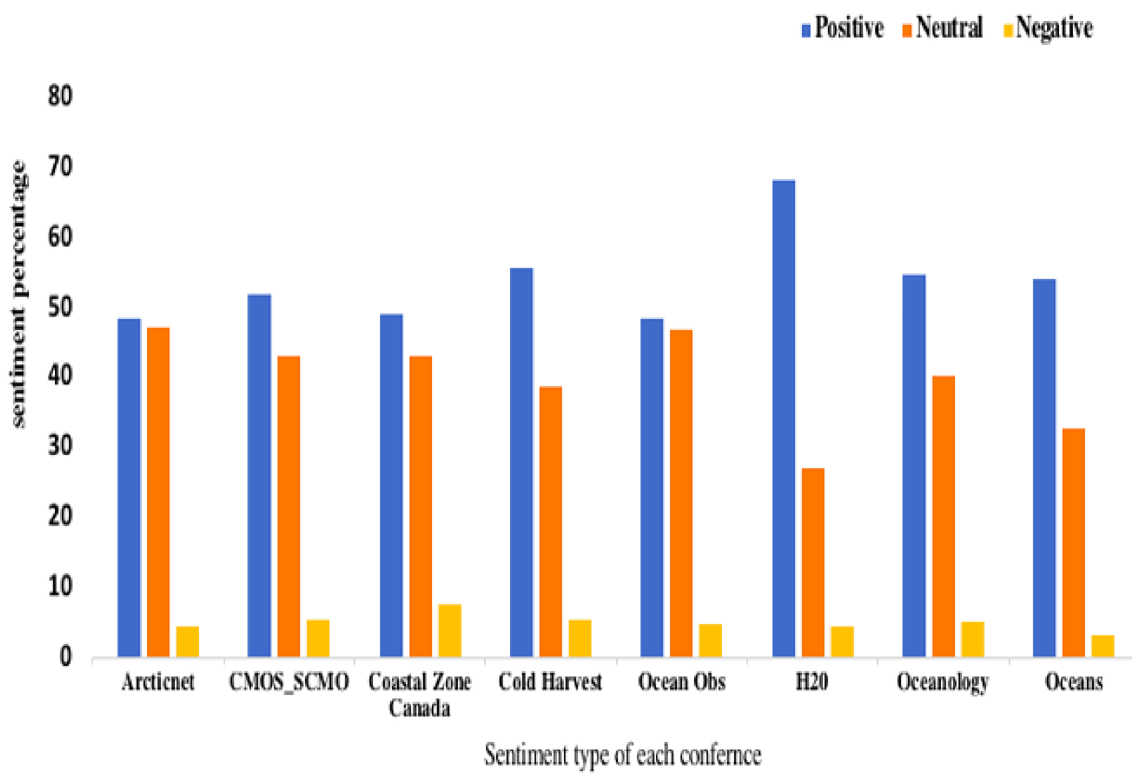


4.2 Sentiment Analysis

4.2.1 Results for Conference

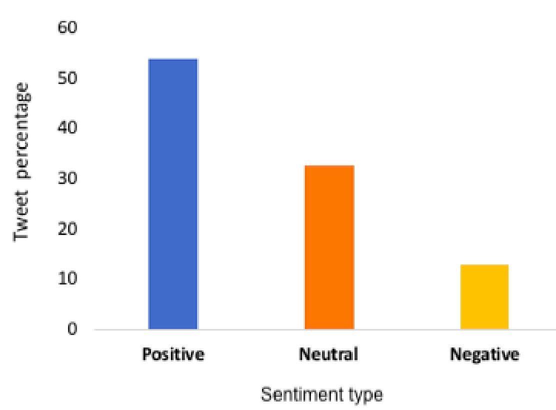
Analyzing 8,715 tweets from eight ocean conferences, the results presented sentiment of each conference form 2018 January to 2020 December. The graph for Figure 4.2 shows the sentiment percentage of eight conferences. It revealed users' opinions on each conference with high positive sentiment over three years. All the conferences had a noticeable neutral sentiment, this could be because of more promotional tweets in the data set.

Figure 4.2: Sentiment graph of eight conferences



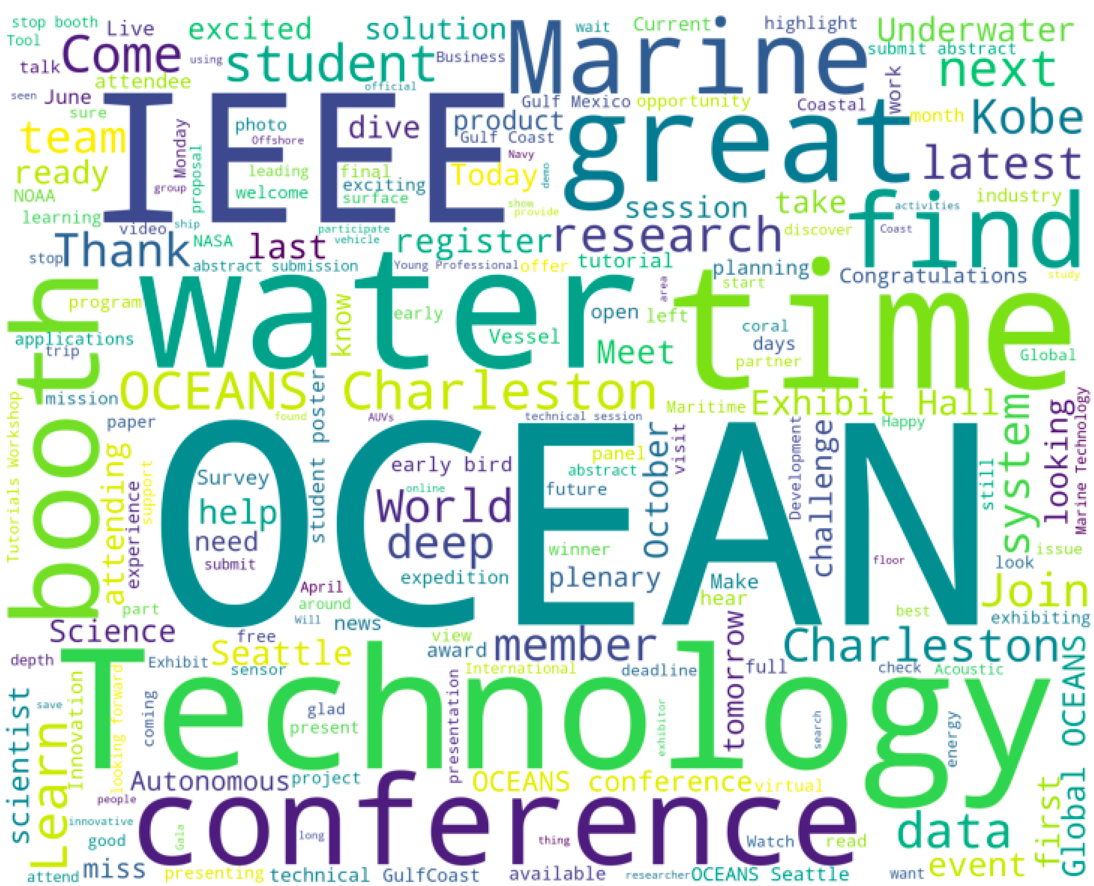
Analysis of individual conference is also presented. The bar graph in Figure 4.3 provided the sentiment analysis of the OCEANS 2020 conference in 2020. It revealed users' opinion on the conference with more positive sentiment with 54.1% of tweets, a neutral sentiment with 32.8%, and a negative sentiment with 13.1%.

Figure 4.3: Sentiment graph of OCEANS 2020



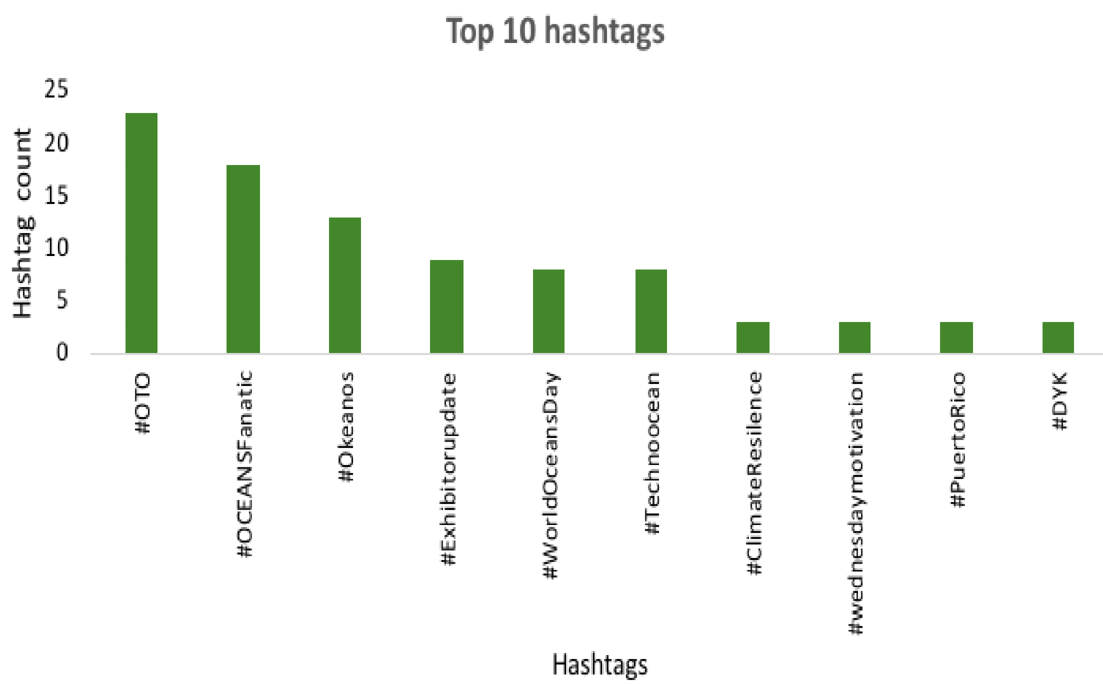
To provide more insights into the users' viewpoint on the conference, the project exposed most frequently used words in the tweets. The frequency of word distribution is important for organizers to analyze a user's interests via a tweet. Figure 4.4 represents the word cloud for the most frequently used words in the OCEANS conference tweets over three years. The most used words in the tweets are ocean, marine, water, technology, and research. The words revealed that most of the conference conversation involved the research topics related to oceans, marine life, and technology.

Figure 4.4: Word cloud of frequently used OCEANS conference tweets



Last, the top 10 frequently used hashtags in OCEANS conference tweets are depicted in Figure 4.5. These hashtags are most commonly specified by the users in the conference tweet. The graph excluded the two common hashtags #oceans and #OCEANS and revealed #OTO (Ocean and Techno-Ocean meetings together) was mostly specified in the tweets.

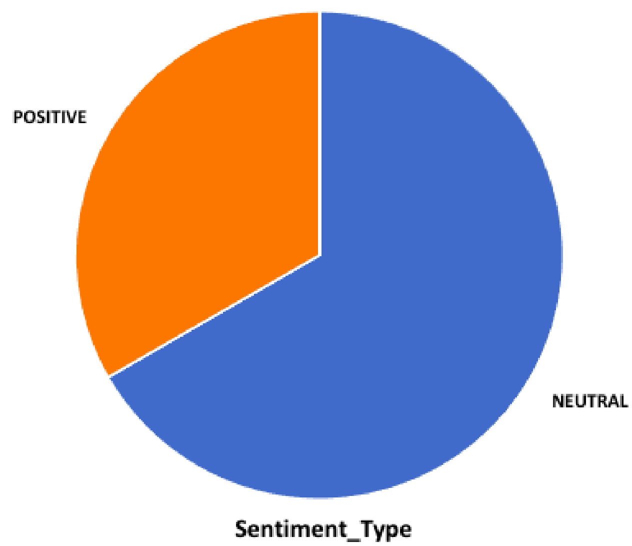
Figure 4.5: Top 10 hashtags used in OCEANS conference tweets



4.2.2 Results for Researcher

Analysis of researchers' conference-related tweets was presented based on two criteria. The results provided an analysis of 410 researchers and user accounts retrieved from the conference tweets. Figure 4.6 represented the sentiment analysis for Steve Gain. The analysis evaluated the sentiment of Steve Gain related conference tweets. The graph resulted in a more neutral sentiment with 66.7%, with positive 33.3%, and zero negative sentiment for the researcher. The dominant neutral sentiment of the tweets exposed that most of Steve Gain’s tweets were related to promoting and advertising conference presentations/events.

Figure 4.6: Sentiment analysis on Steve Gain conference related tweets



4.3 Social Network Analysis

4.3.1 Results for Organizers

The results presents social network analysis for the eight conferences. Figure 4.7 represented the social network connectivity between all the conferences. There are green nodes that indicated user account (person), and pink nodes indicated an organizational Twitter account. There are eight coloured edges connected between each node, each colour represents different conference. The network graph revealed a strong network connectivity among conferences. More specifically, the connectivity between OCEANS, H2O, and Oceanology conference was high whereas ArcticNet is slightly isolated.

Figure 4.7: Social network analysis of eight conferences

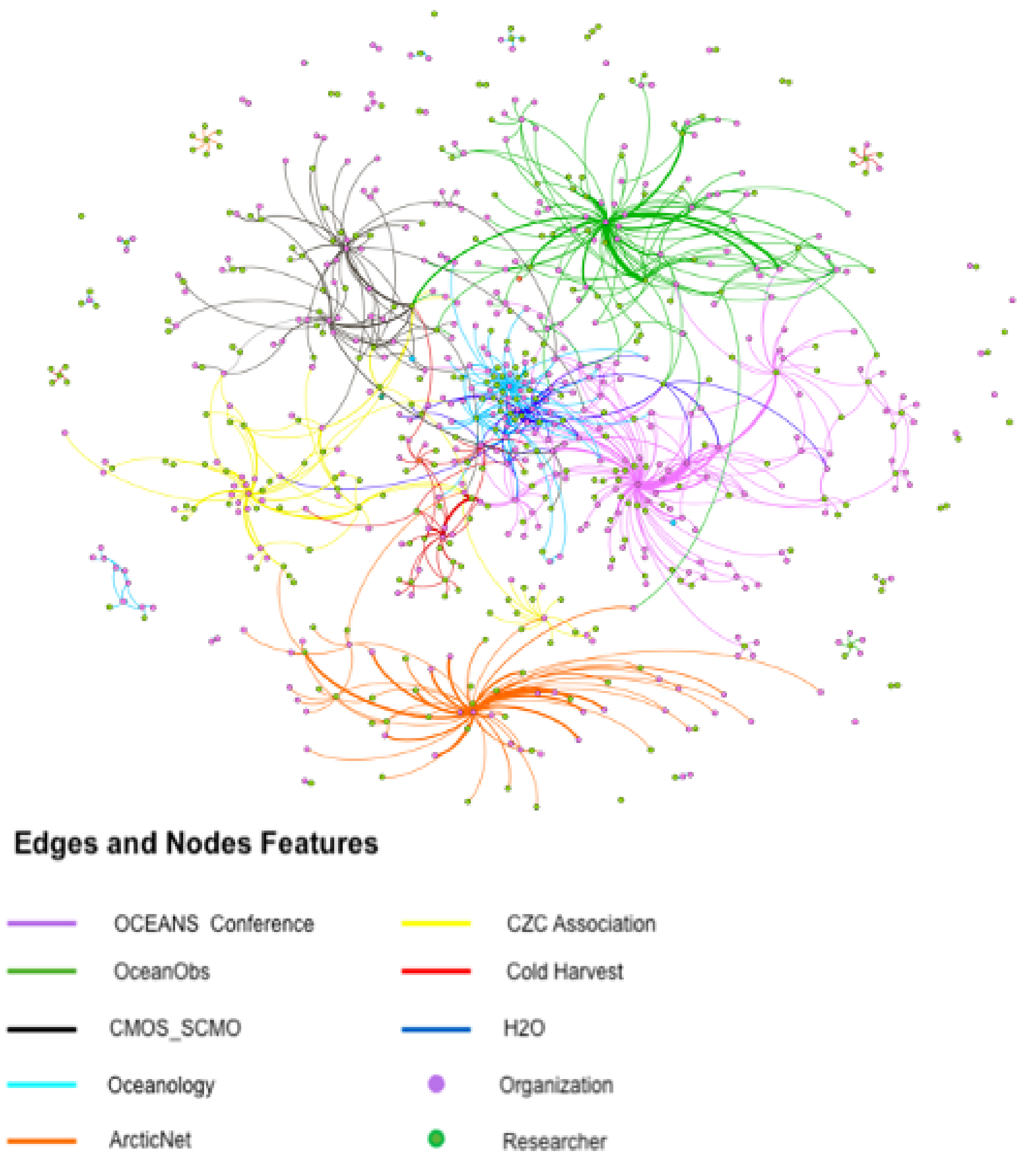


Table 4.2: Network overview of eight conferences

Attribute	Score
Average degree	1.338
Average weighted fegree	1.776
Graph density	0.002
Connected components	38
Average path length	5.147

On the other hand, the analysis provided network structure for each conference separately. Figure 4.8 presents the social network analysis of the OCEANS conference. It illustrated connections and conversations produced around a conference. The graph revealed that most of the users that tweeted for the conference are mostly interconnected. Each colour represented sub-groups conversation related to the conference. There are 30 sub-groups formed within the OCEAN conference conversation. The network statistics in Table 4.3 shows the high connected components and modularity.

Figure 4.8: Social network analysis for OCEANS conference

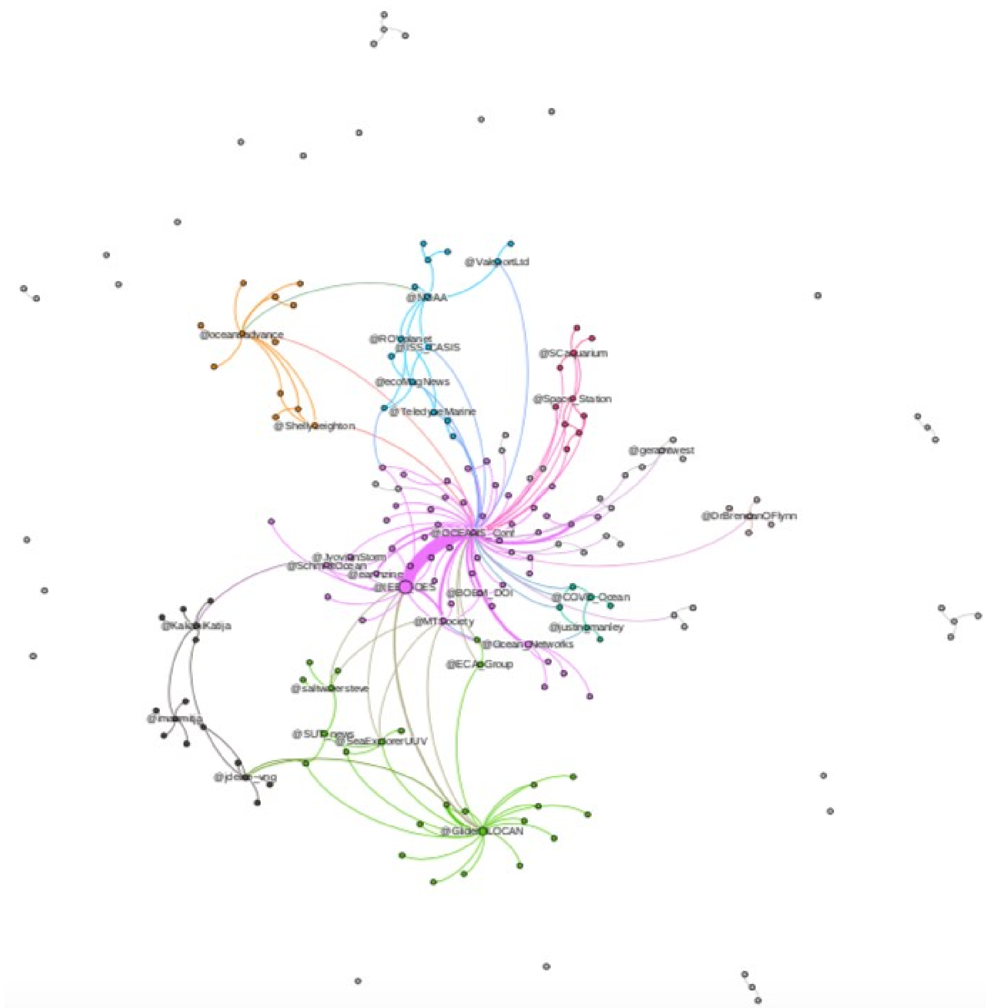


Table 4.3: Network Overview of OCEAN Conference

Attribute	Score
Avg. degree	1.209
Avg. weighted degree	1.346
Modularity	0.58
Network diameter	3
Connected components	22

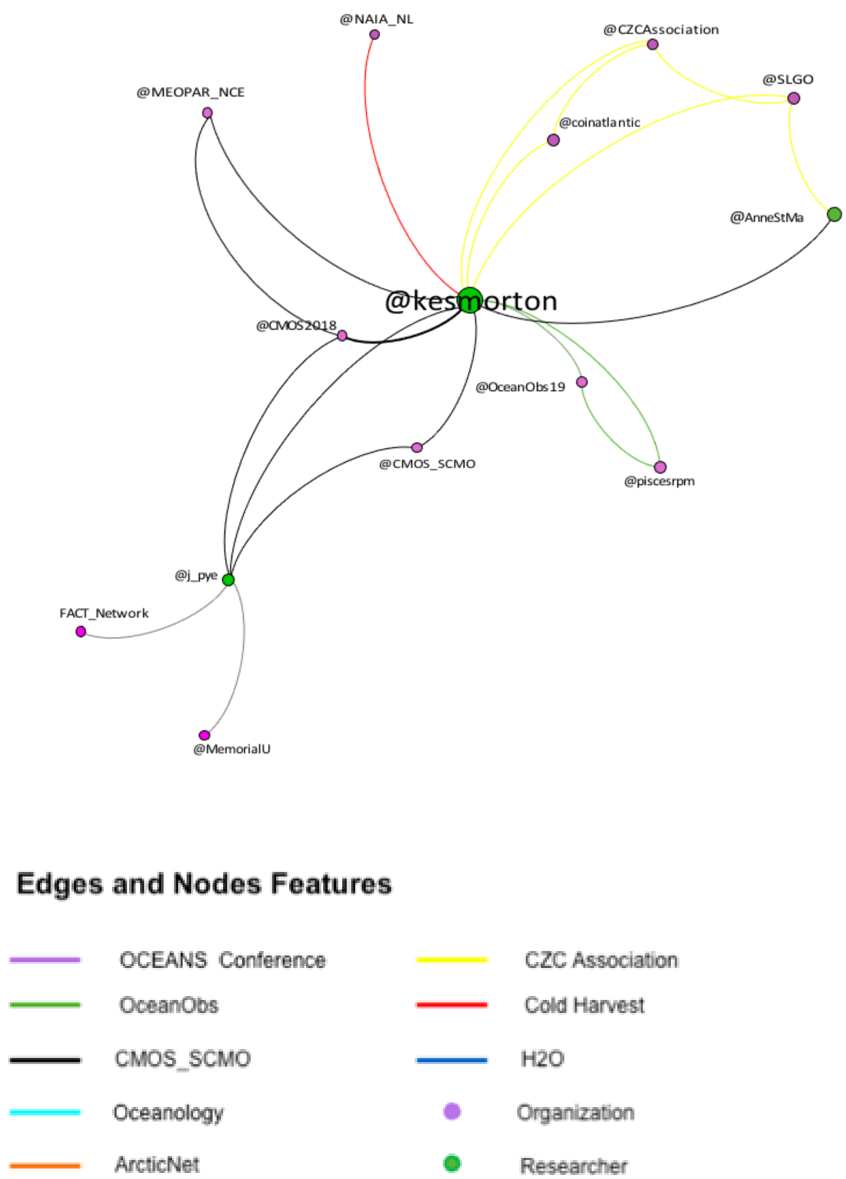
4.3.2 Results for Researchers

The social network of the researcher is evaluated based on the user accounts specified in the tweets. Figure 4.9 represents the social network of @kesmorton conference participation. It illustrated connections produced around the researcher using conference tweets. A different coloured edge represents relationships with different conferences. The graph revealed that @kesmorton is connected with four conferences: @CZC Association; @CMOS_SCMO; @OceanObs; and @NAIA_NL; six organizations; @piscesrpm;@FACT_Network; @colnaltantic; @SLGO; @MemorialU and @MEOPAR_NCE; and two users. Additionally, the network statistics of the graph in Table 4.4 provided more insights into the analysis with a high weighted degree with moderate modularity and low average path length.

Table 4.4: Network overview of @kesmorton

Attribute	Score
Avg. Degree	1.467
Avg. Weighted Degree	12.067
Modularity	0.5164
Network Diameter	2
Avg. path length	1.436

Figure 4.9: Social network analysis of @kesmorton



Chapter 5

Discussion

Twitter can be a valuable information source for both conference organizers and researchers. The research provides a data collection process that can lead to a robust tool for measuring conference participation. The scholarly content on Twitter assisted to analyze and measure user engagement and social relationships. The amount of engagement for each conference is presented using Twitter API; Implementing Sentiment analysis provided individual user's viewpoints on a specific conference or on a researcher; Word clouds provided a quick overview of frequent words present in the tweets; social network analysis exposed the interactions and social connectivity among conferences, researchers and organizations.

If we revisit the outlined objectives, and literature of the study in Chapter 1 and Chapter 2 we notice that the findings significantly contributed to it with few limitations. We present below the reflections on each objective, research contributions, and limitations in detail.

5.1 Reflections on Objectives

Objective 1: To determine the amount of user engagement of a conference

The amount of user engagement of a conference is revealed by a descriptive statistics table for each conference tweets. The analysis could help conference organizers to understand the level of user engagement of the people in the conference. The number of users and tweets show how many people are involved in the conference interaction and the findings support that the level of engagement of each conference varied from other. However, the maximum value of the eight conferences was high, and it is known that most of the tweets posted were from their conference account (organizer) for promotions, speaker's introduction and other conference related information. For instance, the OCEANS conference had 1,192 tweets from 219 users. Out of 1,192 tweets, 260 tweets were posted from @Oceans_Conf. The spread and variations of tweets is calculated by standard deviation and the score is 21.1. The average tweet activity per person in OCEANS conference is 5.44. The values are high because the large number of users on an individual level tweeted only once. These findings assist conference organizers to understand the diversity of the amount of participation in conference interaction.

Objective 2: To determine the sentiment of conference tweets

The results from the sentiment analysis provided the nature of user engagement of the conference. However, it should be interpreted as an approximation because VADER is a lexicon-based analysis and may not be based on true emotions. The findings interpreted high positive and low negative sentiment towards each conference. There were also a more neutral sentiment percentage because of more conference promotional tweets. However, the analysis would be more powerful if there were more conversational tweets rather than promotional tweets. Additionally, we have observed the number of tweets for the OCEANS conference gradually decreased from 2018 to 2020. Potentially, the reason could be the introduction of dedicated mobile applications for conference discussion in later years. Furthermore, the results from the findings provided more insights into the user engagement by showing most frequently used words and top hashtags. This may assist conference organizers to analyze/understand the most promising event, conference topic, presentation, and other key points of the conference

Similarly, the sentiment analysis for the researcher provided user engagement. However, there were a few tweets on the researcher/research due to the low social media presence. For an instance, only 21 conference-related tweets were retrieved about researcher Steve Gain from 2018 to 2020. More tweets would help provide more accurate sentiment analyses.

Objective 3: To determine the interactions of Twitter users

Results on our investigation on interaction revealed social relations among users. The developed social network analysis model focused on the level of interconnectivity and social key concepts of the interactions. The analysis assists conference organizers to identify the information flow and level of interconnectivity. Further, the graphs of the network analysis may help the organizer to recognize the strong and weak connections to make better decisions on upcoming conferences. The analysis facilitates strong relational support for a conference and helps to strengthen their social relationships wherever required. The findings revealed the OCEANS conference network analysis identified 30 strong connections and 32 unconnected nodes. In addition, the network provided a highly centrality node because of more tweets from the conference promoters.

On the other hand, the investigation on the network analysis of the researcher exposed the social relations with users. It may help researchers to recognize and understand their tweets' flow and social relationships. The findings exposed the social connectivity and conference participation of the researcher @Kesmorton. The graph may assist the researcher to visualize the social media network and spread of their tweets. Most importantly, the analysis explores their Twitter network. The network graph is presented using @ (Twitter handles). However, it would be more efficient if we present the connections by researcher's names rather than Twitter accounts.

5.2 Research Contribution

This research study makes three novel contributions to the existing literature. We present the contributions in detail.

Contribution 1: No prior study has been done on Twitter use in ocean conferences. This work represents a first attempt on describing the Twitter use in ocean discipline research. Existing Twitter analysis on conferences were implemented using several other fields such as humanities, medicine, and computer science (Al-Aufi & Fulton, 2014; Desai et al., 2012; Wen et al., 2014). This thesis explored how Twitter provided conference participation analysis to ocean conference organizers and researchers.

Contribution 2: The study is beneficial for the oceans discipline and makes recommendations to improve their conferences. The research is helpful as the results interpreted that conferences are having positive but also neutral conversations, which recommends ocean researchers and attendees should have more excitement in their conferences. Additionally, there are spikes of engagement that are run by the conference organizers which further exposes that most of the attendees are disengaged with conference conversations. The findings encourage people in oceans to actively participate in conference conversations to improve their conferences. The discussions not only help to enhance engagement but also assist to reach the public on current research studies in the ocean discipline.

Contribution 3: To the best of our knowledge prior work has been implemented a detailed study on measuring conference user engagement via sentiment analysis and visualizing interactions for conference participation. This study extends existing literature by combining two methods and using natural language processing techniques. It focused on analyzing social connectivity via Twitter interaction. Using Gephi network tool, we analyzed the conference, researcher, and organization's social relations on Twitter. The results show overlap between the conferences and reveal that the oceans community is tight, and they are supportive of one another.

5.3 Limitations

The investigation has few limitations. The reason for each limitation is described in detail with overcome solutions.

5.3.1 Low Social Media Presence

Low social media presence of researchers and conference organizers provided little data for the analysis. If there was more information and interaction about conference on Twitter, the analysis will be more effective. The research analysis is limited to Twitter platform, there may be comments in other Web and data sources. In addition, the personal discussion about the conference and researcher talk that was made during or around the conference are not considered for the analysis.

The study further continues to analyze conference user engagement using several other data sources such as online surveys and other social media platforms. Online surveys would overcome the limitation on analyzing actual users' emotions, whereas exploring multiple social media sources would expand the presence of conference and help to improve the sentiment score. Additionally, future studies will investigate in more details the conversations and engagement on Twitter during or around conferences.

5.3.2 Sentiment Analysis

Sentiment analysis provided positive, neutral and negative sentiment based on lexicon analysis applied on each tweet with no true emotions. While it is highly accepted in the research field in analyzing user engagement, it misses human-eye. To overcome this limitation, the future research work modifies VADER dictionary with more collection of words to improve sentiment score.

5.3.3 Limitations in Data Collection

Tweets not deposited in the conference timeline such as (@OCEANS, #Oceans2019, #Oceans2020) were not considered for the analysis. However, the results may slightly vary if there exists such information on Twitter. Similarly, the tweets related to a researcher not specified in conference and researcher timeline are not considered. On the other hand, researchers are provided with an analysis (user engagement and conference participation) only if they/their research data is included in the conference related tweets. While there are researchers who may not be active on social media but have extensive public outreach.

Further research in this subject should incorporate all the tweets related to conferences other than from timeline and hashtag account and list of researcher names participated in the conference that help to overcome these limitations.

5.3.4 Privacy Implications

The data collected for the study was extracted from social media — a publicly available source. We explored and measured individuals' private information without their explicit consent. However, Twitter API provides keys to authenticate and utilize the social data of users. Moreover, data on Twitter is publicly available data and users give explicit consent to provide their data to the public.

Chapter 6

Conclusion and Future Work

In this paper, we showed how scholarly content on Twitter assisted to analyze and measure user engagement and social relationships. The conference data available from Twitter API helped us to build a data collection process that captured the amount of user engagement, sentiments, and network interconnectivity. We conclude that the research not only provides an overview of conference experience using scholarly interaction but also proposes a system that provides robust indicators to measure conference participation.

This paper extended existing literature by implementing sentiment analysis and social network analysis to assist conference organizers and researchers. The experiment results show that performing sentiment analysis revealed public sentiment on the conference. The findings provided positive, negative and neutral sentiment. Applying frequency distribution on conference twitter data revealed most trending conference topics and popular hashtags to assist organizers to make better decisions on future events. Finally, the network analysis exposed the network connectivity to understand Twitter interactions and social relationships of conference and researcher. Thus, the study provided information source for conference organizers and researchers to analyze user engagement on Twitter.

Using Twitter conference data, this project constructed a data collection process that can be valuable for developing a powerful tool for Twitter analysis. We recommend additional research to propose a software system that provides potential indicators to measure conference participation and the social impact of a researcher. Measuring conference participation and assessing social impact is important for researchers to understand user engagement of their research, this paper suggests that scholars and conference organizers need to interact more actively on Twitter. This helps further work in this discipline to explore more and provide more effective analysis.

Besides, the study recommends extending the data for the project by collecting the conference related tweets from all the international ocean conferences from past years. This could offer improved results on the user engagement and network graph. Additionally, we could also improve the size of data by collecting data not only from Twitter but from other data sources such as online surveys, in-house comments, and other social media sources.

Furthermore, this paper recommends additional research in analyzing conference experience. As observed in the results section, the number of tweets for a few conferences reduced gradually from 2018 to 2020. The reason behind the reduction in tweets may be the introduction of dedicated mobile applications for conference interaction. Further studies could investigate in more detail the conversations and engagement via conference mobile applications.

References

Abascal-Mena, R., Lema, R., & Sèdes, F. (2015). Detecting sociosemantic communities by applying social network analysis in tweets. *Social Network Analysis and Mining*, *5*(1), 1–17. DOI: 10.1007/s13278-015-0280-2

Al-Aufi, A., & Fulton, C. (2014). Use of social networking tools for informal scholarly communication in Humanities and Social Sciences disciplines. *Procedia , Social and Behavioral Sciences*, *147*(C), 436–445. https://doi.org/10.1016/j.sbspro.2014.07.135

Al-Daihani, S. M., Al-Qallaf, J. S., & Alsaheeb, S. A. (2018). Use of social media by social science academics for scholarly communication. *Global Knowledge, Memory and Communication*, *67*(6/7), 412–424. https://doi.org/10.1108/GKMC-11-2017-0091

Andy W., Chen. (2018). Sentiment analysis of news articles and its comments: A natural language processing application. *Global Journal of Advanced Engineering Technologies and Sciences*, *5*(5). https://www.gjaets.com/Issues%20PDF/Archive-2018/May-2018/3.pdf

Barthelemy, M. (2004). Betweenness centrality in large complex networks. *The European Physical Journal. B, 38*(2), 163–168. https://doi.org/10.1140/epjb/e2004-00111-4

Bastian, M., Heymann, S., & Jacomy, M. (2009). Gephi: An open source software for exploring and manipulating networks. International AAAI Conference on Weblogs and Social Media. https://gephi.org/publications/gephi-bastian-feb09.pdf

Bert, F., Zeegers Paget, D., & Scaioli, G. (2016). A social way to experience a scientific event: Twitter use at the 7th European Public Health Conference. *Scandinavian Journal of Public Health*, *44*(2), 130–133. https://doi.org/10.1177/1403494815612932

Bird, S., Klein, Ewan, & Loper, Edward. (2009). *Natural language processing with Python*. Beijing ; Cambridge [Mass.]: O'Reilly.

Borg, A. & Boldt, M. (2020). Using VADER sentiment and SVM for predicting customer response sentiment. *Expert Systems with Applications*, *162*, Article no. 113746. DOI: 10.1016/j.eswa.2020.113746

Bornmann, L., Haunschild, R., & Marx, W. (2016). Policy documents as sources for measuring societal impact: How often is climate change research mentioned in policy-related documents? *Scientometrics*, 109, 1477–1495.DOI:10.1007/s11192–016-2115-y

Calma, A., & Davies, M. (2015). Studies in higher education 1976–2013: A retrospective using citation network analysis. *Studies in Higher Education (Dorchester-on-Thames)*, *40*(1), 4–21. https://doi.org/10.1080/03075079.2014.977858

Chandrasekaran, R., Mehta, V., Valkunde, T., & Moustakas, E. (2020). Topics, trends, and sentiments of tweets about the COVID-19 pandemic: Temporal infoveillance study. *Journal of Medical Internet Research*, *22*(10), E22624. DOI: 10.2196/22624

Cherven, K. (2013). Network graph analysis and visualization with Gephi: Visualize and analyze your data swiftly using dynamic network graphs built with Gephi (Community experience distilled). Packt Publishing.

Desai, T., Shariff, A., Kats, A., Fang, M., Christiano, X., Ferris, C., & Gupta, V. (2012). Tweeting the meeting: An in-depth analysis of Twitter activity at Kidney Week 2011. *PLoS ONE*, *7*(7), e40253. DOI: 10.1371/journal.pone.0040253

Dimitrios, N. K., Sakas, D. P., & Vlachos, D. S. (2014). Modeling publications in academic conferences. *Procedia Social and Behavioral Sciences*, *147*(C), 467–477.

Dicle, M. F., & Dicle, B. (2018). Content analysis: Frequency distribution of words. *The Stata Journal*, *18*(2), 379–386. https://doi.org/10.1177/1536867X1801800205

Franceschet, M. (2010). The role of conference publications in CS. *Communications of the ACM*, *53*(12), 129–132. https://doi.org/10.1145/1859204.1859234

Guoyong, M., & Ning, Z. (2013). Analysis of average shortest-path length of scale-free network. *Journal of Applied Mathematics*, *2013*, 1–5. DOI: 10.1155/2013/865643

Hajian, B., & White, T. (2011). Modelling influence in a social network: Metrics and evaluation. Conference: PASSAT/SocialCom 2011, Privacy, Security, Risk and Trust (PASSAT), 2011 IEEE Third International Conference on and 2011 IEEE Third International Conference on Social Computing (SocialCom), 9–11 Oct., Boston, MA. DOI: 10.1109/PASSAT/SocialCom.2011.118

Han, H., Zhang, Y., Zhang, J., Yang, J., & Zou, X. (2018). Improving the performance of lexicon-based review sentiment analysis method by reducing additional introduced sentiment bias. *PloS One*, *13*(8), E0202523. DOI:10.1371/journal.pone.0202523

Hansen, D., Schneiderman, B., & Smith, M. A. (2011). Analyzing social media networks with NodeXL: Insights from a connected world. Morgan Kaufmann.

Hassan, S., Aljohani, N., Idrees, N., Sarwar, R., Nawaz, R., Martínez-Cámara, E., . . . & Herrera, F. (2020). Predicting literature's early impact with sentiment analysis on Twitter. Knowledge-based Systems, 192, Knowledge-based systems, 15 March 2020, *192*.

Haustein, S., Bowman, T. D., & Costas, R. (2016). Interpreting 'altmetrics': Viewing acts on social media through the lens of citation and social theories. In C. R. Sugimoto (ed.), *Theories of Informetrics and Scholarly Communication* pp. 372–406. de Gruyter. http://arxiv.org/abs/1502.05701

Holmberg, K., & Thelwall, M. (2014). Disciplinary differences in Twitter scholarly communication. *Scientometrics*, *101*(2), no. 8, 1027–1042. DOI: 10.1007/s11192-014-1229-3

Hongbo, W. (2017). A Conference to #SaveOurOcean. *UN Chronicle.* https://www.un.org/en/chronicle/article/conference-saveourocean

Hutto, C. J., & Gilbert, E. (2014). VADER: A parsimonious rule-based model for sentiment analysis of social media text. In Proceedings of the Eighth International Conference on Weblogs and Social Media (ICWSM-14), Ann Arbor, MI, June 2014. http://comp.social.gatech.edu/papers/icwsm14.vader.hutto.pdf

Jacomy, M., Venturini, T., Heymann, S., & Bastian, M. (2014). ForceAtlas2, a continuous graph layout algorithm for handy network visualization designed for the Gephi software. *PloS One*, *9*(6), E98679. doi: 10.1371/journal.pone.0098679

Kim, E. H-J., Jeong, Y. K., Kim, Y., Kang, K. Y., & Song, M. (2016). Topic-based content and sentiment analysis of Ebola virus on Twitter and in the news. *Journal of Information Science*, *42*(6), 763–781. https://doi.org/10.1177/0165551515608733

Kharde, V. A., & Sonawane, S. (2016). Sentiment analysis of Twitter data: A survey of techniques, *139*(11), 5–15. https://arxiv.org/ftp/arxiv/papers/1601/1601.06971.pdf

Li, J., & Greenhow, C. (2015). Scholars and social media: Tweeting in the conference backchannel for professional learning. *Educational Media International*, *52*(1), 1–14. https://doi.org/10.1080/09523987.2015.1005426

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. DOI: 10.1007/s11192-017-2413-z

Lemay, D. J., Basnet, R. B., Doleck, T., & Bazelais, P. (2019). Social network analysis of twitter use during the AERA 2017 Annual Conference. *Education and Information Technologies*, *24*(1), 459–470. https://doi.org/10.1007/s10639-018-9787-2

López-Goñi, I., & Sánchez-Angulo, M. (2017). Social networks as a tool for science communication and public engagement: Focus on Twitter. *FEMS Microbiology Letters*, *365*(2). DOI: 10.1093/femsle/fnx246

Lubchenco, J., & Grorud-Colvert, K. (2015). OCEAN. Making waves: The science and politics of ocean protection. *Science*, *350*(6259), 382–383. doi: 10.1126/science.aad5443

Luzón, M., Albero-Posac, S.(2020). 'Had a lovely week at #conference2018': An analysis of interaction through conference tweets. *RELC Journal*, *51*(1), 33–51. https://doi.org/10.1177/0033688219896862

Mastley, C. P. (2017). Social media and information behavior: A citation analysis of current research from 2008–2015. *The Serials Librarian*, *73*(3–4), 339–351. https://doi.org/10.1080/0361526X.2017.1356420

McNamara, P. (2013). #ACMHN Looking back at the 2013 consultation liaison/Perinatal infant conference through a social media lens [Web log comment]. https://meta4rn.com/2013/06/12/noosa/

Neill, A., Cronin, J., Brannigan, D., O'Sullivan, R., & Cadogan, M. (2014). The impact of social media on a major international emergency medicine conference. *Emergency Medicine Journal*, *31*(5), 401–404. DOI: 10.1136/emermed-2012-202039

Neumann, B., & Unger, S. (2019). From voluntary commitments to ocean sustainability. *Science*, *363*(6422), 35–36. DOI: 10.1126/science.aav5727

Oester, S., Cigliano, J. A., Hind-Ozan, E. J., & Parsons, E. C. M. (2017). Why conferences matter—An example from the International Marine Conservation Congress. *Frontiers in Marine Science*, *4*, 257. https://doi.org/10.3389/fmars.2017.00257

Our Ocean Conference 2017: Final report. (n.d.). Our Ocean Malta 5–6 October 2017. https://www.ourocean2017.org

Pulido, C. M., Redondo-Sama, G., Sordé-Martí, T., & Flecha, R. (2018). Social impact in social media: A new method to evaluate the social impact of research. *PloS One*, *13*(8), E0203117. DOI: 10.1371/journal.pone.0203117

Salzmann-Erikson, M. (2018). Mental health nurses' use of Twitter for professional purposes during conference participation using #acmhn2016. *International Journal of Mental Health Nursing*, *27*(2), 804–813. https://doi.org/10.1111/inm.12367

Shiffman, D. (2012). Twitter as a tool for conservation education and outreach: What scientific conferences can do to promote live-tweeting. *Journal of Environmental Studies and Sciences*, *2*(3), 257–262. DOI: 10.1007/s13412-012-0080-1

Sugimoto, C., Work, S., Larivière, V., & Haustein, S. (2017). Scholarly use of social media and altmetrics: A review of the literature. *Journal of the Association for Information Science and Technology*, *68*(9), 2037–2062.

Tattersall, A., & Carroll, C. (2018). What can Altmetric.com tell us about policy citations of research? An analysis of Altmetric.com data for research articles from the University of Sheffield. *Frontiers in Research Metrics and Analytics*, 2. 9. https://doi.org/10.3389/frma.2017.00009

Varlamis, I., Eirinaki, M., & Louta, M. (2010). A study on social network metrics and their application in trust networks. 2010 International Conference on Advances in Social Networks Analysis and Mining, 168–175. DOI:10.1109/ASONAM.2010.40

Waltman, L., & Costas, R. (2014). F1000 recommendations as a potential new data source for research evaluation: A comparison with citations. *Journal of the Association for Information Science and Technology*, 65, 433–445. DOI:10.1002/asi.23040

Webb, S. (2016). Twitter use in physics conferences. *Scientometrics*, *108*(3), 1267–1286. DOI: 10.1007/s11192-016-2031-1

Wen, X., Lin, Y., Trattner, C., & Parra, D. (2014). Twitter in academic conferences: usage, networking and participation over time. In *Proceedings of the 25th ACM Conference on Hypertext and Social Media*, of *HT '14*, 285–290, New York, NY, USA, 2014. ACM. DOI 10.1145/2631775.2631826

Xie, Q., & Luo, T. (2019). Examining user participation and network structure via an analysis of a Twitter-supported conference backchannel. *Journal of Educational Computing Research*, *57*(5), 1160–1185. https://doi.org/10.1177/0735633118791262

Zappavigna, M. (2012). *The discourse of Twitter and social media: How we use language to create affiliation on the web*. Continuum International.

Appendix

Database Entity-relationship Diagram of the Project