ISTANBUL TECHNICAL UNIVERSITY ★ GRADUATE SCHOOL OF SCIENCE ENGINEERING AND TECHNOLOGY

UNDERSTANDING TWITTER USERS' BEHAVIOUR BY SOCIAL NETWORK ANALYSIS DURING DISASTERS

M.Sc. THESIS

Gözde Merve DEMİRCİ

Department of Industrial Engineering

Industrial Engineering Programme

JULY 2020

ISTANBUL TECHNICAL UNIVERSITY ★ GRADUATE SCHOOL OF SCIENCE ENGINEERING AND TECHNOLOGY

UNDERSTANDING TWITTER USERS' BEHAVIOUR BY SOCIAL NETWORK ANALYSIS DURING DISASTERS

M.Sc. THESIS

Gözde Merve DEMİRCİ
(507171141)

Department of Industrial Engineering

Industrial Engineering Programme

Thesis Advisor: Prof. Dr. Şule Itır SATOĞLU
Co-Advisor: Assoc. Prof. Dr. Gülüstan DOĞAN

JULY 2020

ISTANBUL TEKNİK ÜNİVERSİTESİ ★ FEN BİLİMLERİ ENSTİTÜSÜ

AFET DURUMUNDA TWITTER KULLANICILARININ SOSYAL AĞ ANALİZİ İLE DAVRANIŞINI ANLAMA

YÜKSEK LİSANS TEZİ

Gözde Merve DEMİRCİ
(507171141)

Endüstri Mühendisliği Anabilim Dalı

Endüstri Mühendisliği Programı

Tez Danışmanı: Prof. Dr. Şule Itır SATOĞLU
Eş Danışman: Doç. Dr. Gülüstan DOĞAN

TEMMUZ 2020

Gözde Merve DEMİRCİ, a M.Sc. student of İTU Graduate School of Science Engineering and Technology student ID 507171141, successfully defended the thesis entitled “UNDERSTANDING TWITTER USERS’ BEHAVIOUR BY SOCIAL NETWORK ANALYSIS DURING DISASTERS”, which she prepared after fulfilling the requirements specified in the associated legislations, before the jury whose signatures are below.

Thesis Advisor :	**Prof. Dr. Şule Itır SATOĞLU** İstanbul Technical University	....
Co-advisor :	**Assoc. Prof. Dr. Gülüstan DOĞAN** Yildiz Technical University	....
Jury Members :	**Asst. Prof. Dr. Emine YAYLALI** Istanbul Technical University	....
	Asst. Prof. Dr. Ömer Faruk BEYCA Istanbul Technical University	....
	Assoc. Prof. Dr. Gülfem TUZKAYA Marmara University	....

Date of Submission : 10 June 2020
Date of Defense : 17 July 2020

To my beloved family,

FOREWORD

I would like to give my sincere appreciation to my advisors, Prof. Dr. Şule Itır SATOĞLU and Assoc. Prof. Dr. Gülüstan DOĞAN for their helpful advice and encouragement during this journey. Their valuable insights directions gave me a great ambition to complete this thesis. I have gained so much knowledge with their valuable help, and guidance. Other important people that I have gratitude are my colleagues Dilara GÜNDÜZ and Mert Can GİRGİN in my work area. They helped me a lot in this journey with their moral support.

Finally, I want to indicate my deepest gratefulness to my family. They always supported me. I am so appreciative of their always encouragement, patient, and support with full motivation. They always believed in me, and thank you so much for that.

July 2020 Gözde Merve DEMİRCİ

TABLE OF CONTENTS

Page

ABBREVIATIONS

API	**:** Application Programming Interface
GSM	**:** Global System for Mobile Communications
NLP	**:** Natural Language Processing
SA	**:** Sentiment Analysis
SMA	**:** Social Media Analytics
SMS	**:** Short Message Service
SNA	**:** Social Network Analysis

LIST OF TABLES

Page

LIST OF FIGURES

Page

UNDERSTANDING TWITTER USERS' BEHAVIOUR BY SOCIAL NETWORK ANALYSIS DURINNG DISASTERS

SUMMARY

Fast response and assistance is a crucial step in disaster situations. With the use of the internet and social media in recent times, moving disaster management to digital platforms has become logical and interesting. Besides, the fact that social media plays a big role in daily usage for the majority of people provides many conveniences both during and after the disaster and even in the process of preparing for disaster. In the literature, it has been proved that social media plays an effective role in a natural disaster situation. However, no implementation has been made to use this interaction so far. Piece by piece there are many articles that are studied on disaster management. From these works, it is seen that the results of these implementation made are positive for disaster management ideas. Social media, especially Twitter, has a very big reaction in terms of disaster time. Both of the communities who are affected and non-affected from the disaster would like to share their thoughts, beliefs. People tend to share the news they have found about the disaster and sometimes these news may not trustable. Unfortunately, sometimes these shares might direct sufferers in the wrong way. In order to manage and correct these mistakes, it is necessary to analyze the reliability and reputation of users on social media.

In this thesis, Social Network Analysis was applied in user "reply" interaction on Twitter. The main aim is to find a distinct user list and/or categories that can be useful in disaster management by sharing important messages. In disaster moments, accurate messages should be spread to the majority of the community in the fastest way. For this, we analyzed the behavior of social media before, during, and after the disaster moment and see if there is a significant user category to influence. We first retrieved raw data from Twitter and pre-processed the necessary attributes. We then created a node and edge tables which are essential for social network analysis. With the selected centrality measurements, we have compared and analyzed user categories and decide on important users.

In the first part of the thesis, a literature review and prior works about the topic have been mentioned. Related works and their solutions were discussed. From the related works, the most similar three articles were mentioned and the significant difference in our study was stated. The purpose of the study and after implementation, what solution will be get was explained. All possible research questions were stated and the hypothesis was shown. After the literature part, the term Social Media Analytics was detailly explained with equations. Social Network Analysis tools and their measurements, centrality calculations were stated. In this part, all related and necessary centrality descriptions in SNA and graph mining were indicated and their formula equations were shown. Also, the concept of disaster management was explained as well as information spreading term.

In the methodology and implementation parts, Python language and Gephi application were used. We first prepared the collected data to SNA implementation in Python, then

with Gephi SNA visualizations and centrality calculations were made properly. Since the visualizations are an essential and critical part of this study, different types of plots were shown to decide the true category. From the results, it is proved that some apparent types of users react the same, and they are very influential when it comes to reaching the community. From the results, it is seen that statesmen -governmental- accounts have a great impact on message spreading in the disaster moments. For this purpose, eigenvector centrality is more suitable and gives proper results. Also, apart from statesmen, celebrities with verified accounts and media accounts also serve as a bridge in spreading messages. To make sure from the results, we have compared the other three similar disasters and found that the same categories and even the same user accounts are getting influential in disaster management. From these results, it can be said that in terms of disaster management, the important messages should be shared by statesmen, celebrities, and media.

AFET DURUMUNDA TWITTER KULLANICILARININ SOSYAL AĞ ANALİZİ ILE DAVRANIŞINI ANLAMA

ÖZET

Hızlı müdahale ve yardım, afet durumlarında çok önemli bir adımdır. Son zamanlarda internet ve sosyal medya kullanımı ile afet yönetiminin dijitale taşınması mantıklı ve ilgi çekici bir hale gelmiştir. Ayrıca, sosyal medyanın günlük kullanımda büyük rol oynaması, hem afet sırasında hem de sonrasında ve hatta afete hazırlık sürecinde bile birçok kolaylık sağlamaktadır. Literatürde, sosyal medyanın doğal afet durumlarında etkili bir rol oynadığı kanıtlanmıştır. Ancak, şimdiye kadar bu etkileşimi kullanmak için herhangi bir uygulama yapılmamıştır. Parça parça afet yönetimi üzerine incelenen birçok makale var ve sosyal medyanın afet yönetiminde başarılı olduğunu kanıtlayan olumlu sonuçlar mevcuttur. Sosyal medya, özellikle Twitter, önemli ve büyük olaylarda çok hızlı ve fazla etkileşim yapan bir platformdur. Twitter'da saniyede ortalama 900.000 adet tweet atıldığı bilinciyle bakarsak, Twitter'da biriken bu verilerin bilimsel olarak kullanılması çok yararlı bir durumdur. Topluluğun etkilenen ve etkilenmeyen tüm kısımları düşüncelerini, inançlarını ve bazen güvenilir olmasalar da buldukları haberleri paylaşmak ister. Ne yazık ki, bazen bu paylaşımlar afetten etkilenen kazazedeleri yanlış yönlendirebilir. Örneğin yakın zamanda yaşanan deprem afetinde, verilen yanlış bilginin ilgi çeekmesi ve birçok kez retweet (tekrar tweet) ve reply (cevap) almasından dolayı, kışın evleri yıkılan depremzedeler, aslında açık olmayan bir fabrikaya sığınmak için gitmiş ve yanlış yönlendirilmişlerdir. Bu hataları yönetmek ve düzeltmek için, kullanıcıların sosyal medyadaki güvenilirliğini ve itibarını analiz etmek gerekir.

Bu tezde Twitter'da kullanıcıların birbirlerine "cevap" ile oluşan etkileşimlerindeki Sosyal Ağ Analizi uygulanmıştır. Temel amaç, önemli mesajları paylaşarak afet yönetiminde yararlı katkılar sağlayabilecek belli bir kullanıcı listesi ve / veya kategoriler bulmaktır. Afet zamanında, doğru mesajlar topluluğun çoğunluğuna en hızlı şekilde yayılmalıdır. Bunun için afet anından önce, afet sırasında ve sonrasında sosyal medya kullanıcılarının davranışını analiz ettik ve etkileyecek önemli bir kullanıcı kategorisi olup olmadığına baktık. İlk olarak Twitter'dan ham verileri çektik ve gerekli nitelikleri saklayarak yararlı hale gelmesi için ön-işleme aşamasından geçirdik. Daha sonra sosyal ağ analizi için gerekli olan düğüm ve kenar tablolarını oluşturduk. Seçilen merkezi ölçümler ile kullanıcı kategorilerini karşılaştırıp analiz ettik ve önemli kullanıcılara karar verdik.

Tezin ilk bölümünde literatür taraması ve konuyla ilgili önceki çalışmalardan bahsedilmiştir. Konu ile ilgili makaleler araştırılarak katkıları üzerinde durulmuş ve yöntemleri ile sonuçlarından bahsedilmiştir. Bu ilgili çalışmalardan bu çalışmaya en benzer üç makale açıklanmış, ve çalışmamızdaki anlamlı farklılık belirtilmiştir. Çalışmamızın amacı ve yapılacak uygulamalardan sonra hangı çıktının elde edilmesi planlandığı hakkında bilgi verilmiştir. Çalışma için araştırma soruları belirtilmiş ve tez boyunca bu sorulara cevap bulacak uygulamalara yer verilmiştir. Tezin ikinci bölümünde Sosyal Medya Analitiği hakkında genel bilgiler verilerek, ne amaçla

yapıldığından bahsedilmiştir. Sosyal Medya Analitiğinin bir parçası olan Sosyal Ağ Analizi tanıtılmış ve bu çalışma için kullanılan araçlardan bahsedilmiştir. Bu bölümde, SNA ve grafik madenciliğinde ilgili ve gerekli tüm merkeziyet tanımları belirtilmiş ve formül denklemleri gösterilmiştir. Ayrıca, özellikle üzerinde çalıştığımız için, afet yönetimi kavramı ve bilgi yayma terimi açıklanmıştır.

Metodoloji ve uygulama bölümünde Python dili ve Gephi uygulaması kullanılmıştır. Uygulama için Türkiye'de meydana gelen Elazığ depremi ele alınmış ve yaşanan Elazığ depremi tarihlerinin 3 gün önce ve sonrası arasındaki tweet'ler \#deprem ve \#elazig filtreleri ile toplanmıştır. Toplanan ham veriler, önce veri keşfi ile önemli özellikler çıkartıldıktan sonra veri önişlemesi yapılmıştır. Veriyi sosyal ağ analizinde daha verimli sonuç ve genel bir kural oluşturabilmek adına Twitter kullanıcı adına göre atama yapılabilecek 6 kategori -Devlet, Medya, Deprem, Kurum, Birey, Diğer- oluşturulmuştur. Veri çekmesi, hazırlığı, önişlemesi ve görselleştirmsi Python dili ve JupyterLab defterinde yapılmıştır. Tüm kullanıcıların metinleri ile genel bir kelime bulutu oluşturulmuş ve genel anlamda felakeet anında kullanıcıların tepkileri analiz edilmiştir. Aynı zamanda kategori bazlı kullanıcılar için kelime bulutu oluşturularak kategori bazlı farklı bir etkileşimin olup olmadığına bakılmıştır. Sonrasında sosyal ağ analizi için gerekli olan düğüm (node) ve kenar (edge) listeleri oluşturulmuştur. Bu listelerde gerekli görülen özellikler de sütun olarak eklenmiştir. Sosyal ağ analizi, graf görselleştirmesi ve merkeziyet hesaplamaları Gephi uygulamasında yapılmıştır. Görselleştirmeler bu çalışmanın önemli ve kritik bir parçası olduğundan farklı değişkenler üzerinden şekiller oluşturulmuştur. Gephi'de elde edilmiş hesaplama tablosu Python ile görselleştirme için kullanılmıştır. Merkeziyet hesaplamaları ile bazı belirgin kullanıcı türlerinin aynı tepki gösterdiği ve topluluğa ulaşma konusunda çok etkili oldukları kanıtlanmıştır.

Sonuçlardan, devlet adamlarının (Devlet kategorisi) hesaplarının afet anlarında mesaj yayılması konusunda çok hızlı ve geniş kitleye ulaşabildiği gözlemlenmiştir. Açıklanan merkeziyet hesaplamalarından özvektör merkeziliği ve arasındalık merkeziliği göz önüne alınarak çalışmalar yapılırken, özvektör merkeziliğinin afet yönetiminde etkin olan kullanıcı gruplarını bulmada daha uygun olduğu da gözlemlenmiştir. Amacımız iletilecek bir mesajın Twitter topluluğuna olabilecek en hızlı şekilde ve maksimum topluluğa ulaşması olduğu için, ego ağ analizi ile özvektör merkeziliği ve arasındalık merkeziliği karşılaştırılmıştır. Ayrıca, devlet adamlarının yanı sıra, doğrulanmış hesapları olan ünlüler ve medya hesapları da mesajların yayılmasında bir köprü görevi gördüğü kanıtlanmıştır. Son olarak, sonuçlardan emin olmak adına Elazığ depremine benzer son 5 yılda yaşanmış olan 3 deprem olayları için aynı çalışmaları yaparak, kategorilerin ve hatta aynı kullanıcı hesaplarının afet yönetiminde etkili olduğunu gördük. Bu sonuçlardan afet yönetimi açısından önemli mesajların devlet adamları, ünlüler ve medya tarafından paylaşılması gerektiği söylenebilir.

Her ne kadar arasındalık merkeziyet hesaplaması, problemimizde bulmak istediğimiz kullanıcı tipini bulmaya yardımcı olacak şekilde olsa da, arasındalık merkeziyeti ile çıkan sonuçlarda genel olarak “Diğer” kategorisi baskın çıkmıştır. Bu da tahmin edilemeyen bir kullanıcı kategorisi olması ile birlikte, afet yönetiminde yardımcı olabilecek bir sonuç değildir. Bununla birlikte, kategori atamalarında detaylandırma ile çıkan kategori çeşitliliği sayesinde, başka bir çalışma yapılabilir ve bu çalışmada arasındalık merkeziyeti tekrardan göz önüne alınabilir.

Bu tezin sonuçlarından sonra, gelecekte çalışmayı daha genişleterek, kapsamlı bir uygulama çıkması hedeflenmektedir. Bu çalışmada detaylı olarak Twitter sosyal meedyasında kullanıcıların ağ analizi ve kullanıcı tiplerinin ağa etkisi incelenmiştir. Gelecekte, her kullanıcı için duygu analizi ve sahte haber keşfi ile, iletilen mesajların anlık güvenirliği ve paylaşım hacmi bulunabilecektir. Bunun için, Türkçe Doğal Dil İşleme kısmına daha ağırlık verilerek, metin kısımlarının da yararlı bilgilendirme ve makine öğrenmesi ile yararlı sonuçlar bulunabilecektir. Aynı zamanda, diğer sosyal medyalar ile de çalışma birleştirilerek, ana kitleye daha yakın bir çalışma yapılabilecektir. Sadece Twitter değil, Facebook ile entegrelenecek bir uygulama ile afet yönetimi için daha çok ve doğru topluluğa ulaşılabilmesi hedeflenmektedir. Bu şekilde organizasyonel bir çalışma yapılabilecektir.

1. INTRODUCTION

Over the past decades, natural or human disasters have increased very sharply. Therefore, the key term of "disaster management" has been important and caretaken. According to the International Disaster Database, it is enough for 10 people to die, 100 people to affected or the country to declare a state of emergency in order to consider a situation as a disaster [1]. After a disaster happens, the disaster-help phase is a very critical process to save lives fast and accurately. There are lots of needs for disaster management to help the casualties. Also, there are many works and articles about disaster management. As time passes by, technology-based solution ideas started to be increased in disaster management. One reason for these solutions increased is that the applications known as social media have been using excessively in the last 10 years. If they can be used strategically, using social media platforms can be advantageous in terms of disaster management phases.

Social media applications help people to interact and share content, communicate easily at any time they want, and work together [2]. Using social media is getting increased very sharply. This leads to investigate and improve the analysis of human behavior. After social media is considered normal in daily life, it has become available for decision making for specific topics. There are various social media applications today. The most used ones are generally Facebook and Twitter [3]. These platforms help people to share photographs and videos, as well as, share their opinions on the story wall.

Twitter is a very used microblog among other social media applications. It is launched in 2006, and since then it is assumed that approximately 8,898 tweets are tweeted each second, and, there are around 145 million active users monthly [4,5]. A tweet is a message area that the users may express themselves within 280 characters (previously, it was 140 characters). They can post images and/or videos as well. Users can follow other users to see their posts. However, users do not have to follow every user who follows them [6]. On Twitter, there is an option to open the account to everyone or make private it. Private account users’ posts only can be seen to the users following them. Since there is a huge number of tweets tweeted each day, Twitter is one of the most used social media to analyze scientific projects. The fact that tweets are the mirror of opinion, these unstructured data can help to analyze the problem and predict the results. From the Science Direct database, for only the year 2020, there are 1,679 articles about Twitter and its data.

In disaster management, social media analytics plays a big role. However, the more these platforms are used, the more possible false news may spread. The increase in social media usage leads also misdirections. In this thesis, we aim to analyze social media in disaster cases and find the type of users who affect the community more and spread accurate knowledge about the disaster. For this reason, we implemented an analysis for a real-life earthquake disaster that happened on 23rd January 2020 in Elazıg city. We used Twitter social media application data for this work. In disaster time, there is a higher chance to reach people via the internet rather than calls or Short Message Service (SMS). Therefore, projects about using social media for disaster management should be supported, since the internet is way handier in such cases.

However, in the last five years, the articles published for scientific purposes have not focused on disaster management in social media enough. In Table 1.1, there is the number of articles published with a certain keyword given in the header, alongside the "social network analysis" keyword. Science Direct database was used to extract the number of published articles for each year. The other majors apart from "disaster management" is also given to compare the importance of the lack of focus on this topic.

Table 1.1 : Table 1.1 : Published article numbers.

Year	Disaster Management	Marketing	Production	Management
2015	8	167	227	479
2016	15	181	251	485
2017	24	174	290	561
2018	20	226	291	588
2019	22	243	344	714
2020	11	107	184	339

In Figure 1.1, the percentages of the article numbers for each year and for each major are shown. Disaster management has not an essential priority for article writers. This article focuses on this shortcoming of work related to disaster management.

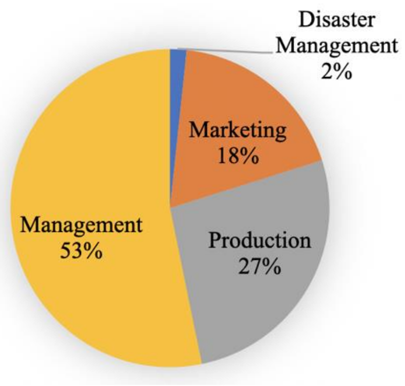

Figure 1.1 : Proportion of article topics published.

1.1 Purpose of Thesis

From the last 4 years of disasters, the internet has had a life-saving effect during the disaster. The recent Istanbul earthquake (magnitude of 5.8) showed that GSM operators are of no use in disaster time. During the earthquake in Istanbul on September 26, 2019, all GSM operators were affected even though there were no casualties or damage. Moreover, some of the operators did not work well for a week after the earthquake. This situation got people to ask whether it is worth trying to call ambulance or police when it comes to disaster moments. In addition, it was seen that the internet attracted very well during the earthquake and that everyone reached people via WhatsApp call or Twitter. GSM operators even stated that they focus on mobile internet rather than GSM functions.

Not long after the Istanbul incident, an earthquake occurred in Elazıg˘ on January 24, 2020, with a magnitude of 6.8. In this disaster, 41 people died and more than thousands of people were injured. There were also thousands of buildings destroyed [7]. During and after the earthquake, people shared information about casualties, help, and shelters via Twitter. Twitter became an information and news source.

This thesis aims to create an influencer-users list as well as deciding the best suitable category of users from Twitter to use in disaster management. In disaster moments, accurate news should be spread in the fastest way to the majority of society. As mentioned before, due to the wide range and number of users of the Twitter platform, important messages in the disaster moment can be spread. We will extract the data from Twitter for the Elazıg earthquake and create social network analysis. Along with the centrality algorithms given in the next chapters, considering their text spreading speed we will show the important users.

The contributions of this thesis to the literature can be summarized as follows:

- In the event of Turkey-focused natural disasters, the impact of social media.
- Implementations of Social Network Analysis, Graph Mining, and Text Visualiza- tion from Twitter data.
- User types which are more effective in disaster in social media.
- Important centrality measurements in terms of disaster-related social media implementations.

1.2 Literature Review

In the literature, there are various Sentiment Analysis (SA) works on Twitter for disaster management. Authors have realized the power of social media and used them. Sentiment analysis gives a desirable output of the emotions of the sentences given. It can be binary results such as 0 for negative, 1 for positive, or it can also be a multi-categorized version which can be given as a scale and the severity of the emotion. However, Social Network Analysis (SNA) or Graph Mining is not used as much as SA's. In Table 1.2, there is a summary table about the papers studied about disaster management using social media.

There are basically three phases of disaster management. These can be counted as preparedness, response, and recovery phases. In the academical world, authors focus on either one of these phases or multi and sometimes all phases. Some papers about disaster management using social media studied in preparedness and early warning phase [8-10], during the disaster [11,12], and in post-disaster phase [13] to analyze the community reaction.

In the paper of [14], authors wanted to see whether they can extract three phases of disaster management from the tweets. For that, they have gathered data for three hurricane occurrences and then manually label them as preparedness, response, or recovery. They also used a keyword list to decide whether the tweet is relevant for the specific phase or not. After that, they used machine learning tools to see accuracy. They have found around 85% accuracy.

In another article, it is shown that Twitter is fast and useful for both getting response and recovery. They collected data before and after the disaster, and they proved that Twitter usage mostly with geo info helps to see the real-time situation so that authorized people can respond fast. They have used Hurricane Sandy disaster as an implementation [15].

In [16], authors extracted Twitter data for a certain 10 disaster hashtags which are avalanche, drought, earthquake, flood, hurricane, landslide, tornado, tsunami, volcano, and wildfire. After gathering these data, they have clustered the tweets to use a machine-learning algorithm to see the most negative and positive tweets about disasters. And they plotted the word cloud of the influencers. As a result, the most dominant region and disaster were found. In this way, they proved that tweets and their users can be used to accumulate knowledge.

Also, as a graph mining example, the study applied social network analysis to analyze and comment on disaster moments. They gathered data from Facebook and classified the networks as three entities as individuals, emergency agencies, and organizations. They used 4 centrality algorithms and ranked the users to see relations between entities. These centrality algorithms are in-degree, out-degree, eigenvector, and betweenness centrality. From results, it concluded that social media can be used for emergency management and information posting [3].

In another paper [17], a new approach has been used to analyze Twitter users' behavior in the disaster recovery phase. For this purpose, tweets about Hurricane Sandy were extracted via Twitter API. For tweets, the possible location information also gathered. From the data; demographic, socioeconomic and spatial attributes are constituted about tweet users. A Dirichlet regression model has been applied for both disaster experienced and non-disaster experienced users. From the result, it is shown that in the disaster recovery phase, disaster experienced users tend to tweet about the disaster and they think the disaster (Hurricane Sandy) will affect their daily life. However, the non-disaster experienced users did not care about the disaster and they thought their daily life will not be affected at all.

In [18], a proposed framework was used to understand the behavior of social media groups in disaster moments. They created two main purposes; group behavior interactions and measurement of the tweets related disaster. Social Network Analysis was applied and nodes separated into 5 groups. With 4 centrality functions, the top 20 users showed the relations between groups. Tweets gathered from the groups were clustered by their semantic quantities and they have visualized the words more used. As a result, the individuals and the other group users have related interactions in disaster problems. Also, [11] studied the propagation of information on social media during the disaster. In the paper, tweets and users' information have gathered for the time Cindy storm. They investigated the functioning of social media and the reliability of it. Hence, social network analysis and thereafter text analysis were applied. Users are formed as groups of 6 and their interactions and dominancy were examined. After SNA, they also created word clouds and the frequency of words used right before and after the disaster. From the results, news and weather agencies are the main actors in spreading the information, whereas the public group is found as generally observer and retweeter for the disaster follow-up. Other additional articles about this topic [19-21] are given in Table 1.2.

The subject and course of this thesis are similar to articles [3,18,11] whose work have been summarized above. However, at the end of this thesis, we aim to interpret whether there is a certain category group that we can use in disaster management. We aim to find such a category of users that can help to spread accurate information to the community before false, or fake news diffuses during the disaster. At the end of this thesis, we aim to decide whether we can use social media for emergency management.

1.3 Research Question

In this thesis, we aim to gather information and answers the questions below:

- Which type of users has more impact in disaster moment?
- Which type of users spread the information faster?
- Can social media be used for emergency management strategically?
- Does social media react dramatically in the disaster phases?

Table 1.2 : Article summaries.

#	References	Type	Approach / Tools	Environment	Libraries	Social Media Used	Data Size	Situation
1	Khaleq, A. & Ra, I., 2018	Book Chapter	Cloud-based System Framework Classification	Microsoft Azure ML Studio	Gnip Twitter API	Twitter	2,311	Hurricane Mathew, 2016 Harvey & Irma, 2017
2	Zou L. et al., 2018	Published Article	Lexicon Based Approach Ratio/Correlation	Mongo DB	archive.org	Twitter	4 Million	Hurricane Sandy, 2012
3	Şimşek, M. U., 2012	Master Thesis	N/A	C# SPSS	Twitter API	Twitter	1.9 Million	Turkish Stock Market
4	Savaş, S. & Topaloglu, N., 2019	Published Article	N/A	Python R	Twitter API Zemberek	Twitter	158,463	Terror Attacks
5	Aziz, K., et al., 2019	Article Notification	Bayes' Theorem Naive Bayes Classifier	R Studio	Twitter API	Twitter	32,117	Hashtag: avalanche, drought, earthquake, flood, hurricane, landslide, tornado, tsunami, volcano, and wildfire
6	Kim, J. & Hastak, M., 2018	Published Article	Centrality Measurements	N/A	N/A	Facebook	1,171 Nodes 27,515 Edges	Louisiana Flood, 2016
7	Jamali M., et al., 2019	Published Article	Drichlet Regression	N/A	Twitter API	Twitter	26,148 Nodes 1,911,733 Edges	Hurricane Sandy, 2012
8	Lu, X., & Brelsford, C., 2015	Published Article	New Framework	N/A	Topsy API	Twitter	14.2 Million	Earthquakes and tsunamis in Japan, 2011
9	Kim, J. & Park, H., 2020	Published Article	LDA-based Topic Model. Proposed Framework	Python	Twitter API	Twitter	18,499 Nodes 21,330 Edges	Hurricane Harvey, 2017
10	Carley K., et al., 2016	Published Article	N/A	N/A	Twitter API	Twitter	13,670,165	Tsunamis in Padang
11	Xiao, Y., et al., 2015	Published Article	MMAM Model	N/A	N/A	Twitter	1,763,141	Hurricane Sandy, 2012
12	Landwehr, P., et al., 2016	Published Article	TWRsms	Python	TweetTracker	Twitter	2,617,208	Tsunamis in Padang
13	Kim, J., et al. 2018	Published Article	Centrality Measurements	R Studio	Twitter API	Twitter	14,094 Nodes 16,269 Edges	Storm Cindy, 2017
14	Chen, W., et al., 2020	Published Article	Centrality Measurements Clique Analysis SVD Analysis	Gephi 0.92, CFinder, UCINET 6, and Ora 3.0	-	Search Engine	3,443 Nodes 3,852 Edges	Wenchuan Earthquake, 2008

2. SOCIAL MEDIA ANALYTICS

Social media is a technology to help the users interact with each other and share their thoughts, beliefs, interests using internet Web 2.0. It varies according to users' aim to use. It can be either private chat applications (WhatsApp, FB Messenger, etc.) or it can also be a platform where users can share their knowledge and thoughts with everyone (Facebook, Twitter, etc.). Contrary to popular belief, the concept of social media has been around for a long time. However, in the 2000s, social media has been a preferred platform where people interact with each other. From the individual to the organization, users started to share all their announcements, information, and opinions by using social media.

Usage of social media like Twitter is easy and most requested. There are good reasons for this state. The speed of incoming data from social media each second is over hundreds of thousands, and also this helps to get normally distributed crowdsourcing populations. Mostly social media platforms have a mobile device option as well, and this helps to get a quick reaction for a specific event. Statistically, the average number of messages coming from social media is 500 million [4]. From this, it is obvious that the amount of incoming data from social media is tremendous. This data greatness help scientist to work on projects such as understanding disaster management, user reaction, impact analysis [18].

From the article of [22]; social media is divided into 4 dimensions in issues such as disaster management. They may call as area-zone, time-zone, content, and network. These four dimensions; individually, simultaneously, or separately-combined analyzed, and any scientist may work whichever dimension combinations they want. For Twitter data, these dimensions may call as place/locations, created_at, tweet, and retweet/reply/like respectively. These dimensions' summary of explanation is given below:

Place/locations: In Twitter data, there are two ways to extract space information. One is the exact coordinate information which gives longitudes and latitudes of the user's location. The other one is toponyms such as place name that the user gives. For the coordination duals, they give a very accurate conclusion, however, it is very rare to find user's coordinates since there is a necessity for that like using mobile devices, and positioning systems should be turned on. As the other way -the place names-, it comes from the user's profile page which means that the user writes where she/he is by hand. This may lead to misdirection. A user may write "Istanbul" in her toponyms, but she may also write meaningless descriptions such as "top of the world".

Time Zone: All tweets have their "created_at" tag in the type of date and hour. This helps to track the basic hashtag about a disaster and see the changes over time. Some works showed that the frequency of tweets about a specific event increases in response time and then decreases again.

Content: For Twitter, tweets have critical meaning for disaster management. Tweets can give important vision and help to monitor the situations in disaster phases. There are a few types of text classifications. They can be classified as either topic-based or sentiment-based. With topic-based, it is easy to distinguish important topics from the word vectorization and relations. In sentiment-based, the sentiment of the texts may give an aspect of the user-based opinions.

Network: The last dimension was the least mentioned and applied one. Network dimension is very crucial for disaster management since the information spreading is life-saving if it reaches the majority of the users. In this dimension, the user information and their response for the other users can show the relations and importance of the user.

In this thesis, the main purpose is to create social network analysis that is related to the "network" dimension of social media data. The combination of the other dimensions will also used as a sub-factor.

Social media analytics (SMA) is a way to understand the interaction in social media. It is the whole process of obtaining and processing the data accumulated in the social network and inferencing from this data in accordance with the desired or expected result. Since it involves statistics, computer science, sociology, mathematics, and many other fields, it can be said that social media analytics is an interdisciplinary science.

All social media contents come from individual users, that is why it shows that users express their feelings without filtering. This makes social media unique and special. It helps to get very important information. It is an important source of information as well as the difficulty of processing text, videos, and photos created by users.

2.1 Social Network Analysis

In favor of the interactions in social networks, many inferences and information can be obtained about the users in those networks. These network analyses can give answers to multiple questions such as:

- Who are the most powerful/important/influential users?
- What are the roles of people on the network?
- How is information spread across the network?
- How can social communities in the network be identified?
- What content are people interested in?

The answer to all these and many other questions can be given through social network analysis (SNA). All investigations of users and their interaction on the basis of relations are called SNA. In other terms, Social Network Analysis (SNA) pays attention to find out the relations between users. These users could be individual person, groups, community, organizations, etc [23].

SNA is a sort of sub-branch of social media analysis. It deeply focuses on social media networks and the relations between users. It can be useful to analyze specific events or occasions. Especially in disaster moments, SNA offers a big opportunity to manage.

In SNA, there are two key terms: Node and Edge. These are denoted as users and the relation between two users in the network. Nodes are users whose behavior is analyzed within the network. The purpose of the analysis is to find the answer to the question we are looking for with specific calculations for each node. The edge is called the defined relationship between the two nodes which are connected within that network. The edge may vary differently according to the network type. For example, on Twitter, all of the retweets, mentions, comments, followings between two users are edges separately. As such, each of the interactions between the user on Facebook is an edge. Both edges and nodes can have attributes. It is not necessary to have it, but the attributes give unique information for the node or edge that it represents. These attributes may vary such that it can be the date, number, name, or even picture.

In literature, the network where involves all nodes and edges is called a graph. The graph helps to calculate the outputs. It is in one sense a mathematical representation of the network [24,25]. To generate the mathematical results for a network, there should be a represented graph that has nodes and edges inside of it.

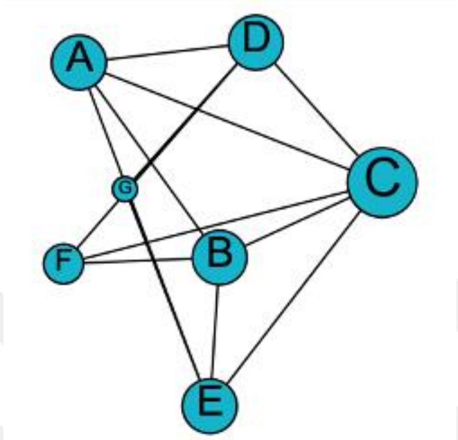

Figure 2.1 : Example of a graph.

In Figure 2.1, there are 7 nodes (A, B, C, D, E, F, G) and 13 edges. This is a plain sample of a network graph. Assuming that the graph's name is assigned as G, the graph is defined as G(V , E) where V denotes each node i ni and E denotes all edges between node i and node j ei j in the graph. In this example, the graph G has a set of nodes V = A, B, C, D, E, F, G and set of edges E = (A, B), (A, C), (A, D), (A, G), (B, C), (B, E), (B, F), (C, D), (C, E), (D, G), (D, F), (F, G).

The edge distribution gives an opinion about the graph given. Since the edge means relation or interaction between pair nodes, we can also say that the more edge a node has, the more interactive it is. In Figure 2.1, the nodes' sizes proportioned by their relation number. Node C has the highest relation among all, so the biggest node is C. The detail about sizes, relations, and calculations of the graph will be given in the following chapters.

2.1.1 SNA tools

In real-life problems, the nodes and edges in the graph will be more than thousands. Calculations for each node and the whole graph will be neither easy nor fast for human beings. In this situation, software programs or libraries help to analyze fast and accurately. There are tens of tools found on both the internet and literature that can be used for SNA. However, most of these software are not current and inclusive. In [26], authors emphasize the noisy crowd in SNA tools and a list of them. They also provide fifty-six selected programs and tools for social network analysis.

There are basically two categories as an SNA tool. There are open-source software and commercial/academic application software. Some of the most popular toolkit and their properties are given in Table 2.1.

Table 2.1 : SNA software tools.

Tool	Type	Language
Ucinet	Application	-
NetMiner	Application	Python-based
Gephi	Application	Java-based
Igraph	Open Source	R
NetworkX	Open Source	Python
ORA	Application	-

Since there are so many application attempts to make social media analysis, some features have started to be searched for when choosing these applications. As mentioned before, for the real-time problems, we are facing with hundreds of nodes and edges, and this big data may force the tool and the result may take a long time. Additionally, after the calculations, visualization is a significant factor to decide.

In this thesis, we will use Gephi and Python software to analyze and visualize the networks.

2.2 Graph and Measures

As mentioned before, the graph is a mathematical representation of a network. Nodes and edges are mandatory elements of the graph. Node and edge information is sufficient to analyze a social network. However, some information about the graph constitutes different analysis methods. There are two basic graph types. These are called a directed graph and an undirected graph. In the directed graph, the edge is oriented in a certain direction. It does not have to be driven back by the node it is directed to. The basic example of this graph is Twitter where one can follow a user regardless of being followed back from that user. On the other hand, in the undirected graph, the edge is symmetrical and bidirected. An example of an undirected graph is Facebook where once a user becomes a friend with another user, they become friends together.

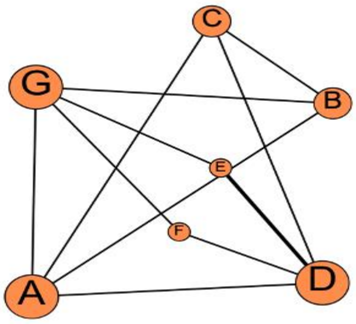

Figure 2.2 : Undirected graph.

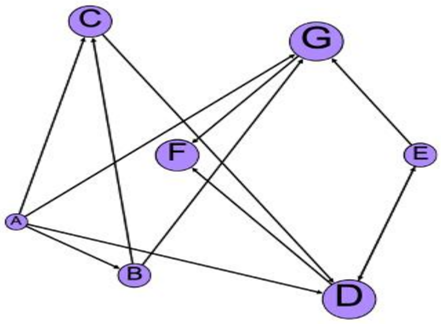

Figure 2.3 : Directed graph.

Apart from graph type, there is another concept to consider. The weight of the edges may vary according to the density of the relation for two nodes. Weight is a numerical representation of two nodes' relations. For example, the comment number for a post of a user may be calculated as the weight of the edge. As a more detailed example; on Twitter, let's assume that node A replied node B 3 times, and node C 1 time. This number can tell us about the relation between the users.

In Figures 2.2 and 2.3, undirected and directed graph examples given respectively. Node sizes were arranged according to incoming edges attached to the node. Both graph's edge and node table are the same, but deciding the direction affects the graph visualization and calculation properties. Weight attribute's action changes according to graph type. In Figure 2.2, the edge thickness changed between node E and node D, because in the edge table these nodes interact twice. However, in Figure 2.3, this thickness and weight do not appear. The reason is that in the edge table, node E interact with node D, and node D interact node E only once. In an undirected graph, this is counted as two, but in the directed graph, it is counted as bidirectionally one. This does not mean that directed graphs have no weights. If there would be more interactions between node D and node E or vice versa, the edge thickness would increase associated with the weight value.

When an edge starts from one node and ends in the same node, this is called self-loop. In some networks, self-loops are not used e.g. a user will not follow itself on Twitter. However, in some cases self-loops are acceptable. A person can email to herself. So, the self-loop case varies according to the network type and its purpose.

There are different ways (such as adjacency matrix, adjacency list, incidence matrix, edge tables, etc.) to represent graphs. The aim of these display figures is to give the mathematical format of the network to the graph. All of these ways result in the same if the edges are identical. These options will not be explained in detail here.

2.2.1 Centrality measures

There are some metrics and calculations to analyze both the whole network and its nodes in SNA. These calculations are in the form of numerical results for comparing the nodes in the networks. The concept of centrality is essential to determine the most important nodes according to the intention of the network. Centralities affect each individual node inside of the network and show how central the node is. They determine how important, influential, popular, or connected the node is [23,27].

The important thing to know about centralities is that each centrality measurement gives a different aspect of importance. In the analysis, the criteria are determined according to the purpose of the analysis, and making interpretation of these criteria provides a more accurate output. It also must be known that the calculated centrality criteria can only be compared for nodes in that network. Centrality criteria values will change in another network. Each centrality criteria below address different parts of importance in the network.

2.2.1.1 Degree centrality

Total degree centrality is calculated for each node basically sum the all connected edge to that node. According to the type of the network and correspondingly the graph, degree centrality is separated as in-degree and out-degree centrality. While in the undirected graphs, degree centrality is calculated as total degree centrality, in the directed graphs it has in-degree and out-degree calculations. Because there is a direction in the directed graph, the total incoming edges to a node are in-degree, and the total outgoing edges from the node are out-degree centrality of that node.

$$C_D(n_i) = \sum_{j=1}^{n} E_{ij} \qquad (2.1)$$

In Equation 2.1, the degree centrality calculation is given, where (n_i) is the node i we are calculating and E_{ij} is the edge from node i to node j for every node in the network. E_{ij} is a binary value, and if there is an edge between node i and j E_{ij} is 1, otherwise, it is 0. N is the total number of nodes within the network. For the in-degree and outdegree calculation, the same equation is applied. Only the formula representation changes to C_{id} and C_{od}.

This centrality indicates that the higher number of degree centrality, the more important the node is. In in-degree centrality, this representation shows as more prominence, while in out-degree higher centrality gives the more social users.

2.2.1.2 Closeness centrality

Closeness centrality calculates the shortest path from each pair of nodes to every other node in the network and takes the reciprocal of the average shortest path of the node. This centrality shows how close the node to the other nodes in the network. Hence, this measure is the only centrality where the smaller calculation is better. This means that according to this centrality, the more central a node, the faster it can reach all other nodes.

$$CC(n_i) = \frac{N-1}{\sum_{j=1}^{n} d_{ij}} \tag{2.2}$$

In Equation 2.2, N is the number of nodes, and d(i,j) is the shortest path between node i and j given in the network.

There is a very similar centrality to closeness centrality which is called Eccentricity Centrality. The only difference from closeness centrality is that d(i,j) is the maximum shortest path between one node to other all nodes instead of the average shortest path.

We can give a simple example of these centralities as creating a hub where all parcels must deliver to customers as fast as possible. Closeness and eccentricity centralities show the optimum node for the hub. For social media, this centrality could be useful to find a node that shares the information to all other nodes fast.

2.2.1.3 Betweenness centrality

Betweenness centrality looks at whether a node is in the shortest path that two different nodes have. Differently from closeness centrality, it does not look for the shortest path, it counts and sums the paths. It becomes in some way a bridge among the network. It looks for the shortest way in every pair of nodes, and then for each individual node, the formula checks whether there is that node for the node pairs. In this way, it captures the bottleneck of the network. It is a measure of a node’s influence. It measures the gatekeeper node in the network [11].

Assuming there are 2 shortest paths between node A and node B, if node C is included in one of the shortest paths between node A and B, node C's betweenness measure is calculated as where 1 is the number of showing of C in the path, and 2 is total shortest path number between node A and node B. For node C, each pair are calculated and a total of the measures, the result becomes node C's betweenness centrality. As it is seen from the explanation, the division cannot be a compound fraction. Nodes with a higher betweenness centrality tend to be a really important node in the network that controls the flow of the discussion. It also the one which connects different groups through the network.

$$C_{BTW} = \sum_{i=1}^{n} \sum_{s,t \varepsilon N} \frac{\delta_{s,t}(n_i)}{\delta_{s,t}} \tag{2.3}$$

In Equation 2.3, n is the number of the node in the network, $\delta_{s,t}$ is the number of shortest path between s and t, and $\delta_{s,t}(n_i)$ is the shortest path between s and t passing over node i.

This centrality is one of the most used. The calculated measure gives the node which controls the flow of the information over the network. However, since the formulation takes into account the shortest path for all pairs in the node, its calculation can be slow. This problem can also be said for closeness centrality.

2.2.1.4 Eigenvector centrality

An eigenvector measure is a slightly different centrality from among all others so far. This centrality shows the importance of the node. In eigenvector centrality, the node is important not only by the number of its connected nodes but also its neighbors' reputation in the network. It creates a matrix where a node that is linked to by more influential nodes gets higher weight. This method believes that if a node is important, its connected "friends" are highly important in the network. It is calculated with eigenvalue in linear algebra on the adjacency matrix for each node and its neighbors.

$$C_{EIG} = \frac{1}{\lambda} \sum_{j \varepsilon G} \alpha_{i,j} C_{EIG}(n_i) \tag{2.4}$$

In Equation 2.4, λ is the eigenvalue in the adjacency matrix, αi, j is the binary value for connection between i and j if there is an edge between node i and node j the value is 1, otherwise it is 0.

Eigenvector centrality is important for the network where influential nodes should be found. From the calculations, higher eigenvector value is the more influential and prestige a node is in the network. There is another centrality called PageRank Centrality which acts the same with the eigenvector centrality. It is used on the Google web page ranking.

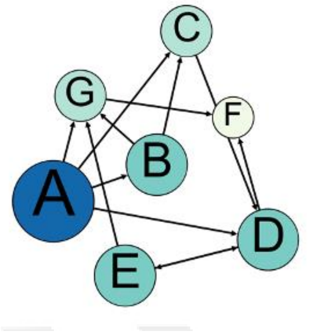

Figure 2.4 : Outdegree centrality.

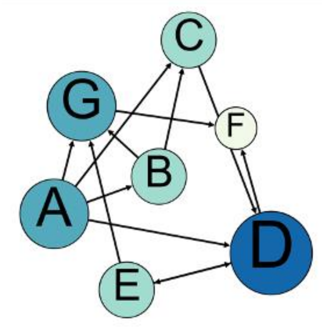

Figure 2.5 : Degree centrality.

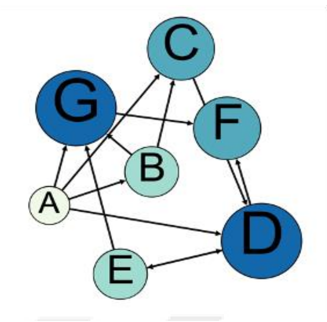

Figure 2.6 : Indegree centrality.

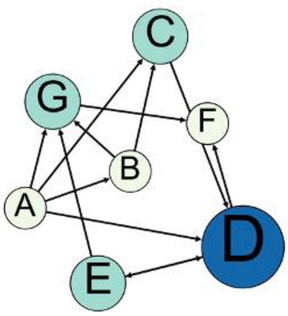

Figure 2.7 : Betweenness centrality.

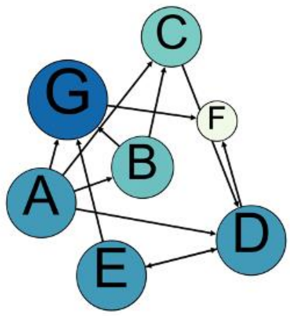

Figure 2.8 : Closeness centrality.

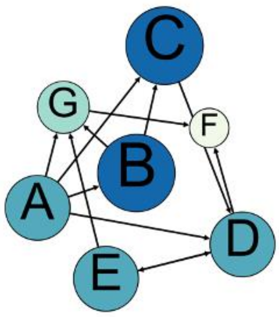

Figure 2.9 : Eccentricity centrality.

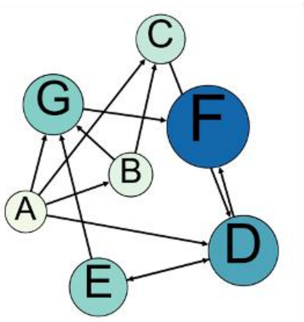

Figure 2.11 : PageRank centrality.

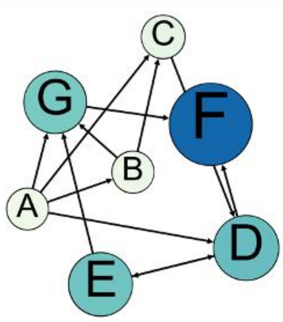

Figure 2.10 : Eigenvector centrality.

From Figure 2.4 to Figure 2.11, all types of centrality mentioned above were applied to the same node-edge adjacency matrix. The node color and node size were adjusted by centrality applied. From the results, it is seen that each node has a different outcome and importance. That is why knowing the purpose of the network and problem is very essential to get more accurate results. In some cases, different types of centralities can be compared to get healthy interpretations. However, from the figures, it is easy to see that some centralities are very similar to compare. From the experiments [28], it is given that degree centrality is strongly correlated with betweenness and eigenvector centralities separately. There are other correlated centralities like eigenvector and PageRank centralities, closeness and eccentricity centralities, etc. This correlation information helps us to which centralities we should combine to compare. Since high correlations would give the same results we cannot have a unique comparison. There are other centralities to measure, but the commonly used centralities are given above. These centralities can help to analyze the network at the desired value.

Last but not least graph measure to explain is graph density. The graph density shows how connected the network is. It is measured by the fraction of the total edge number to possible maximum edge number that the graph may have. Edges are important for this measure. For the directed and undirected graphs, the graph density calculations are given in Equation 2.5 and Equation 2.6 respectively. It is known that in real-life examples the possible measure of graph density is almost zero. Even though there are thousands of nodes, the edge number is not close to the possible edge numbers at all. The reason is that there are always outliers in the network which cover most of the network and these users do not form any interaction, they usually only watch the flow of the discussion e.g. they do not have edges.

$$\frac{m}{n \times (n-1)} \tag{2.5}$$

$$\frac{m}{n \times (n-1)/2} \tag{2.6}$$

As we can see from the equations, while the number of nodes is fixed, the graph density is higher in the graph with more edges.

2.3 Information Spreading

Detection of a user who influenced society more and disseminated the topic is getting essential as social platforms are daily usage in our lives. The information spreading term is easy to understand. It studies the process of spreading information among any communication tools. Each different type of field is interested in seeing the spreading flow, speed, and users for specific content. Such as, the marketing field could use information spreading to see which product of the company should be targeted to innovate, the media field could use it to decide which category they should focus on to get the attention of their subscriber or increase their subscription number, etc. Apart from the private sector and business purpose, information spreading can be used in disaster management to determine who should be chosen to influence the community to spread information disaster-related messages as far as possible [29].

Especially, for the recent 15 years, all big and effective disasters have been discussed in social media by millions of people more than traditional media such as televisions, radios. These discussions bring ample data that saves very subjective information and discussion flow from users. There are plenty of studies about this data to see social media impacts on this kind of situation. As mentioned in the previous chapter, from these studies general idea is comprised that people use and react to events in social media much enough to understand, analyze, and interpret on behalf of disaster management [30].

In Twitter, each of the attributes of the retweet, mention, and the reply is a way to see the flow of information spreading. Most of the studies focused on retweet attribute, however, in this thesis we will focus on mention and reply attributes. Information spreading flow will also give influential users, and this will help to assign important users/people in terms of disaster management. The basic structure of the information spreading sample is given in Figure 2.12 [31].

From the figure, it is shown that the information is started from a user and every time the content gets commented, shared, forwarded, or mentioned the spreading is comprised. The speed can be calculated as the differences between the last user's sharing and the first user's sharing. In the example, the information is spread three times (3 depths) and the speed is 16 hours and 10 minutes. In [31], they indicated that almost 95% of the data has a depth of 2 which means that the most important and useful data is comprised of 5% of the whole data. This knowledge makes sense since the real-world data has a big proportion of noisy data and irrelevant.

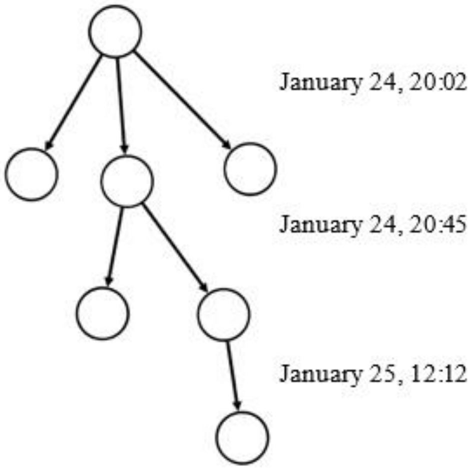

Figure 2.12 : Information spreading structure sample.

In SNA, there is no exact and official calculation of information spreading measurements. However, we can say that influential users help to spread the information fast and widely. Hence, centrality measurements which give the most important node as influential can be used. Eigenvector centrality may help to extract the information spreader in addition to influential users.

2.4 Disaster Management

A disaster by definition is the impact of a natural hazard with a casualty, damages, and/or chaos. It happens quickly and in an instant that has no warning before. This situation leaves authorities with no clue to foreseen [32]. Since natural disasters are inevitable to happen, the critical behavior is to prevent and overcome them as fast and with less casualty as possible.

The term disaster management is a type of management with certain steps that came out to help reduce these difficulties. In disaster management, there are generally four phases: Prevention, Preparedness, Response, and Recovery. The purpose of disaster management is to make conscious and preparedness for disaster by making plans before the disaster, to make the most efficient help and management to reduce the losses during the disaster, and to make important decisions to provide quick support to casualties after a disaster.

Social media usage effects to decide on situations. Especially in marketing, for a specific product they measure social media users' reactions. Social media platforms have countless attributes to extract and analyze. By knowing that millions of people use social media during the day, inferences can be made with this collected data. And analyzing these data not only for marketing but also for disaster management makes sense. From the numerous papers, we see that the relationship between social media usage and disaster management has studied and the results show that this relationship helps to improve emergency management [33]. Another benefit of using social media besides data size is that there is various type of users as individuals or organizations in social media. This helps to get real life-based feedbacks in disaster moments. In managing the disaster, these differences help to disperse the information, and if the management is made accurately, spreading accurate information to each person in the community can be easier.

The effectiveness of social media usage in disaster management depends on the quality and quantity of data extracting. Keywords, time-intervals, users, location knowledge, text affect the quality and necessary data we need [34]. Since we mentioned that data in social media is enormous, the relevant data from this pool is very essential to get meaningful results. The time range is given around the disaster date plus/minus 2-3 days as wanted to see both before and after a disaster. Keywords are important to get a relevant topic on these date ranges. And the location is also important to see local users' reactions.

After setting up the proper attributes, there is another important key: quantity. Big data has an essential aspect to understanding approaches accurately. There are two difficulties to get more data: privacy terms and location information.

Privacy became a very important matter particularly in the last 5 years. With the GDPR law, which was adopted with the priority of the European Union, a law was created in which customers, including social media, have the right to protect certain data. In this context, users' information will not be captured by any social media API without the consent of the users or by default. Another way to protect their information, the user can make private their account so that nobody can see their information unless they are not the user's friend. However, this is not a very robust way of protection, because the person who pulls the information can be the friend/follower of that user.

Location information brings another difficulty if the research is made with location extraction. From the researches, it is known that only 1% of all Twitter data was geotagged in 2015 [35]. There are different types of location information like geo, and place. However, the real accurate knowledge can come from geotagged since it gives directly longitude and latitude information while place information only gives users' expression. With the data collected with these features, disaster management will be much faster, more accurate, and successful.

3. METHODOLOGY AND IMPLEMENTATION

In the implementation part, there are roughly two stages. Data retrieving, cleaning, processing, and visualization were done in Python language while SNA implementation, graph mining, and centralization measurements were done in the Gephi application. The reason we used Gephi for the SNA process is that the efficiency of Gephi is way faster than Python libraries.

In Figure 3.1, the steps for both stages are shown. Visualization has two different parts and variables. First, we will visualize the raw data retrieved from Twitter, and then we calculate the centralizations and visualize the centralities and see which users and attributes have more impact.

In the implementation steps, data preparation has very important impact. Data retrieving has brought so many attributes, and deciding which attributes are important to use for the aim of the study is very essential and critical. Therefore, the first four steps after the data store, are not only descriptive parts of the implementation, but also strategically decision part. To get appropriate decisions in the result part, we have created six categories where each user was assigned one of them. After preparing the data, another step was to prepare the necessary attributes for the SNA. In this part, usernames were accepted as nodes, and in each row where there is a reply interaction between two usernames, an edge was created. A node-edge list fulfilled the main implementation necessities. The Node-edge list has stored in CSV file extension which Gephi application also supported. After the first part of Python language implementation, the data became ready with node and edge list, including their node attributes and edge attributes. Social network analysis and centrality measurements were made and the result was decided. All detail about these steps will be given below.

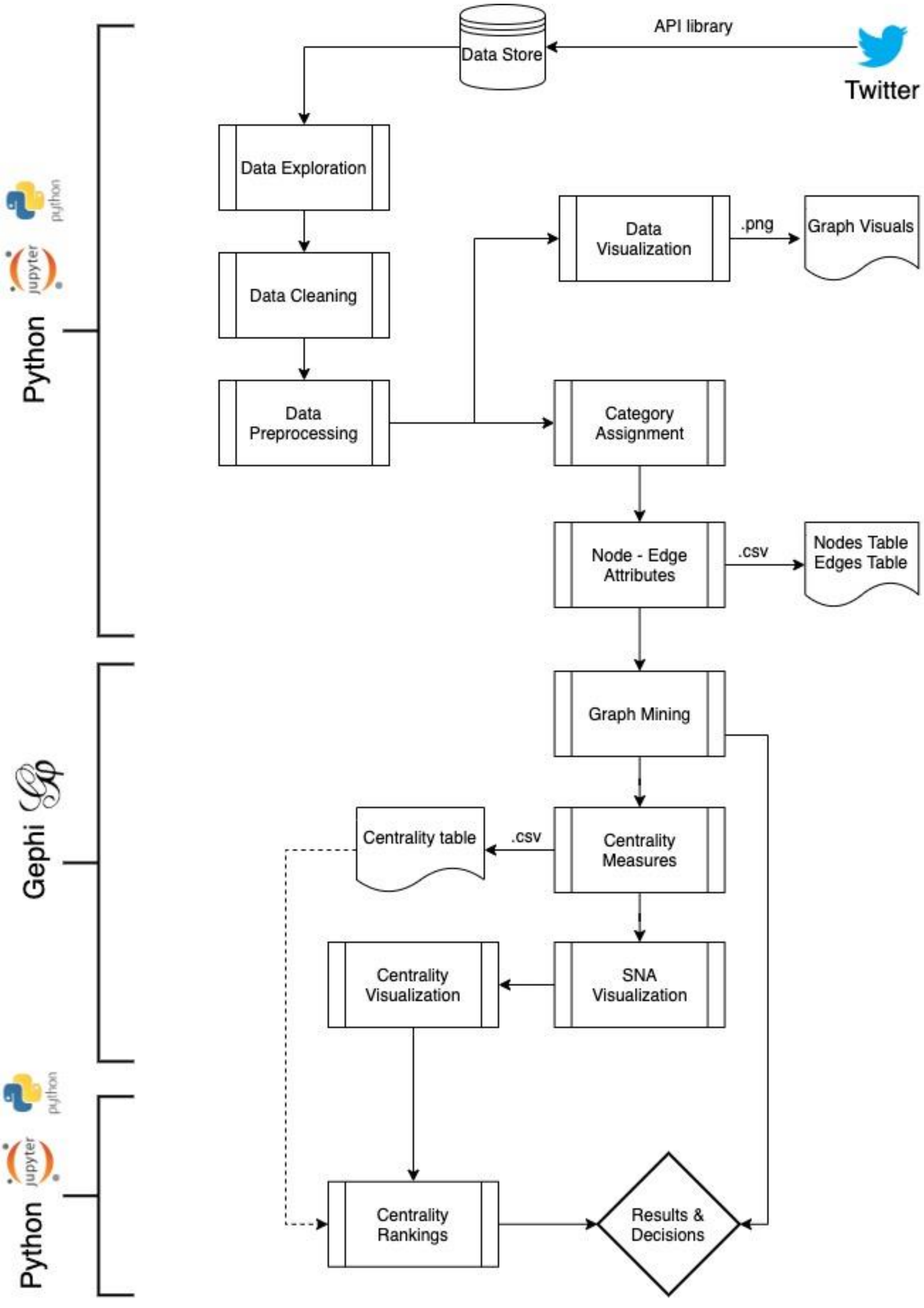

Figure 3.1 : Flow chart of the implementation.

3.1 The Incident

The disaster we are focusing on this thesis is called the Elazig earthquake in literature. It happened in Elazig on January 24, 2020, around 17:55, and the quake took about 40 seconds. The depth of the earthquake was 8.06 km., and the magnitude was recorded as 6.8 [36]. The earthquake's epicenter was in Sivrice which is a town in Elazig, and Sivrice is very close to Malatya which is another city in Turkey. Since the fault line was very close, the earthquake was felt in Malatya as well.

76 buildings were demolished in Elazig and neighboring provinces. During and after the earthquake, a total of 41 people died, 4 of them were in Malatya and 37 of them were in Elazig. Apart from these, a total of 1466 people were injured. After the destructive big quake, a total of 1140 aftershocks occurred, 13 of which were over a magnitude of 4.

This earthquake is the largest earthquake since 2012 based in Turkey. In all social media from all types of users (individuals, institutions, statesmen, etc.) reacted. From international media and statesmen conveyed their condolences.

3.2 Data Retrieving

To examine the problem, 7 days of data were retrieved from Twitter. The incident happened on January 24, 2020. The data includes 3 days before and after the disaster date. By doing so, we intended to analyze the behavior of users towards disasters in detail.

Twitter API could not be used since data acquisition was made a certain time after the incident. Twitter has three different API endpoints of itself which are Standard, Premium, and Enterprise endpoints, and they have some differences in using API. Standard API users have 7-day limits to get data while Premium and Enterprise have 30-day. Premium and Enterprise have another full-archive option where user can get all data between any desired dates

In this work, twint library was used to retrieve data. Twint [42] library is a Python tool that allows us to scrape Twitter in an advanced form. This tool doesn't require to use Twitter API and it allows us to extract all types of user, follower, following, tweet attributes while sparing API limitations. The collected data is 304,374 rows with 24 columns. Each column represents the Twitter attributes collected from one specific interaction. The data stored in .csv format.

From the collected data, we will create graph mining. In this graph mining, the vertex represents Twitter users while the edge represents the relationship created between two nodes. In Twitter, there are various different ways to create relations such as retweets, mentions, likes, and replies. The relationship type used in this study is the reply. We will use replies between two users as a relationship for graphs.

3.2.1 Data attributes

As mentioned before, Twitter has an enormous data comes each second. Data can become noisy and unnecessary if they retrieved without any filter, and it affects the analyzes.

To get the data, we filtered them by keyword and date variables. As keywords, we have used:

- "#deprem OR #elazig" as search keywords.
- 21/01/2020 - 27/01/2020 as date ranges.

From collected data, 24 attributes which are added as columns in table extracted. These attributes are: id, username, conversation_id, date, time, timezone, user_id, place, tweet, mentions, photos, replies_count, retweets_count, likes_count, hashtags, retweet, link, video, user_rt_id, near, geo, source, retweet_date, reply_to. Among these, some attributes dropped since there is no informative data.

In data visualization and exploring parts, the attributes of username, date, time, tweet, replies_count, retweets_count, and likes_count likely used whereas in SNA part, additional created attributes and reply_to attribute was added.

3.3 Data Pre-Processing

After extracting Twitter data, we processed it before analyze to get more accurate results. Data pre-processing part split into 5 steps:

1. ***Text Cleaning:*** One of the attributes in the data frame is “text” which refers to tweets. These tweets can give invaluable information about the incident, people’s behavior, social media affects, etc. There will be 5 steps for the text cleaning part as well:

– *Mention removal:* Mentioned users are shown with the @ sign. Mentions were extracted from the texts. there is a “mentions” attribute that gives a list includes the mentioned users if there is any for each row (e.g. @afad).

– *http link removal:* Links used in tweets will be irrelevant in visualize part. Spaceless characters starting with http were removed. Instead, there is a “link” attribute that gives a list includes the http link if there is any for each row (e.g. https://twitter.com/home).

– *Case sensitive:* In analyze, each word in the text were lowercased (e.g.Twitter → twitter).

– *Punctuations and Numeric removal:* All characters and numeric were removed (e.g. 5 dakika önce deprem mi oldu???? → dakika önce deprem mi oldu).

– *Stopwords removal:* Unnecessary words for analyzing were removed from the text. In this section, redundant words not related to natural disasters were deleted along with conjunctions and prepositions. In this way, sensitive and important words can be spotted more easily. Our aim is to work with only important and vital words e.g. deprem hakkında afad önemli açıklama yapıyor → deprem afad önemli açıklama).

2. ***Hashtag Category:*** In the text part, there are various different hashtags. Some are important and some are irrelevant to the situation. To eliminate this diversity, we created a function that returns only #deprem or #elazig for each row if there is a close hashtag related to either of them. The word “deprem” means an earthquake in Turkish. For these hashtags, the function looks for the closest hashtag words such as:

- #deprem: #deprem, #earthquake, #dprm, #deprm, #dprem, #quake
- #elazig: #elazig, #elezig, #elazığ, #elazg, #elzg, #elaziğ, #elazıg, #elezığ

3. ***User Category:*** We created another function to create a category by user screen name. User behavior will be analyzed by user categories. 6 different categories were created and the function is looking for keywords in the username and returns with the related category:

- **Earthquake:** the exact word list of [’deprem’, ’earthquake’, ’quake’, ’zelzele’, ’emsc’, ’kandilli’, ’deparsis’] was created
- **Media:** All similar words about media were created as a list. These words include journal, news, newspaper, magazine, etc. category names in Turkish.
- **Individual:** The most frequently used male and female names list were created. 70 names were entered. We have also added verified individual users such as celebrities to this list. Although they are in the same category, we wanted to see if their behavior would be different.

- **Governmental:** All word and phrases about governmental accounts was created. Also, the specific usernames such as members of parliament, president, municipalities, governorships, etc. were added.
- **Institutional:** Institutional accounts such as private corporates, benevolent associations, schools, etc. were created as a list. The specific usernames and similar words related to this category were added.
- **Other:** We classified the users who could not fall into these categories above as other.

4. ***Unnecessary Columns Drops:*** In collected data, there are some columns -attributes- that is not related directly to the subject. Some of them because of null values, and some of them are not suitable for this study's purpose. These columns were dropped vertically from the dataset.

5. ***Node - Edge Preparation:*** In the SNA part, two sources of data are needed categorically: Source and Target nodes and their relations. In the dataset, each row has a certain username which is called Source Node for the SNA model, but if that user does not reply to anyone, the "reply_to" cell will be empty and this leads that there is no Target Node. In this case, the relation model would not be generated. Also, in the dataset, if a user replies to multiple users, it is given in a dictionary format in one row. Our aim is to create a function that gives one row for each edge.

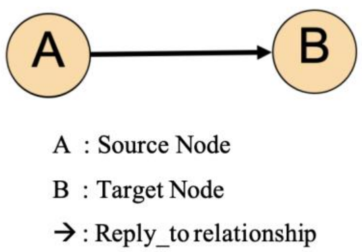

Figure 3.2 : Structure of network.

At the beginning of this part, the total row of data was 304,374. We first expanded the Target Node cell if there are more than one reply_to users in that cell. The length of the data became 657,192. We then drop the empty cell if there is any. The length became 352,818. After these steps, the Source Node and the Target Node were filled. However, there are some rows that both Source Node and Target Node are the same users.

These types of relations are called “Self-Loop”. Self-Loops affects the calculations. Therefore, we deleted rows where both Source and Target nodes are the same users. After this elimination, the length of the data became 48,447.

After all these 5 steps, data changed into the desired table. The steps of implementation will be as visualization of the data, graph mining, centrality measures, centrality visualization, and text mining.

3.4 Data Visualization

Data visualizations help to understand overall aspects. The visualization part of this study has a strong impact on the decision result. In this study, different types of visualizations were made to comment on all attributes.

3.4.1 Tweets plotting

In the text part, visualizations can be made either with text mining or in dimension and descriptive ways. In Figure 3.3, the lengths of the tweets are given. The tweet usage seems in exponential decreasing function format.

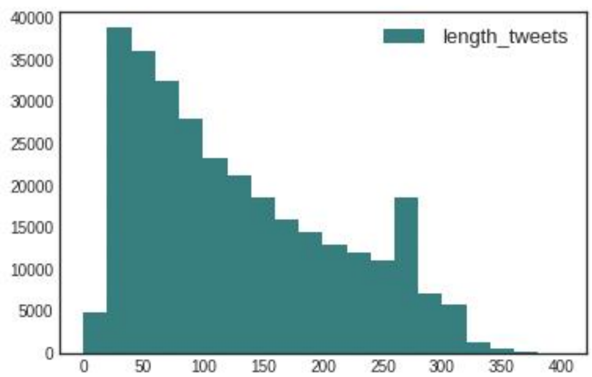

Figure 3.3 : Distribution of length of the tweets.

Apart from descriptive analysis, text mining is the key term for tweets plotting. With text mining, important words and hashtags can be extracted. In Figures 3.4 and 3.5, the hashtags used by users are given. Since the data was extracted with the keyword of #deprem and #elazig, it is understandable that these two hashtags are excessively high rather than the others, e.g. Figure 3.4. To understand local behaves, we removed the first 8 most used hashtags from the list and display them again. In Figure 3.5, the top 20 used hashtags after removing much-used are given. It is seen that hashtags are related to the disaster.

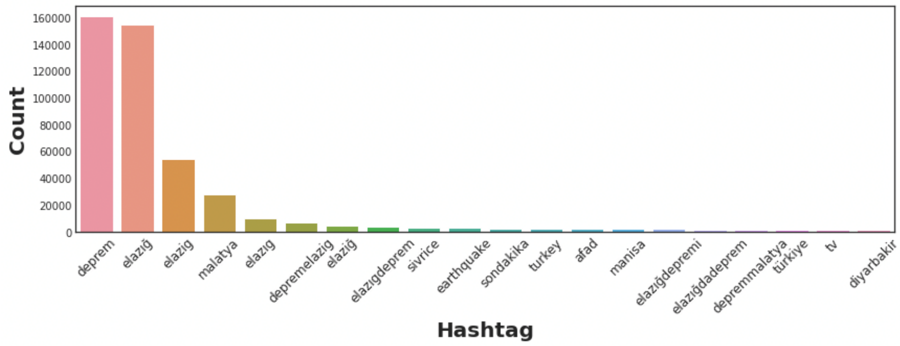

Figure 3.4 : The 20 most used hashtags.

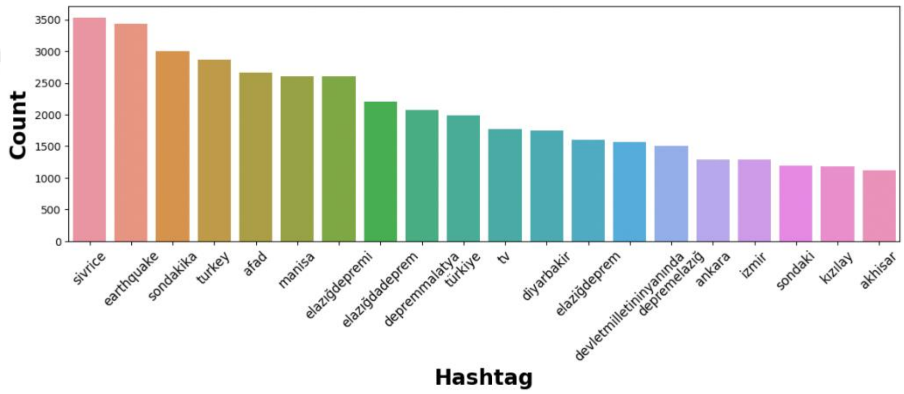

Figure 3.5 : Hashtags after removing excessive ones.

Another way to analyze text for the study is to see important and most-used words in tweets. A word cloud created from the tweets was made. In a word cloud, each word was split and join for all rows. Basically, the function counts each word within all words. By counting, the representation of the word in word cloud resizes. In Figure 3.6, word cloud of the whole tweets in the dataset is given. From the words and phrases, it is seen that most of the words are about sadness and giving condolences rather than giving important messages to earthquake victims.

Figure 3.6 : Word cloud of all tweets.

3.4.2 Users plotting

Users' visualization also gives an important perspective in the analysis. Being mentioned is related to being important in the network. Because getting more mentioned by other users can affect being an important user in the discussion topic. In Figure 3.7, the top 25 most mentioned users are given. From all accounts, it is clearly seen that all verified and well-known accounts are important in social media as well. All top 25 most mentioned users' categories are either "Governmental" or "Media".

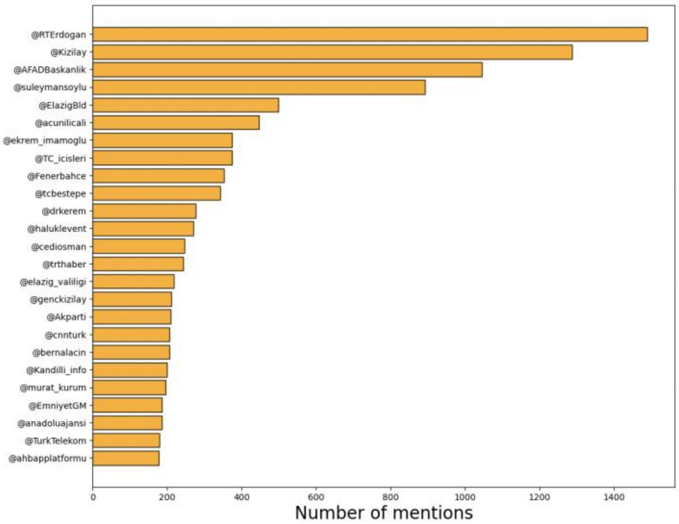

Figure 3.7 : Most mentioned users.

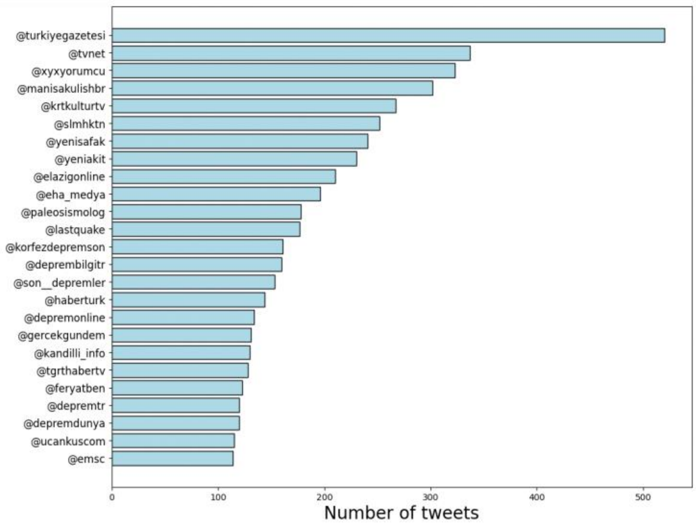

Figure 3.8 : Active users.

Apart from being mentioned, to look up the most active users, the categories and types are totally different. Active users are those who tweeted the most. They represent the more out-degree concept from the centralities. They are more social and talkative people in the network. The users' categories are more likely to "Media" or "Individual".

3.4.3 Time Series plotting

The time series shows the users' reactions toward the incident in the network. This type of visualization is very critical to decide whether social media is really useful for disaster management. In Figure 3.9, daily total tweets are given. Considering that the Elazig earthquake has happened on January, 24 at 17:55, the number of tweets shows that Twitter has an instant reaction in the hazard.

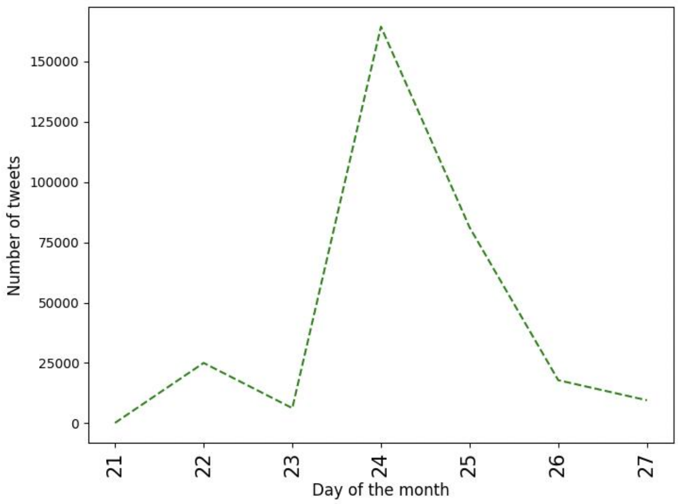

Figure 3.9 : Number of tweets per day.

Since we wanted to be sure that the earthquake affected society instantly, we also checked Google Trends data. We have extracted the Twitter data by filtering #deprem and #elazig. Therefore, each tweet is being related to the earthquake. However, with Google Trend, we can see the overall reaction among all-region and see if there is a similar behavior with the tweets number changing. In Figure 3.10, it is clearly seen that on the disaster day there is a peak search for both "deprem" and "elazığ" keywords. People tended to search about the incident and they care about it on the web.

To see if people talk about a more specific incident which is related to #elazig, or generally which is represented as #deprem, in Figure 3.11 tweets per 5 minutes for each hashtag were given. It is important to note that on January, 23, an earthquake happened in the city of Manisa. The plot shows that society reacts to specific incidents.

In Figures 3.4 and 3.5, the hashtag of Manisa can be shown easily since the earthquake in Manisa happened on the date that collected data extracted.

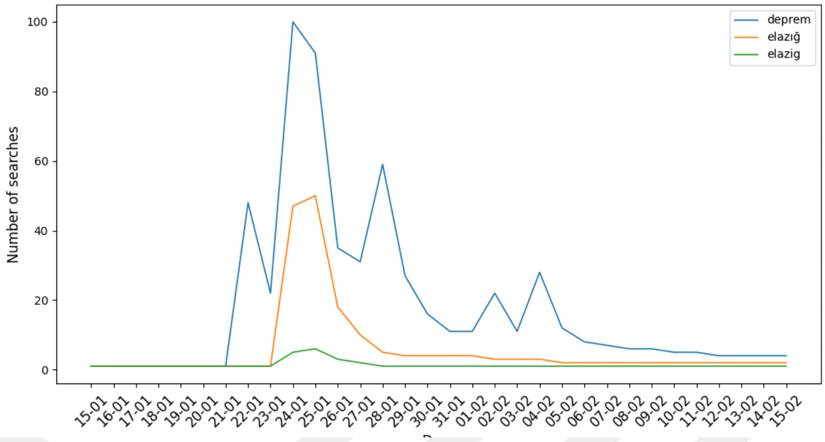

Figure 3.10 : Number of tweets per day.

To see if people talk about a more specific incident which is related to #elazig, or generally which is represented as #deprem, in Figure 3.11 tweets per 5 minutes for each hashtag were given. It is important to note that on January, 23, an earthquake happened in the city of Manisa. The plot shows that society reacts to specific incidents. In Figures 3.4 and 3.5, the hashtag of Manisa can be shown easily since the earthquake in Manisa happened on the date that collected data extracted.

Considering there has been another earthquake on January, 23, both Google Trend results and hashtag tweets results have the same ups and downs. From all time-related plots, the instant reaction is obvious. It leads us to answer one of the research questions: Does social media react dramatically in the disaster phases? The answer is yes. Social media has a reaction particularly in the response phase of disaster management.

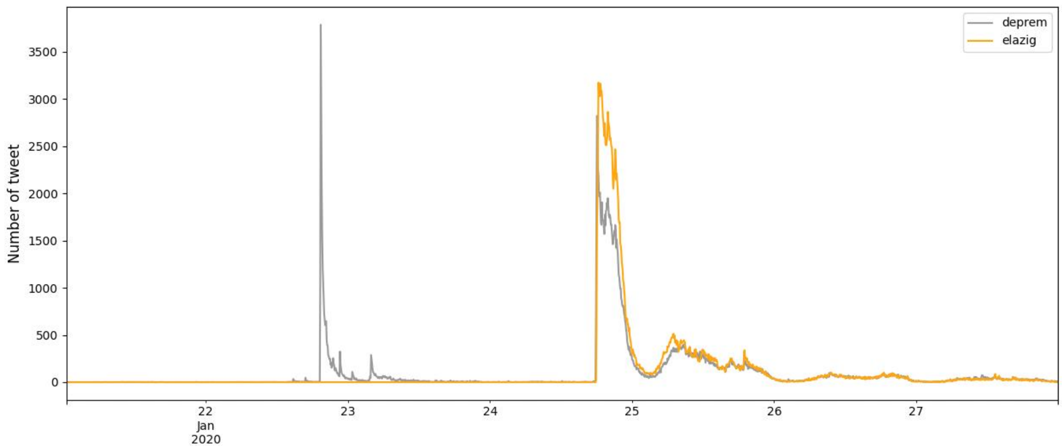

Figure 3.11 : Tweets per 5 min for #deprem and #elazig.

3.5 Social Network Analysis

In the SNA part, Source Node, Target Node, date-time, category, hashtag, number of retweets, likes, and replies attributes were used. The implementation was made in Gephi application and the centrality results were created in .csv format while the graph visualization was created in PNG form.

From the collected data, two files were created as Nodes.csv and Edges.csv. In Nodes file, both source and target nodes were gathered and category of each node, the total number of retweets, likes, and replies were constituted. In Edges file, the “reply to” relationships between two nodes and their date-time, and hashtag were created. We created a graph using these two files and centrality measures were calculated.

After getting centrality results, visual representations of the centralities were made in Python language.

3.5.1 Gephi and network visualization

Centrality statistics measurements, filters, and graph visualizations were made in Gephi. In the graph, the total number of nodes was 24,129, and the total number of edges was 38,328. Each user’s category helped us to comment on the behavior of each type of user in the network. For graph overview, layout type, node size, and node color were chosen. To see the differences, three graphs and centralities were compared with and without filters.

Unfiltered Graph: In Figure 3.12, a complete graph without any filter is shown. Force Atlas layout which is a Force-directed large graph layout algorithm was used to visualize the graph. Force-directed layout is very beneficial for visualizing clusters in the network between nodes. Node sizes and node colors were arranged according to betweenness centrality and modularity respectively. Modularity is a measure that looks for each node and gives a represented cluster by its density of the other nodes in the network.

From the visual, it is easy to see that there are various nodes that do not connect with the core structure. With the graph manipulation, the goal is to find meaningful and plain visualization that can be given.

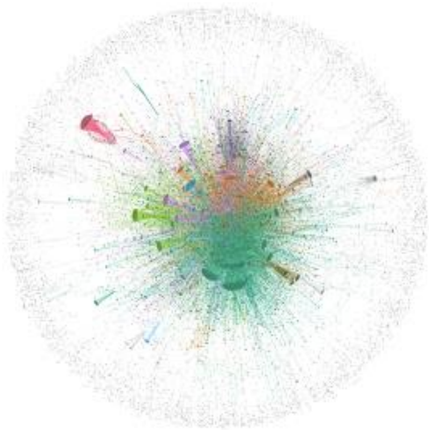

Figure 3.12 : Graph visualization without filtered.

Giant Component Filtered Graph: In Figure 3.13, In Figure 4, the filter of the giant component which filters out all nodes that are not connected into the main cluster was applied to the network. Layout, node size, and node color were set with the same choice in the unfiltered graph. After the giant component filter, the total number of nodes changed into 19,540 which is %80.98 of the entire nodes, and the total number of edges became 35,456 which is %92.51 of the entire edges. From the visual form, it can be seen that small unrelated nodes circled the main graph was removed. The network connections and the cluster can be seen easier now.

There are so many filter options that we can make. However, in this study, we will not emphasize in filtering, since we would like to see each user's behavior.

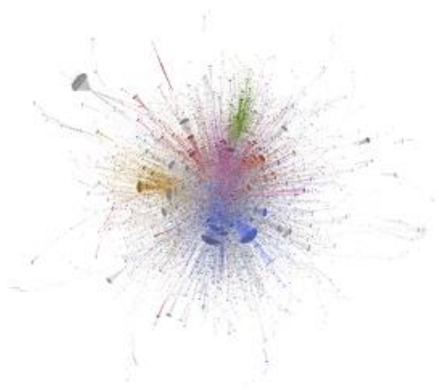

Figure 3.13 : Graph visualization with giant-component filter.

Local Graph: From the centrality results which will be given detailly in the Results section, there is a certain type of category users affect the flow of discussion. To see if there are minor effects that do not seem because of the majority, we created a local graph. For this, we have selected the first 175 users with the highest eigenvector centrality. The reason we have selected eigenvector centrality measure for the elimination is that eigenvector measures have more dominance while they have small quantities. These 175 people were composed of 17 percent of the non-zero eigenvector centralities. The distribution of eigenvector values is power-law; therefore these 175 users generate a big part among all users. In Figure 3.14, the graph visualization is shown. Layout, node size, and colors were chosen the same as the others. From the visual, it is shown that the irrelevant layer of the nodes was increased and the clusters became disorderly. To put in order, we filtered the graph with the giant component to get rid of circled nodes. In Figure 3.15, a giant component filtered local graph is given.

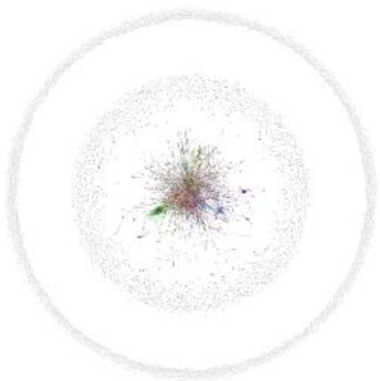

Figure 3.14 : Local graph without filtered.

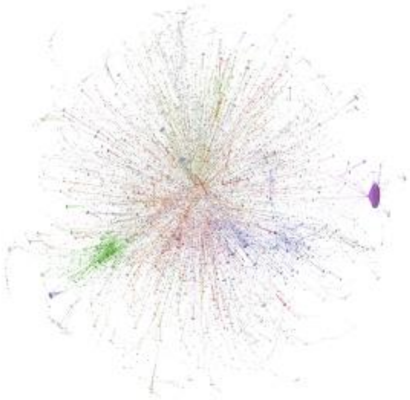

Figure 3.15 : Local graph with giant component filter.

From the results, we can say that once we deleted the important nodes according to eigenvector centrality, the network became scattered. There is no distinct user category or the user itself that we can say is important.

3.6 Calculations

Among all centrality calculations mentioned in chapter 2, we focused on two main centralities which are betweenness centrality and eigenvector centrality. The reason we chose both of them is that our goal is to find the best suitable users list which can lead society in an accurate and fastest way. Therefore, two dimensions, speed, and density are essential.

After we got centralities, two tables according to sorted from large to small for both betweenness and eigenvector centralities were created. In the results, other centralities such as in-degree and out-degree were also considered in interpretation.

3.7 Other Similar Incidents

After calculating the results, before deciding the final situation, we compared the results which came from other and similar incidents. Table 3.1 details the three largest earthquakes that occurred in Turkey in the last 5 years are listed [43]. We processed the same calculation steps for these earthquakes incidents and analyze whether the user types behaviors are similar in a disaster moment.

Table 3.1 : Other incidents information.

Region	Kirmansah	Ege	Ege
Date	12-11-2017	20-07-2017	12-06-2017
Place	Kirmansah/Iran	Mugla/Turkey	Izmir/Turkey
Magnitude	7.2	6.3	6.2
Depth	20.00 km	7.80 km	6.96 km
Deaths	629	0	1
Casualties	>8000	35	10

4. RESULTS

Implementations were made in two applications. Data pre-processing, text mining and all visualization stages were done in Python language Jupyter Lab notebook. Social Network Analysis and centrality calculations were done in the Gephi application. The resulting output and collected data stored in .csv file.

The data extraction part took 5 hours for 304,473 tweets and other attributes. The collected data file size is 134.3 MB.

4.1 Centrality Scores

In Table 4.1 and Table 4.2, the top 15 betweenness centrality and eigenvector centrality scores in descending sort are given respectively.

From the results, it is seen that users who aligned with eigenvector values are more in the corporate category rather than individuals. In Table 4.1, users who are “individual” are verified well-known users e.g. celebrities. Another detail about the eigenvector list is that users do not interact a lot in the network, however, if their out-degree is higher than zero which means if they tweeted, their effect is very high. So many people like, reply, and retweeted their tweets. This means that if these types of users’ tweet, the message they share is very likely to spread fast. And since all these types of users are very well-known and verified, misinformation spreading would be less likely.

In Table 4.2, users who have higher betweenness centrality, are comprised of the individual category or not fully identified which is the “other” category. When the top 15 users are checked, it is also seen that these "individual" users are neither verified nor the celebrity, they are only users with high followers. Unlike eigenvector, we can understand that this type of user interacts a lot because their in-degree and out-degree values are very high. However, their tweets are not as popular as eigenvector users. Their tweets’ reply, retweet, and like counts are not very high -except one user which is in the “Institutional” category. Given our purpose of work, we need to analyze whether such users will be effective.

Table 4.1 : Eigenvector centrality top 15 users.

Id	Label	Category	#replies	#retweets	#likes	indegree	outdegree	betweeness centrality	eigen centrality
1	kizilay	Institutional	0	0	0	1260	0	0	1
2	rterdogan	Govermental	0	0	0	1247	0	0	0.943628
3	afadbaskanlik	Institutional	36	411	1747	894	3	4586.75	0.695753
4	suleymansoylu	Govermental	65	380	2238	770	1	1505.5	0.593564
5	elazigbld	Govermental	0	0	0	454	0	0	0.341823
6	acunilicali	Individual	0	0	0	411	0	0	0.30883
7	ekrem_imamoglu	Govermental	0	0	0	407	0	0	0.300422
8	fenerbahce	Institutional	0	0	0	350	0	0	0.258011
9	haluklevent	Individual	313	4843	60402	332	1	379	0.241577
10	tc_icisleri	Govermental	0	0	0	318	0	0	0.235093
11	bernalacin35	Individual	0	0	0	304	0	0	0.22201
12	drkerem	Individual	21	325	1507	294	1	187.119048	0.218482
13	tcbestepe	Govermental	0	0	0	283	0	0	0.213396
14	cediosman	Individual	0	0	0	268	0	0	0.195076
15	cnnturk	Media	0	0	0	184	0	0	0.178285

Table 4.2 : Betweenness centrality top 15 users.

Id	Label	Category	#replies	#retweets	#likes	indegree	outdegree	betweeness centrality	eigen centrality
1	adayesat	Individual	1	0	3	9	88	14241.10873	0.009073
2	eyup_sagcan	Individual	0	13	36	3	5	13463	0.005405
3	sirvanvural	Other	2	0	6	7	23	13019.16667	0.006814
4	kenan12_erol	Individual	0	0	0	10	62	8401.109127	0.010066
5	hayrolayasemin	Other	1	1	5	8	8	8185.504762	0.00761
6	fevzisevgili	Other	6	39	249	10	5	8185	0.00892
7	mustafaymn	Individual	11	362	472	40	4	7430.5	0.029947
8	atasoner____	Individual	0	0	3	12	10	5774.35	0.010082
9	focalapkulu	Individual	0	23	122	12	5	5711.333333	0.010418
10	capulcu_ihtiyar	Other	0	3	4	7	72	5661.194444	0.007487
11	afadbaskanlik	Institutional	36	411	1747	894	3	4586.75	0.695753
12	erayyynihal	Individual	1	6	10	12	10	4158.833333	0.01005
13	avekremolgac	Individual	4	55	254	14	5	3368.166667	0.01214
14	hilalzcan20	Individual	0	0	0	9	11	3314.235714	0.008184
15	gundogan_reis	Other	0	0	2	6	9	2927.833333	0.005154

4.2 Ego Network Analysis

According to Table 4.1 and Table 4.2, to see which centrality users can help in disaster management, we created ego network analysis. The Ego network is creating a network of an individual node (ego node) and draw the path of its directly connected other users (alters). There are different explanations of ego networks in the literature, however, its idea is the same [37,38]. We have created ego network analysis for both user of "@kizilay" and user of "@adayesat" since they are the highest scored users for eigenvector and betweenness centrality respectively. In both ego networks, the depth which is the alter level was taken as 1. This means only direct relations with the ego user was taken. In Figures 4.1 and 4.2, the visualizations of ego networks are shown. In eigenvector's highest user -@kizilay- reaches 5.23% users of the network, while betweenness highest user -@adayesat- has 0.38% users of the network. This shows that reaching the whole community is easier with the users sorted by eigenvector centrality.

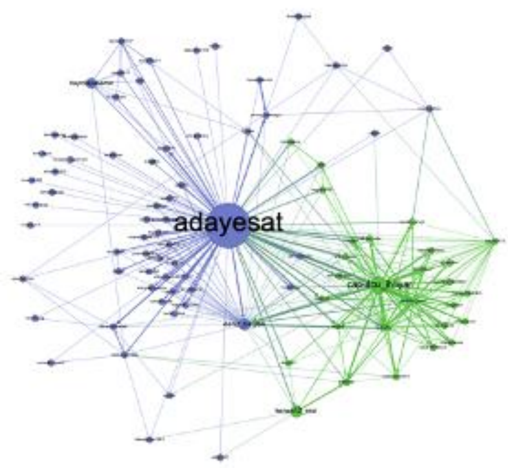

Figure 4.1 : Ego network of the user @adayesat.

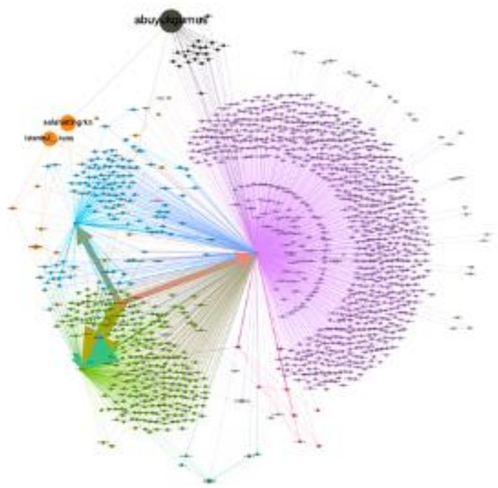

Figure 4.2 : Ego network of the user @kizilay.

The Ego network gives information spreading speed and volume for the node. In some works, the authors agree that eigenvector centrality has a higher influential impact on network [39]. From the ego network visuals and the proportion of spreading intensity, we can say that eigenvector is the best centrality to find an optimum solution in our study.

The ego network results give us an answer to another research question which is: Which type of users spread the information faster? Given the fact that dispersion density and the neighbor users' reliability, users who have higher eigenvector centrality may spread the information faster. And from the previous title, it is clear that users who have higher eigenvector centrality are in the governmental and institutional categories. Hence, corporate type users may help to spread the information fast.

Moreover, the centrality results show the dominance of the user type in the network as well. From the measurement calculations, again corporate type users or well-known individuals like celebrities are more trustable and ruler of the network. This situation gives an answer of another research question which is: Which type of users has more impact on disaster moments? Apparently, community is interested more in well-known and verified accounts. This is a logical and prospective condition.

4.3 Centrality Distributions

Centrality visualizations can help interpret the situation of the network. First, we have created a bar chart that shows total centrality scores of betweenness and eigenvector centrality for each category respectively in Figure 4.3 and Figure 4.4. From the results, the eigenvector centrality score's user category is stable and distinct while the betweenness centrality score's users are highly ambiguous. Other categories were created when the function cannot find any related other four categories for the user's name. These accounts can be more likely to a bot and fake accounts than other category users.

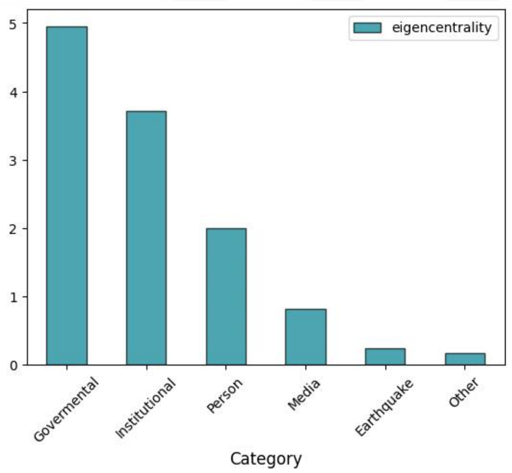

Figure 4.3 : Category dist. by eigenvector centrality.

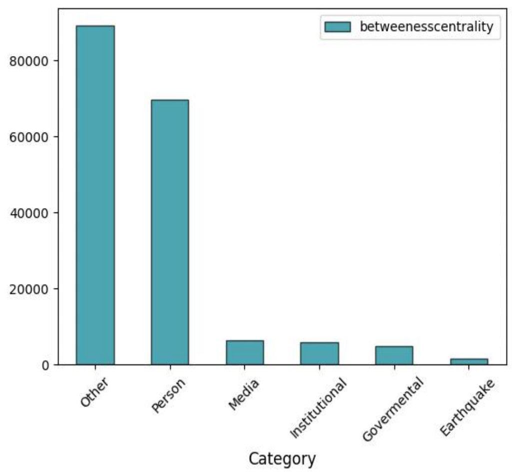

Figure 4.4 : Category dist. by betweenness centrality.

From the category plots, it is seen that eigenvector centrality ranking is way more important than betweenness centrality. Among the other visual and calculation results, categories show very important detail about the behavior of the user types in disaster moment, therefore, this plot says a lot about our research questions.

One of the purposes of our study is to see whether the result we will have can be said as a general concept for disaster management. We would like to see whether this network analysis reacts as a local network or real-world example. To see that, degree distributions can help what kind of network we are working with. In Figure 4.5 and Figure 4.6, eigenvector and betweenness centrality score log-log scale distributions are given. Both of the scales act like Power Law distributions. They are high skew and asymmetric. The power law is a distribution style where two quantities are related to each other by increasing with power one of them [40]. As an example, the red point in Figure 4.6 means that the network has tens of thousands of observations when the centrality score is less than 10.

The power law is a form of a Pareto analysis where literature says that only 80 percent of something caused by 20 percent of the system. In Figure 32, the linear relations in indegree and outdegree distribution also show the Pareto chart visual, which is also Power law.

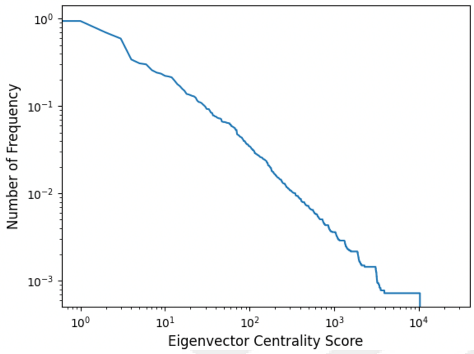

Figure 4.5 : Eigenvector centrality log-log distribution.

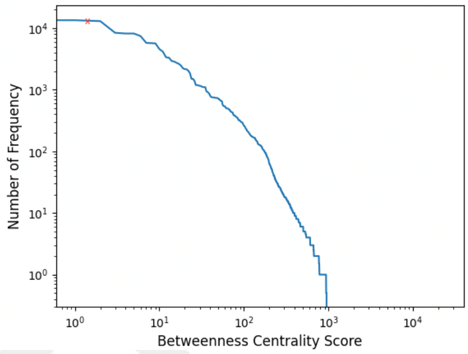

Figure 4.6 : Betweenness centrality log-log distribution.

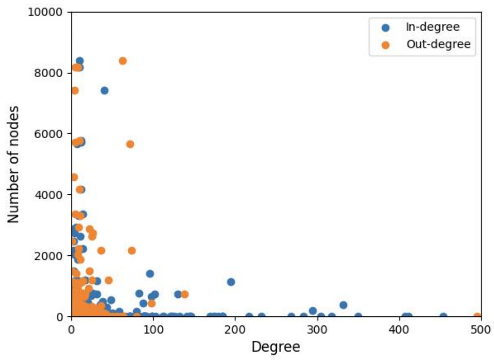

Figure 4.7 : Indegree & outdegree distribution.

The reason we are focusing on Power-law distribution is that this kind of distribution behavior on the network proves that the network is scale-free. Real-world networks have hubs where some nodes have higher degree scores than others, and these nodes bring the long tail distribution. Scale-free networks are power-law because power-law distribution stays the same no matter what scale we are working on. It remains unchanged. In this way, we know that our sample from Twitter shows that the network is scale-free and it can be counted as a real-world network [41].

4.4 Category Analysis

After getting the calculation results, we focused on category-based estimations. The centrality distributions of the categories were given in Figures 4.3 and 4.4 in the previous title. Based on this, it seems once again that corporate accounts are more effective. The commentary about this result was explained in the previous chapters. Figure 4.8 gives category-based tweet counts. The reason why the "Other" category is much more than others is that there are too many user names that do not comply with any rule. This indicates the formation of long-tail users.

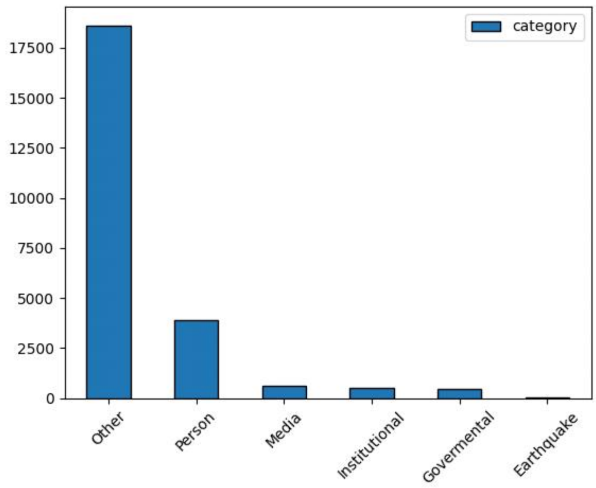

Figure 4.8 : Category based tweet numbers.

Another topic to look detailly in the category is the text users send. Since the disaster management in social media occurs mostly with text, the users’ tweets are also essential to discuss. In this study, we did not focus on analyzing tweets with NLP. Therefore, we will deduct from word clouds and frequently used words. Our goal is to find out if the types of users in any category are sending informative messages. In Figure 4.9 to 4.14, the word cloud of tweets sent by users of each category type was created. Before creating word clouds, text preprocessing, and stopwords removing were done. Dominant words that are apparent in every tweet such as deprem, elazığ, etc. were also added to the list of stopwords. The reason for this is to make other unique words to stand out. Frequently used words in the tweets are written in a larger font in the word cloud. In this way, we can understand whether users of each category act in common.

Figure 4.11 : Media category.

Figure 4.12 : Governmental category.

Figure 4.9 : Institutional category.

Figure 4.13 : Individual category.

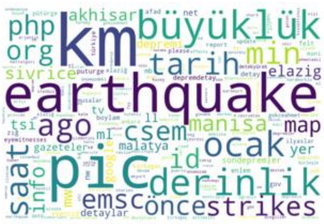

Figure 4.14 : Earthquake category.

Figure 4.10 : Other category.

From the results, it is seen that each user types write on similar topics. Usually, sadness and condolence notifications were repeated. Any user type did not send informative messages during and after the earthquake. Only users in the "earthquake" category (Figure 4.13) shared concrete measurement information (depth, magnitude, etc.) related to the earthquake. It is logical since the account opened only for the earthquake focused on these issues.

4.5 Other Incident Results

In chapter 3, three other highest earthquake incidents were mentioned. The category proportion for both eigenvector and betweenness calculations and results are shown in Table 4.3 and Table 4.4 for eigenvector and betweenness centrality respectively. From the results, it is shown that categorical user behavior stays still in disaster moments. This means that the action plan may be taken having regard to these results.

This generalization and consistent results show that further action may be taken to improve disaster management. Also, the last unanswered question of research questions can be answered after these figures. Social media can clearly be used for emergency management, and this usage would be of service to the majority of the social media community. As a matter of fact, if the information sharing would be focused, we believe that casualty saving would increase.

Table 4.3 : Eigenvector centrality category proportions.

	Izmir	Mugla	Kirmansah	Elazig
Governmental	36.67%	23.33%	46.67%	40.00%
Institutional	13.33%	10.00%	23.33%	26.67%
Media	10.00%	6.67%	20.00%	6.67%
Person	10.00%	23.33%	10.00%	23.33%
Earthquake	0.00%	6.67%	0.00%	3.33%
Other	36.67%	30.00%	0.00%	0.00%

Table 4.4 : Betweenness centrality category proportions.

	Izmir	Mugla	Kirmansah	Elazig
Governmental	23.33%	20.00%	6.67%	6.67%
Institutional	20.00%	6.67%	3.33%	3.33%
Media	13.33%	6.67%	23.33%	0.00%
Person	10.00%	20.00%	30.00%	56.67%
Earthquake	0.00%	10.00%	0.00%	0.00%
Other	26.67%	36.67%	36.67%	33.33%

5. CONCLUSION AND FUTURE WORKS

In this study, social media behavior on disaster were analyzed. The users were categorized by their account profile and screen name. Social network analysis and graph mining were made to get centrality scores. By centrality scores, the final decision and interpretation were made. According to the results, it was decided that it is more appropriate to use the order of the eigenvector value in decision making. Betweenness centrality is higher by more unfamiliar users, and in the disaster moment, the category type of user is important to spread reliable messages. The unrecognized users with a category of "Other" may give unpredictable results. In addition, the ego network result showed that users with high eigenvector centrality can spread information to more people. The high eigenvector of users is mostly governmental and institutional users. This is also a good result that can help prevent the spread of false information.

There are many results that can be interpreted and shown in this study. In the future, studies can be done by entering the details for disaster management with these features obtained from Twitter. Especially by improving the Natural Language Processing (NLP) study in Turkish, focusing on the messages sent during the disaster can be directed towards the community in the network. From this study, mostly seen words indicating sadness, condolences. However, with the advanced NLP studies, unseen detailed messages such as disaster guidance can be extracted.

It is proved that the sample gathered from Twitter shows a scale-free network. We also know that Twitter users act in the same way during earthquakes. General action planning may be created. Also, other social media platforms such as Facebook, Instagram, etc. can be gathered to create more accurate results. Facebook is another mostly-used social media platform. Especially middle-aged people tend to use this platform. Again information sharing and situation analysis may be done with this additional social media and combine them. This way, important health, rescue, or any other likely deliveries in disaster moments would be easier. In addition to this, this work can be widened with image processing from social media posts during a disaster. By this means, the destructive result of the disaster can be seen faster.

A very big problem in social media is fake news. Even in the Elazig earthquake, there were countless fake tweets about casualties, wrecking, even about aid efforts. Stopping this problem would help social media more reliable and useful. Since this problem is a worldwide issue, there are various fake news studies in multiple languages. In the future, a fake news detection study can be added to this study to strengthen the results. In fake news there is a Natural Language Processing (NLP) works for each language, so this part will be related to focusing on the part of the messages.

Lastly, after finishing these two parts of the study, an organizational disaster management application would be helpful to make provision for disasters in any stage of the disaster. Detecting real influencer users, spreading the accurate message, eliminating the fake news will help in saving lives. Giving that certain category of users, we can say that governmental and institutional categories in the emergency management phases will be effective in this organizational disaster management application.

6. REFERENCES

[1] **Caglayan, N., Satoglu, S. and Kapukaya, E.** (2018). Afet Yönetiminde Büyük Veri ve Veri Analitiği Uygulamaları: Literatür Araştırması, *7. Ulusal Lojistik ve Tedarik Zinciri Kongresi*, Bursa, Turkey.

[2] **Solomon, B., Duce, D. and Harrison, R.** (2011). Methodologies for Using Social Media Collaborative Work Systems, *First International Workshop on Requirements Engineering for Social Computing*, Trento, Italy, pp.6–9.

[3] **Kim, J. and Hastak, M.** (2018). Social Network Analysis: Characteristics of Online Social Networks After a Disaster, *International Journal of Information Management*, volume 38, pp.86–96.

[4] **Url-1** <https://www.internetlivestats.com/one-second/#tweets-band>, date retrieved 23.03.2020.

[5] **Url-2** <https://www.oberlo.com/blog/twitter-statistics>, date retrieved 23.03.2020.

[6] **Sakaki, T., Okazaki, M. and Matsuo, Y.** (2010). Earthquake Shakes Twitter Users: Real-time Event Detection by Social Sensors, *WWW2010: The Nineteenth International WWW Conference*, Raleigh, USA.

[7] **Url-3** <https://tr.wikipedia.org/wiki/2020_Elaz%C4%B1%C4%9F_depremi>, date retrieved 23.03.2020.

[8] **Carley, K., Malik, M., Landwehr, P., Pfeffer, J. and Kowalchuck, M**. (2016). Crowd sourcing disaster management: The complex nature of Twitter usage in Padang Indonesia, *Safety Science*, *90*, https://doi.org/10.1016/j.ssci.2016.04.002.

[9] **Landwehr, P., Wei, W., Kowalchuck, M. and Carley, K.** (2016). Using Tweets to Support Disaster Planning, Warning and Response, *Safety Science*, *90*, https://doi.org/10.1016/j.ssci.2016.04.012.

[10] **Chatfield, A. and Brajawidagda, U.** (2012). Twitter tsunami early warning network: A social network analysis oftwitter information flows, *Proceedings of the 23rd ACIS Conference*, pp.1–10.

[11] **Kim, J., Bae, J. and Hastak, M.** (2018). Emergency Information Diffusion On Online Social Media During Storm Cindy In U.S., *International Journal of Information Management*, *40*, 153–165, https://doi.org/10.1016/j.ijinfomgt.2018.02.003.

[12] **Chen, W., Zhang, H., Comfort, L. and Tao, Z.** (2020). Exploring complex adaptive networks in the aftermath of the 2008 Wenchuan earthquake in China, *Safety Science*, *125*, 153–165, https://doi.org/10.1016/j.ssci.2020.104607.

[13] **Xiao, Y., Huang, Q. and Wu, K.** (2015). Understanding Social Media Data for Disaster Management, *Natural Hazards*, *79*, https://doi.org/0. 1007/s11069-015-1918-0.

[14] **Khaleq, A. and Ra, I.** (2019). Twitter Analytics for Disaster Relevance and Disaster Phase Discovery: Volume 1, *Proceedings of the Future Technologies Conference*, https://doi.org/10.1007/ 978-3-030-02686-8_31.

[15] **Zou, L., Lam, N., Cai, H. and Y., Q.** (2018). Mining Twitter Data for Improved Understanding of Disaster Resilience, *Annals of the American Association of Geographers*, *108:5*, 1422–1441, https://doi.org/10.1080/ 24694452.2017.1421897.

[16] **Aziz, K., Zaidouni, D. and Bellafkih, M.** (2019). Social Network Analytics: Natural Disaster Analysis Through Twitter, *2019 Third International Con- ference on Intelligent Computing in Data Sciences (ICDS)*, Marrakech, Morocco.

[17] **Jamali, M., Nejat, A., Ghosh, S. and Cao, G.** (2018). Social media data and post-disaster recovery, *International Journal of Information Management*, *44*, 25–37, https://doi.org/10.1016/j.ijinfomgt.2018. 09.005.

[18] **Kim, J. and Park, H.** (2020). A Framework for Understanding Online Group Behaviors During A Catastrophic Event, *International Journal of Information Management*, *51*.

[19] **Lu, X. and Brelsford, C.** (2014). Network Structure and Community Evolution on Twitter: Human Behavior Change in Response to the 2011 Japanese Earthquake and Tsunami, *Scientific reports*, *4*, https://doi.org/ 10.1038/srep06773.

[20] **Şimşek, M.** (2012). Sosyal Ağlarda Veri Madenciliği Üzerine Bir Uygulama (Master's Degree Thesis, *Gazi University, Graduate School of Natural and Applied Sciences*, Ankara, Turkey.

[21] **Savaş S. and Topaloğlu, N.** (2019). Data analysis through social mediaaccording to the classified crime, *Turkish Journal Of Electrical Engineering Computer Sciences*, *27*, 407–420, https://doi.org/10.3906/ elk-1712-17.

[22] Wang, Z. and Ye, X. (2017). Social media analytics for natural disaster management, *International Journal of Geographical Information Science*, 1–24, https://doi.org/10.1080/13658816.2017.1367003.

[23] **Rodrigueza, R. and Estuar, M.** (2018). Social Network Analysis of a Disaster Behavior Network: An Agent-Based Modeling Approach, *IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*.

[24] **Raj, K., Mohan, A. and Srinivasa, K.** (2018). Practical Social Network Analysis with Python, *Computer Communications and Networks*, 1–24, https: //doi.org/10.1007/978-3-319-96746-2.

[25] **Tyshchuk, Y., Hui, C., Grabowski, M. and A., W.W.** (2011). Social Media and Warning Response Impacts in Extreme Events: Results from a Naturally Occurring Experiment, *45th Hawaii International*

Conference on System Sciences, Maui, HI, pp.818–827, https://doi.org/10. 1109/HICSS.2012.536.

[26] **Huisman, M. and Duijn, M.** (2011). A Reader's Guide to SNA Software, *The SAGE Handbook of Social Network Analysis*, London, U.K.

[27] **Tunali, V.** (2016). Sosyal Ag˘ Analizine Giris¸.

[28] **Li, C., Li, Q., Mieghem, P., Stanley, H. and Wang, H.** (2015). Correlation Between Centrality Metrics and Their Application To The Opinion Model, *The European Physical Journal*, *88*, https://doi.org/10.1140/ epjb/e2015-50671-y.

[29] **Zaman, T., Herbrich, R., Gael, J. and Stern, D.** (2010). Predicting Information Spreading in Twitter.

[30] **Sadri, A., Hasan, S., Ukkusuri, S. and Cebrián, M.** (2017). Understanding Information Spreading in Social Media during Hurricane Sandy: User Activity and Network Properties, https://arxiv.org/abs/1706. 03019.

[31] **Wang, D., Wen, Z., Tong, H., Lin, C., Song, C. and Barabasi, A.** (2011). Information Spreading in Context, *Proc. of WWW*, 735–744, https: //doi.org/10.1145/1963405.1963508.

[32] **Huang, Q. and Cervone, G.** (2016). Usage of Social Media and Cloud Computing During Natural Hazards, *Cloud Computing in Ocean and Atmospheric Sciences*, 297–324, https://doi.org/10.1016/ B978-0-12-803192-6.00015-3.

[33] **Kavota, K.J., Kamdjoug, J.R. and Wamba, S.F.** (2020). Social media and disaster management: Case of the north and south Kivu T regions in the Democratic Republic of the Congo, *International Journal of Information Management*, *52*, 297–324, https://doi.org/10. 1016/j.ijinfomgt.2020.102068.

[34] **R., F., Gei, F., Marabissi, D. and Micciullo, L.** (2016). 2 - The Use of Social Networks in Emergency Managemen, *Wireless Public Safety Networks*, *2*, 25–61.

[35] **Sloan, L. and Morgan, J.** (2015). Who Tweets with Their Location? Understanding the Relationship between Demographic Characteristics and the Use of Geoservices and Geotagging on Twitter, *PloS one*, *10(11)*, https://doi.org/10.1371/journal.pone.0142209.

[36] **Url-4**, https://deprem.afad.gov.tr/depremkatalogu, retrieved from: 13.05.2020.

[37] **Arnaboldi, V., Conti, M., Passarella, A. and Dunbar, R.** (2017). Online Social Networks and information diffusion: The role of ego networks, *Online Social Networks and Media*, *1*, 44–55, https://doi.org/ 10.1016/j.osnem.2017.04.001.

[38] **De Salve, A., M., D., Guidi, B. and Ricci, L.** (2015). The impact of user's availability on On-line Ego Networks: A Facebook analysis, *Computer Communications*, *73*, https://doi.org/10.1016/j.comcom. 2015.09.001.

[39] **Martin, G.** (2016). A social network analysis of Twitter: Mapping the digital humanities community, *Cogent Arts Humanities*, *3*, https://doi. org/10.1080/23311983.2016.117145.

[40] **Bar-Yam, Y.,** Concepts: Power Law. New England Complex Systems Institute, https://necsi.edu/power-law, retrieved from: 18.05.2020.

[41] **DQ, N**., Scale-free networks. From Math Insight, http://mathinsight. org/scale_free_network, retrieved from: 18.05.2020.

[42] **Url-5** < https://github.com/twintproject/twint>, date retrieved 20.05.2020.

[43] **Url-6** <https://deprem.afad.gov.tr/>, date retrieved 20.05.2020.

APPENDICES

APPENDIX A.1: Betweenness Centrality Results Top 150
APPENDIX A.2: Eigenvector Centrality Results Top 150

APPENDIX A.1

Table A.1 : First 150 users with the highest betweenness centrality.

#	Username	Category	Indegree	Outdegree	Betweenness Centrality
1	adayesat	Other	9	88	14241.10873
2	eyup_sagcan	Person	3	5	13463
3	sirvanvural	Other	7	23	13019.16667
4	kenan12_erol	Other	10	62	8401.109127
5	hayrolayasemin	Other	8	8	8185.504762
6	fevzisevgili	Other	10	5	8185
7	mustafaymn	Person	40	4	7430.5
8	atasoner____	Media	12	10	5774.35
9	focalapkulu	Other	12	5	5711.333333
10	capulcu_ihtiyar	Other	7	72	5661.194444
11	afadbaskanlik	Institutional	894	3	4586.75
12	erayyynihal	Other	12	10	4158.833333
13	avekremolgac	Other	14	5	3368.166667
14	hilalzcan20	Person	9	11	3314.235714
15	gundogan_reis	Other	6	9	2927.833333
16	gullerevurgun4	Other	3	22	2870
17	ihsanemirgazili	Person	4	26	2739.666667
18	mywayturkey	Other	11	25	2635.16746
19	benastrea	Other	3	1	2476.783333
20	kadturay	Other	14	9	2232.779762
21	tc_berfin	Other	6	74	2161.685317
22	coskuner_rabia	Other	3	36	2156.441667
23	cankayhaan	Person	4	8	2040
24	mfatihkayhan	Person	8	11	1865.833333
25	suleymansoylu	Govermental	770	1	1505.5
26	tavuskuu4	Other	3	22	1502.133333
27	malatyabeltr	Govermental	96	6	1402.583333
28	slmhktn	Other	17	25	1194.666667
29	deniz_ba_ran	Person	7	45	1193.95119
30	mmt__21	Other	4	14	1175
31	gkhnkhrman	Other	31	14	1158.833333
32	elazig_valiligi	Govermental	195	2	1149.166667
33	terebenti	Other	6	1	1120.333333
34	kemalbekirhan2	Other	3	8	1105.5
35	ibrahimicgil	Other	3	12	1105
36	a_sgzc	Other	10	1	1101.416667
37	tayyarozcan	Person	2	7	944
38	soltarafim123	Other	4	21	926.416667

Table A.1 (cont.) : First 150 users with the highest betweenness centrality.

#	Username	Category	Indegree	Outdegree	Betweenness Centrality
39	bernado0707	Other	3	8	894.733333
40	burcugunden	Other	2	5	860.35
41	ali_ilbas	Person	2	5	796
42	35_lazuri_53	Other	2	15	762
43	m_thk_2	Other	3	15	757.116667
44	akut_dernegi	Other	83	5	753
45	selahattingrkn	Other	26	5	750.666667
46	ziyaselcuk	Other	130	1	743.533333
47	deparsis	Earthquake	4	138	739
48	depremdairesi	Earthquake	102	3	737.166667
49	sevdaloji_35	Other	4	3	736.85
50	yusufyl47083485	Other	5	17	726.84127
51	elaziggm	Other	31	12	723
52	ak_murad	Other	5	5	694.333333
53	veliagbaba	Other	24	9	688.783333
54	oznurcalik	Person	17	16	666.166667
55	durgunnecmiye	Other	3	6	660
56	trthaber	Media	98	2	652.666667
57	mardinilmem	Other	10	9	572.5
58	ferhat_korcak	Other	2	8	551.733333
59	ihtiyacharitasi	Other	48	1	543
60	arzudevrim22	Other	2	5	540.47619
61	hasanncakar	Other	1	8	535.833333
62	tuba5409	Other	1	7	512
63	cevher_ciftci	Other	5	4	500
64	12numaraorg	Other	38	7	497.733333
65	lemislam1	Other	2	6	496.666667
66	sukruuras	Other	4	4	495
67	mehmetcinar44	Person	9	3	486.75
68	sevdaturkusev	Other	20	6	473.716667
69	abkarabulut	Other	3	7	472.666667
70	elazigonline	Other	4	98	451.333333
71	ahmetcakir44	Person	9	2	441.5
72	kotexturkiye	Other	88	1	440
73	fatmacumhurefe	Person	12	11	431.669048
74	seherdogand	Other	6	6	430.938095
75	yasmin1903bjk1	Other	2	10	420.433333
76	elazigmem	Other	20	12	392.833333
77	vatan_sancak	Other	8	5	389.5
78	abuyukgumus	Other	35	3	379.5

Table A.1 (cont.) : First 150 users with the highest betweenness centrality.

#	Username	Category	Indegree	Outdegree	Betweenness Centrality
79	haluklevent	Other	332	1	379
80	cursebru	Other	4	14	376.424603
81	millibeka	Other	8	13	370
82	hudeydankayaalp	Other	3	5	361.5
83	harmoni1938	Other	2	12	359.666667
84	genclikbirligi	Other	8	36	353.65
85	umit_turhan	Other	2	10	351
86	mehmetfendogluu	Person	8	18	342.25
87	sukutualemm	Other	5	25	323.333333
88	msefikerden	Other	1	5	321
89	hasantahsinpasa	Other	1	5	321
90	fatihkadiri	Person	1	5	321
91	rembey2	Other	1	4	314
92	gonulnimetullah	Other	4	5	309.166667
93	kaptan1_gul_tc	Other	1	17	301.833333
94	rhisarciklioglu	Other	43	3	300
95	farisakgul	Other	1	5	296.333333
96	fatmasahin	Other	36	4	294.666667
97	m_akgun56	Other	4	5	294.5
98	zllktlr	Other	5	9	287.145238
99	paleosismolog	Other	10	28	276
100	zabeyazgul	Other	39	6	267.666667
101	mhptbmmgrubu	Other	10	7	261
102	cengiz_gokcel	Other	7	8	259.283333
103	hzpollyana15	Other	2	3	256
104	mucize_rte	Other	1	1	250
105	istanbul__kusu	Other	12	15	245.833333
106	trthabercanli	Media	6	2	237
107	who98408150	Other	24	2	231.5
108	gaziantepbeld	Other	18	6	226.666667
109	kutupps	Other	1	3	225.5
110	erenerdemnet	Other	14	3	216.4
111	kosal_21_12	Other	5	4	213
112	fuatoktay	Other	34	2	209.583333
113	tvnet	Media	7	3	208.166667
114	ekremkocaesk	Other	3	2	206.75
115	adilaysur	Other	2	2	202.5
116	miragest	Other	1	30	198
117	23yanilmaz	Other	4	8	194.333333
118	mirzakayhan	Other	3	5	190.5
119	denizfeneriorg	Person	7	6	187.583333
120	drkerem	Other	294	1	187.119048

Table A.1 (cont.) : First 150 users with the highest betweenness centrality.

#	Username	Category	Indegree	Outdegree	Betweenness Centrality
121	drsinanogan	Other	27	4	181.4
122	oasadikoglu	Other	3	16	179
123	alimahir	Person	12	4	177.5
124	goncalar07	Other	5	12	177
125	kadikoybelediye	Govermental	58	2	174
126	gsbgenclik	Other	18	2	174
127	kocaelibld	Govermental	8	5	173
128	serhatkayatc	Other	7	4	172.583333
129	evrenselgzt	Other	10	2	170
130	sondur_ozcan	Person	1	4	169
131	mardinilsaglik	Other	1	4	168
132	abdulah_topkaya	Other	2	4	167
133	aydiner_sah	Other	1	4	167
134	sevgili_kamuran	Other	1	6	167
135	haberturk	Media	80	2	166
136	tcmmvar	Other	3	7	165.665079
137	muratsahin2023	Person	9	8	158.5
138	mehmetsekmen	Person	14	3	158.333333
139	mhilmiguler	Other	31	2	154.5
140	yenisafak	Other	13	1	153
141	gokhanveysi	Other	4	11	149
142	malatyatso1923	Other	5	3	148.5
143	eha_medya	Media	3	27	146.566667
144	akosmanli_nihat	Other	1	36	146
145	denizzbarlas	Person	5	22	143.833333
146	ceyhunirgil	Institutional	9	8	140
147	zaferyagantekin	Other	2	1	134
148	ayseucar5	Other	8	2	132
149	devletimalatya	Other	7	4	131.5
150	gaziantepgm_	Other	14	13	131

APPENDIX A.2

Table A.2 : First 150 users with the highest eigenvector centrality.

#	Username	Category	Indegree	Outdegree	Eigenvector Centrality
1	kizilay	Institutional	1260	0	1
2	rterdogan	Govermental	1247	0	0.943628
3	afadbaskanlik	Institutional	894	3	0.695753
4	suleymansoylu	Govermental	770	1	0.593564
5	elazigbld	Govermental	454	0	0.341823
6	acunilicali	Person	411	0	0.30883
7	ekrem_imamoglu	Govermental	407	0	0.300422
8	fenerbahce	Institutional	350	0	0.258011
9	haluklevent	Person	332	1	0.241577
10	tc_icisleri	Institutional	318	0	0.235093
11	bernalacin35	Person	304	0	0.22201
12	drkerem	Govermental	294	1	0.218482
13	tcbestepe	Govermental	283	0	0.213396
14	cediosman	Person	268	0	0.195076
15	cnnturk	Media	184	0	0.178285
16	kandilli_info	Earthquake	232	0	0.169179
17	dyson_lin	Media	217	0	0.157615
18	elazig_valiligi	Govermental	195	2	0.149678
19	mansuryavas06	Govermental	184	0	0.137788
20	turktelekom	Institutional	184	0	0.136031
21	ahbapplatformu	Institutional	179	0	0.133489
22	akparti	Govermental	175	0	0.13078
23	emniyetgm	Govermental	173	0	0.12927
24	turkcell	Institutional	169	0	0.125113
25	murat_kurum	Person	146	0	0.115854
26	jandarma	Govermental	81	0	0.111589
27	drfahrettinkoca	Govermental	145	0	0.110541
28	nacigorur	Person	144	0	0.108372
29	vodafonetr	Institutional	141	0	0.104688
30	ziyaselcuk	Person	130	1	0.101654
31	meral_aksener	Govermental	132	0	0.098241
32	istanbulbld	Govermental	126	0	0.09291
33	tv8	Media	123	0	0.09238
34	atuncayozkan	Person	123	0	0.092229
35	besiktas	Institutional	121	0	0.08778
36	kilicdarogluk	Govermental	113	0	0.083939
37	saglikbakanligi	Govermental	112	0	0.0835
38	ahbap_elazig	Institutional	83	0	0.078059

Table A.2 (cont.) : First 150 users with the highest eigenvector centrality.

#	Username	Category	Indegree	Outdegree	Eigenvector Centrality
39	trthaber	Media	98	2	0.077815
40	tskgnkur	Other	29	0	0.076125
41	orhanaydin6	Person	103	0	0.075877
42	depremdairesi	Earthquake	102	3	0.074179
43	afadturkey	Institutional	92	0	0.073303
44	herkesicinchp	Other	99	0	0.072513
45	youtube	Institutional	98	0	0.072321
46	csbgovtr	Govermental	99	0	0.07196
47	malatyabeltr	Govermental	96	6	0.071121
48	dbdevletbahceli	Govermental	90	0	0.066652
49	tk_tr	Other	83	0	0.066221
50	fahrettinaltun	Person	90	0	0.065692
51	orkid	Institutional	91	0	0.065649
52	molfix	Institutional	91	0	0.065649
53	ismailsaymaz	Person	89	0	0.065054
54	molpedturkiye	Institutional	90	0	0.064927
55	sleepykizlari	Other	89	0	0.064206
56	sleepybebekleri	Other	89	0	0.064206
57	galatasaraysk	Institutional	88	0	0.064031
58	kotexturkiye	Institutional	88	1	0.063484
59	unileverturkiye	Institutional	88	0	0.063484
60	abdulhamitgul	Person	83	0	0.061163
61	akut_dernegi	Institutional	83	5	0.061101
62	iyiparti	Govermental	80	0	0.058864
63	tff_org	Other	81	0	0.058763
64	haberturk	Media	80	2	0.058041
65	ntv	Media	79	0	0.05766
66	mgulluoglu	Other	72	0	0.05689
67	beratalbayrak	Govermental	73	0	0.056534
68	trabzonspor	Institutional	77	0	0.056037
69	06melihgokcek	Govermental	72	0	0.053634
70	tcmeb	Govermental	70	0	0.053568
71	ceydak	Other	72	0	0.053449
72	mutlumutfaklar	Person	66	0	0.047775
73	anadoluajansi	Media	57	0	0.047099
74	valicetinoktay	Govermental	57	0	0.046893
75	ankarabbld	Govermental	62	0	0.046381
76	hmbakanligi	Govermental	59	0	0.046045
77	adalet_bakanlik	Govermental	60	0	0.04499
78	kadikoybelediye	Govermental	58	2	0.044824

Table A.2 (cont.) : First 150 users with the highest eigenvector centrality.

#	Username	Category	Indegree	Outdegree	Eigenvector Centrality
79	gsbasketbol	Institutional	60	0	0.044368
80	selcukozdag	Other	60	0	0.044151
81	dibalierbas	Person	55	0	0.043701
82	demetakalin	Person	59	0	0.043536
83	tobbiletisim	Other	56	0	0.042476
84	tuncsoyer	Other	56	0	0.041746
85	gokhanozoguz	Person	57	0	0.041175
86	cumhuriyetgzt	Media	55	0	0.039898
87	ihtiyacharitasi	Other	48	1	0.03989
88	fatihportakal	Person	55	0	0.039844
89	mhp_bilgi	Other	53	0	0.039519
90	_tjk_	Other	43	0	0.038395
91	showtv	Media	50	3	0.038193
92	kasapoglu	Other	47	0	0.037186
93	ordubbld	Govermental	46	0	0.03706
94	eytsyddernegi	Other	49	0	0.036488
95	iletisim	Other	49	0	0.036438
96	sahin_serifoglu	Other	45	0	0.036376
97	metroturizm	Other	50	0	0.036233
98	gazetesozcu	Media	48	0	0.035232
99	ihhinsaniyardim	Other	45	0	0.03483
100	gibsosyalmedya	Media	44	0	0.034782
101	sputnik_tr	Other	48	0	0.034738
102	themarginale	Other	47	0	0.034377
103	diyanetbasin	Other	45	1	0.034003
104	tele1comtr	Other	46	0	0.033647
105	hdpgenelmerkezi	Other	46	0	0.033401
106	foxhaber	Media	45	2	0.032626
107	rhisarciklioglu	Other	43	3	0.032317
108	mevlutcavusoglu	Other	44	0	0.032233
109	ankara_kusu	Other	44	0	0.031742
110	zabeyazgul	Other	39	6	0.031734
111	t24comtr	Other	41	0	0.031147
112	genckizilay	Other	43	0	0.031129
113	ahaber	Media	42	1	0.030975
114	profdrdemirdag	Other	42	2	0.030461
115	mustafaymn	Person	40	4	0.029947
116	cmylmz	Other	40	0	0.028967
117	hilal_kaplan	Other	40	0	0.028911
118	diyanetvakfi	Media	37	0	0.028881

Table A.2 (cont.) : First 150 users with the highest eigenvector centrality.

#	Username	Category	Indegree	Outdegree	Eigenvector Centrality
119	omerrcelik	Person	39	0	0.028459
120	haberturktv	Media	39	0	0.028405
121	cavs	Other	39	0	0.028297
122	melekle14337784	Other	37	0	0.028184
123	by	Other	38	1	0.027902
124	selvacam	Other	37	0	0.027673
125	hasandogan	Other	38	0	0.027632
126	12numaraorg	Other	38	7	0.027468
127	forzabesiktas	Other	38	0	0.027468
128	abuyukgumus	Other	35	3	0.027317
129	fatmasahin	Other	36	4	0.027052
130	tcmalatyav	Other	34	0	0.026946
131	birgun_gazetesi	Media	37	0	0.026884
132	erkankandemir	Other	36	0	0.026457
133	kanald	Media	36	0	0.026359
134	elazigspororgtr	Other	35	0	0.026124
135	ymskulubu	Other	36	0	0.026079
136	izmirbld	Govermental	36	0	0.026025
137	hasan_guzeloglu	Other	29	0	0.025964
138	fuatoktay	Other	34	2	0.02585
139	afadelazig	Institutional	35	0	0.025742
140	bekirpakdemirli	Other	33	0	0.02574
141	tbmmresmi	Other	35	0	0.025694
142	deprem	Earthquake	33	0	0.025564
143	milenyumkahin	Other	35	0	0.025468
144	yhyustun	Other	33	0	0.025016
145	selahattingrkn	Other	26	5	0.024931
146	mikdatca	Other	33	0	0.024746
147	atvcomtr	Media	33	0	0.024655
148	sevannisanyan_r	Other	34	0	0.024638
149	gencfborg	Other	31	0	0.024581
150	elaziggm	Other	31	12	0.024552

CURRICULUM VITAE

Name Surname : Gözde Merve Demirci

Place and Date of Birth : Istanbul, 09-12-1993

E-Mail : gozdemervedemirci@gmail.com

EDUCATION :

- **B.Sc.** : 2016, Istanbul Kultur University, Faculty of Engineering, Industrial Engineering
- **M.Sc.** : 2020, Istanbul Technical University, Faculty of Management, Industrial Engineering

PROFESSIONAL EXPERIENCE AND REWARDS:

- 2017-2020 ISBAK Istanbul IT and Smart City Technologies Inc at Supply Chain Department.
- 2020 he Scientific and Technological Research Council of Turkey (TUBITAK) 2224-A. International Scientific Activities Support Program
- 2016 The Scientific and Technological Research Council of Turkey (TUBITAK) 2241-A. Industry Oriented Undergraduate Thesis Support Program
- 2015 Turkish National Agency, Erasmus+ Programme Scholarship
- 2013-2014 Turkish National Agency, Erasmus Programme Scholarship
- 2011-2016 Assessment Selection and Placement Center (OSYM), Education Scholarship – 100%

PUBLICATIONS, PRESENTATIONS AND PATENTS ON THE THESIS:

- **Demirci, G.M., Keskin, R., Dogan, G.,** 2019. Sentiment Analysis Using Deep Learning, *2019 IEEE International Conference on Big Data*, 2215-2221. Los Angeles, CA, USA.