Social Network Analysis: Understanding User Behavior in Threads

Diana Wahyu Lestari, Rita Ambarwati
Management Science Doctoral Program, Pasundan University, Bandung, Indonesia
E-mail: ritaambarwati@umsida.ac.id

ABSTRACT

The presence of threads has enhanced users' ability to share stories, information and views in greater detail and effectively in the online world. Social media platforms allow users to interact, share content, and connect online. Therefore, the purpose of this research is to create a Social Network Analysis (SNA) network visualization and categorize the data to contribute to the understanding of user interaction on this platform. The research method in this study is the Social Network Analysis (SNA) approach to analyze the relationship between individuals in social networks. The results show that user participation in the Threads platform has a significant positive impact. The data shows a high level of participation, providing an understanding of user engagement in various topics. Social network analysis revealed characteristics of user interactions, such as the number of nodes, average connectedness, and relationship complexity. In addition, word categories and network visualizations provide insights into relationship patterns and topics in conversations.

Keywords: Threads, Twitter, Media Social, Social Network Analysis

INTRODUCTION

The digital era has evolved and the presence of social media platforms has transformed society into an information platform where they can share their views, experiences and opinions on various topics (Samrin & Akbar, 2023). The emergence of the internet provides a new space for people to communicate to the current group of Indonesians who use social media widely not only to receive information, but also as a space to discuss, share information, and communicate (Rofidah, 2021). Social media users are widely used by the general public as information, expressing opinions and other things that attract their attention, so it can be seen from the many applications created to fulfill the wishes of the community (Hadna et al., 2016). Applications used to access information. One of them is Threads, Instagram's new social media platform. The threads feature includes a search for the hottest and most talked about topics. In addition, threads provide information about what is happening (Rhein Rahmahsy Reshany & Santi Indra Astuti, 2023).

In July, a new social media platform, Threads: an Instagram app, took the internet by storm. The emergence of Threads created a lot of buzz. Threads are considered to be similar to Twitter. Of course, it made many users curious. Apparently, in less than a week, Threads surpassed the 100 million user mark (Ragam et al., n.d.). Threads are similar to Twitter in that they allow users to share text-based posts. However, the appearance and features of Threads are slightly different from Twitter, although some features are similar. Threads has several features such as viewing profiles, checking mentions, and more (Samrin & Akbar, 2023). Some of the reasons for Threads users to review the description in the Threads app are to find a collection of topics of conversation that are very popular at the moment in Threads. Threads is basically similar to Twitter, so it is a microblogging platform (a content that contains short information in the form of text). It allows users to express their thoughts and feelings in text form. Users will get feedback from other users. The response given can be in the form of re-sharing or commenting on the text given (Ragam et al., n.d.).

In previous research (Samrin & Akbar, 2023) has discussed Sentiment Analysis of Threads.
Application User Comments on the Google Play Store Using the Multinomial Naïve Bayes Algorithm. On the other hand, other studies have focused on Twitter Interaction Networks in Customer Engagement in E-Commerce and E-Health Performance Competition in Indonesia, using the Social Network Analysis (SNA) approach in Qualitative Descriptive Research (Navisha et al., 2023). In addition, there is also previous research that discusses Twitter as a medium for self-expression among millennials, by conducting a qualitative analysis using Johari Window self-disclosure theory (Mutiara et al., 2020).

The research being conducted by the author has similarities with previous research (Navisha et al., 2023) because both use the Social Network Analysis (SNA) method. However, the main difference from the research conducted by the author is the use of the Multinomial Naïve Bayes algorithm to perform sentiment analysis. Meanwhile, previous research focused more on the Social Network Analysis (SNA) approach in the type of qualitative descriptive research. Therefore, the research conducted by the author combines elements of sentiment analysis with the SNA approach to provide a more comprehensive insight into the data under study. Meanwhile, previous research (Mutiara et al., 2020) used Johari Window self-disclosure theory. Thus, the main differences between the two studies lie in the research methods, data collection techniques, and theories that serve as the basis for analysis. This research applies an assumption that threads users give reviews to others, so that two nodes with two edges are formed and connected to each other (Susanto et al., 2012). Therefore, a suitable method is needed to analyze the interaction of a large number of threads user reviews. Social network analysis (SNA) was chosen because it is able to process large amounts of data to provide an overview of user interactions in the network, and can quantify through network properties consisting of size, edges, density, modularity, diameter, average path length, average degree (Alisya Putri Rabbani et al., 2020).

Although SNA has great potential, there has been no significant use of Social Network Analysis (SNA) in the field of information and communication technology so far. Social media platforms allow users to interact, share content, and connect online. Therefore, the purpose of this research is to create a Social Network Analysis network visualization and perform data categorization to contribute to the understanding of user interactions on these platforms.

**Problem Statement**: SNA network analysis and categorizing data on Review Threads

**Research Question**: How to make SNA network visualization and categorize data on Review Threads?

**SDGs Category**: This research falls into the 17th category of partnerships for the goals, namely increasing global partnerships in achieving sustainable development goals [https://sdgs.un.org/goals/goal17](https://sdgs.un.org/goals/goal17)

**LITERATURE REVIEW**

**Social Media**

Defined as a software that allows users to exchange information, share photos, videos, text, and the like online by utilizing forums or blogs to create a virtual world space called social media. Another term for social media is social networking, where most people in the world use this device to make it easier to share and create content (Watie, 2016). Based on this description, the conclusion that can be drawn regarding the definition of social media is a device that utilizes the internet network in supporting communication activities, collaboration, interaction, sharing, and expressing themselves with other users to create virtual social ties (Puspitarini & Nuraeni, 2019).

The features available in social media are communication features (voice, video, chat, images, tags, like, share, tweet/retweet, and many more). Many people from various backgrounds, ages, as well as individual and organizational aspects, use social media to exchange information. With the increasing number of social media applications in society,
social media communication networks are also increasing. Social interaction plays an important role including communication in work, education, business, and many other fields (Ariyanti, 2022).

Google Play Store

A device that provides google digital content in various categories ranging from books, music, movies, games, applications, and the like is called the google play store. Users can access this software by using the android application or website. Google play store also features reviews and ratings from other users (Herlinawati et al., 2020). Google Play Store has various features. One of them allows users to leave reviews about the application. Thus allowing other users to get information from these reviews (Nishfi et al., 2023). This feature can influence potential users because the tendency of users before downloading will see the review section as a benchmark for whether the product is good or not.

App Store

The media provider of applications needed by the iPhone OS (IOS) based Operation System is the App Store. The party that manages and develops the app store is Apple Inc. The existence of this service makes it easier for Apple users to browse and download the applications they need. The App Store has applications organized by categories such as games, learning, and entertainment (Sandag, 2020). In addition, there are free and paid applications available. The categories on this app allow users to easily search for the apps they need (Hadi et al., 2020).

Social Network Analysis (SNA)

The scientific field that studies the relationship between the application of theory and humans is called SNA (Social Network Analysis). SNA is described through a network representation where information sharing in a social network is based on two fundamental elements, namely edges (relationships) and nodes (actors). The relationships formed between nodes can be analyzed in detail by involving visualization to obtain accurate and relevant information to the interests of users (Akbar et al., 2022). The function of SNA is to facilitate the mapping of relationships to optimize the formation of knowledge management. The components contained in the SNA network include average path length, diameter, average degree, edges, and nodes (Bratawisnu & Alamsyah, 2019).

METHOD

Social Network Analysis (SNA) technique is the method used in this research. SNA is associated with understanding social relationships that involve the correlation of edges (correlation between users as lines) and nodes (correlation between users as points) (Bratawisnu & Alamsyah, 2019). One type of qualitative research is research whose purpose is to analyze events or events in terms of actions, motivations, behaviors, perceptions, and others in detail expressed in the form of a series of words (Hukum et al., 2013). Researchers obtained secondary data sources in the form of data covering review content. This research requires the use of several supporting programs, including jupyter notebook, Sns scrape library, notepad++, wordij and gephi. The concept of the research flow is as follows:
Problem Identification
Observe and discover the Threads phenomenon

Data Collection
This stage of the research includes data collection. The purpose of data collection is to gather information that will be used to analyze a research. Running tools, setting up searchable subjects, and using text codes to create articles that are retrieved by the Google Colab app are the steps involved in data collection. Next, the researcher used crawling coding, written in Python, and used Gephi tools to build a social network model from the collected data. The researcher searched for datasets through Kaggle that were relevant to the subject.

Data Pre-Processing
Data preprocessing is an important part of preparing raw data for social network analysis. This process involves various methods to transform the raw data into a more structured and analyzable format. In addition, the preprocessing process on social media data such as Twitter also includes things like changing the base word form, removing unimportant words, removing affixes, and conjunctions from the tweet document. The process of preprocessing:

a. Case Folding
Case Folding is where all characters are converted to lowercase. The purpose of case folding is to eliminate the inconsistent use of uppercase and lowercase letters in user reviews. This process is done removing unused characters from the data such as (&,//*,0,(),etc).

b. Tokenizing
Tokenizing is separating words in a sentence for further text analysis process.

c. Stopword Removal
The stage of removing words that are meaningless and unrelated but often appear is called stopword removal.

d. Normalize
A series of activities that are part of the research stages that require researchers to change abbreviated words to the original word form and make standard words from non-standard words is called word normalization. How to find the word is classified as a standard word or not through KBBI (Kamus Besar Bahasa Indonesia), where researchers must check whether the word is listed in KBBI or not. Then the word change is based on the data set used in the research.

Network Modeling
Related to the stage of reprocessing data that has been processed by applying the Gephi application. This application makes it easy to visualize the network model by implementing an
undirected graph that ignores the direction of the relationship.

**Network Property Identification**

Each network model processed by the Gephi application has several properties that will be calculated. The network properties that will be calculated are: nodes, edges, average degree, modularity, and modularity.

**RESULT and DISCUSSION**

The crawling results in Table 1 show the number of tweets in threads, illustrate the level of user participation and engagement in various topics, provide insights into conversation trends, communication patterns, and the potential impact and dynamics of social networks; while further analysis can provide an in-depth understanding of the level of interaction, user responses, and characteristics of conversations on the platform.

<table>
<thead>
<tr>
<th>Focus</th>
<th>Source</th>
<th>Date</th>
<th>Amount of Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Review</td>
<td>Threads</td>
<td>July-August 2023</td>
<td>8.140</td>
</tr>
</tbody>
</table>

Table 1 shows that the number of tweets obtained from crawling threads with the keywords to be researched focuses on koten threads such as reviews. This study collected data from July to August 2023, as much as at least 1,000 data per year for each criterion, researchers only got the data in 2023 because the application became popular in July 2023, so researchers explored related to the application “threads” focusing on reviews. Then processed using Jupyter Notebook software. The data was obtained from the review keyword as much as 8,140 data in July to August 2023.

After collecting data (crawling), the next step is to preprocess the data. The help of a dictionary that aims to eliminate words that are not relevant to the analysis process, this process is called the filtering process, which removes unused words and then processes them using Wordij (Anjani & Ambarwati, 2023).

Table 2. Data Preprocessing Results

<table>
<thead>
<tr>
<th>Focus</th>
<th>Date</th>
<th>Amount of Data</th>
<th>Unique</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Review</td>
<td>July-August 2023</td>
<td>28.265</td>
<td>1.375</td>
<td>20.5564</td>
</tr>
</tbody>
</table>

Table 2 shows the results of data preprocessing using Wordij. This data was obtained from the preprocessing done using Jupyter Notebooks, which was saved as a CSV file. The next process involved reprocessing the data using the Wordij application, which enabled data visualization using Gephi. The Wordij application produces various output files, including files such as net, stp, sw, wrd, wtg, pr, and so on (Chamila & Sukmono, n.d.). In the visualization stage, the data used comes from the stw.csv file. This file contains information about the total number of words that appear from the keyword “review” threads in 2023, with a total of 28,265 words. Of these, there are 1,375 unique words, and the average word that appears is 20.5564. Visualization of this data is an important step after the preprocessing process to provide a clearer picture of patterns or trends that may be present in the data.

The analysis method applied in this research is Social Network Analysis (SNA). SNA is used to analyze social media activities, especially in modeling interactions. The interaction is represented as a network of relationships between users, represented by nodes and edges. This analysis is important because it provides a deeper understanding of the social interaction patterns of individuals or communities. Some of the network properties used in Social Network Analysis (SNA) involve nodes, edges, average degree, average weighted degree, diameter, modularity and average path length (Azmi et al., 2021). These properties help in mapping the relationships in the network, thus making a significant contribution in improving knowledge management and understanding of the dynamics of social interactions in the context of reviewing “threads” applications in 2023.
Table 3. Review Network Properties

<table>
<thead>
<tr>
<th>NO</th>
<th>Network Properties</th>
<th>Value (Review)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Nodes</td>
<td>276</td>
</tr>
<tr>
<td>2</td>
<td>Edges</td>
<td>1317</td>
</tr>
<tr>
<td>3</td>
<td>Average Degree</td>
<td>2,858</td>
</tr>
<tr>
<td>4</td>
<td>Average Weighted Degree</td>
<td>29,383</td>
</tr>
<tr>
<td>5</td>
<td>Network Diameter</td>
<td>7</td>
</tr>
<tr>
<td>6</td>
<td>Modularity</td>
<td>0.217</td>
</tr>
<tr>
<td>7</td>
<td>Average Path Length</td>
<td>2,356</td>
</tr>
</tbody>
</table>

Nodes are representations of user actors in the review social network. The data in table 3 shows that there are 276 nodes that interact with each other in the social network, with a total of 1317 edges or relationships. Average Degree shows the average number of connections between each node and other nodes. In the social network review, nodes have an average connectedness of 2,858 (Digital et al., 2022). Average Weighted Degree which is the average weighted relationship between nodes, reaches 29.383 on the review network (Putri et al., 2023). Network Diameter is the farthest distance between two nodes, in the review social network it reaches 7. Modularity reflects the ability of actors or users to form different groups in the network. In the review social network, the modularity value is 0.217. Furthermore, Average Path Length, which is the average number of nodes that must be passed to reach a certain node, on the review social network has a value of 2.356 (Bratawisnu et al., 2018).

Data Categories

The category analysis process in this study was categorized based on certain criteria, and the results of the analysis of the relationship between words formed four main categories, namely App, Action, Reaction, and Object. Then grouped by these categories, reflecting the general characteristics of the review threads observed in this study.

Table 4. Data Category Identification

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>app</td>
<td>10.6%</td>
<td>Delete</td>
<td>0.85%</td>
</tr>
<tr>
<td>twitter</td>
<td>6.62%</td>
<td>Follow</td>
<td>0.83%</td>
</tr>
<tr>
<td>instagram</td>
<td>3.20%</td>
<td>Option</td>
<td>0.76%</td>
</tr>
<tr>
<td>threads</td>
<td>2.74%</td>
<td>Post</td>
<td>0.72%</td>
</tr>
<tr>
<td>account</td>
<td>2.13%</td>
<td>download</td>
<td>0.50%</td>
</tr>
<tr>
<td>thread</td>
<td>0.79%</td>
<td>Login</td>
<td>0.45%</td>
</tr>
<tr>
<td>application</td>
<td>0.69%</td>
<td>Review</td>
<td>0.42%</td>
</tr>
<tr>
<td>meta</td>
<td>0.67%</td>
<td>Feed</td>
<td>0.38%</td>
</tr>
<tr>
<td>social</td>
<td>0.53%</td>
<td>Worst</td>
<td>0.33%</td>
</tr>
<tr>
<td>insta</td>
<td>0.48%</td>
<td>Posts</td>
<td>0.33%</td>
</tr>
</tbody>
</table>

Table 4 shows 10 words for each topic. “App” (Topic 1), “Action” (Topic 2), “Reaction” (Topic 3), and Object (Topic 4). Then, “App” (Topic 1) is categorized as a word because it refers to a specific field or topic, in this case related to technology or software. For example, the words “app” (10.6%), “twitter” (6.62%), “instagram” (3.20%), “threads” (2.74%) appear most frequently. Meanwhile, “Action” (Topic 2) is included in the category of words because its meaning is related to activities or actions, including words such as “delete”(0.85%), “follow”(0.83%), “option”(0.76%), “post”(0.72%). The use of these words provides...
a deep understanding of the activities that dominate conversations and the extent to which user interactions involve actions.

Furthermore, "Reaction" (Topic 3) belongs to the category of words as it refers to a response or response to a stimulus or event. It can be used in various contexts, such as in science, psychology, and everyday conversation, to describe how someone or something responds or reacts to a situation or change. For example, in the words "nice"(1.44%), "love"(1.06%), "bad"(0.07%), "amazing"(0.62%). Then, "Object" (Topic 4) belongs to the category of words because it can refer to a physical object or concept that is the focus of attention. This word has many usages involving various contexts, such as in programming languages, math, or everyday conversation, where "object" can refer to something concrete or abstract that is the topic of discussion or attention. For example, "people"(1.02%), "time"(0.66%), "features"(0.56%), "data"(0.51%)

Network Model Visualization

The next process is data visualization, the concept used in Social Network Analysis (SNA) is graph theory, consisting of nodes (points) connected by edges (lines). The visualization process requires supporting tools through Gephi software.

The network visualization in Figure 1 provides insight into the review topics in the Threads application. The graph is structured with five large nodes providing a strong visual representation of the centers of conversation among Threads users. These main clusters, marked with large nodes, include key words such as ("app", "twitter", "instagram", "threads", and "account"). This gives an idea of the main focus of the conversation, which is clearly centered around social media platforms and app features. The different colors of the nodes and edges play a key role in unlocking additional layers of information. With five different colors: purple, gray, green, blue and orange. This visualization signals specific groups and their relevance to the discussion. The largest group, colored purple, shows a primary focus on social media platforms, with words like ("app", "twitter", "meta" and "application"). This suggests that the discussion will include consideration and comparison of various features and functions of social media applications, such as Threads.

The second group, marked in green, emphasizes how important millennials are in the conversation. Words like ("instagram"
In the context of visualization “threads” indicate the intensity of related discussions, which means that social media platforms to thread content that often appear and are related. The two nodes “threads” and “Instagram” indicate how often “threads” and “Instagram” are used in related applications. This can provide insight into user trends and interests in conversations on these social media platforms. Thus, the threads application has significant value in various aspects of business and product development. This is in line with research that has stated that users provide an in-depth view of the product (Samrin & Akbar, 2023).

In social network analysis, the nodes that often appear are the platforms “Twitter” and “Instagram”, which are frequently used by millennials. The focus is on understanding how information and interactions spread among users, including community identification, influence, retweet or share patterns, and analysis of shared content. The research being conducted by the author has similarities with previous research (Navisha et al., 2023) because both use the Social Network Analysis (SNA) method. However, the main difference from the research conducted by the author is the use of the Multinomial Naïve Bayes algorithm to perform sentiment analysis. Meanwhile, previous research focused more on the Social Network Analysis (SNA) approach in a qualitative descriptive research type. In the previous study (Navisha et al., 2023), the main emphasis was on using the Social Network Analysis (SNA) method to understand interaction patterns and user behavior on social media platforms that identify user communities of “twitter” and “instagram” applications, measure the influence between users, analyze retweet or share patterns, and conduct analysis of shared content to understand certain trends or patterns in online interactions.

Along with previous studies, the research being conducted by the author also uses the Social Network Analysis (SNA) approach to understand the dynamics of social networks on the “Twitter” and “Instagram” platforms.
However, the main difference lies in the focus of the analysis. While the previous study highlighted more on network description and mapping as well as user behavior. In general, the author's research adds the dimension of sentiment analysis by using the Multinomial Naïve Bayes algorithm. This allows the authors to not only understand the structure and dynamics of the network, but also to explore how users respond to and perceive the content they use or share on the platform.

Thus, while there are similarities in the use of Social Network Analysis (SNA) methods, the authors' research adds value by expanding the scope of the analysis into the sentiment domain, which can provide more comprehensive insights into user behavior and preferences on the "Twitter" and "Instagram" platforms.

CONCLUSION
The results show that user participation in the Threads platform has a significant positive impact. The data shows a high level of participation, providing an understanding of user engagement in various topics. Social network analysis revealed characteristics of user interactions, such as the number of nodes, average connectedness, and relationship complexity. In addition, thematic word categories and network visualizations provide insight into the patterns of relationships between keywords and topics in conversations. In light of this, the Threads platform has rapidly gained popularity, exceeding 100 million users in a short period of time, demonstrating the potential for user growth and becoming the first choice for users to interact by sharing social content.

However, this study has limitations in the data available, the analysis methods used, and the overall research results. Although data mining was done accurately, the limited data may not cover all topics in Threads, affecting the representation of the research results. The use of Social Network Analysis (SNA) method, and Multinomial Naïve Bayes algorithm may have limitations in capturing the complexity of social interactions and it is difficult to generalize the research results. Therefore, the development of more modern analysis methods and wider data collection and validation of research results are required. This reinforces the importance of understanding the limitations of the study and encourages further research to provide deeper insights for decision makers in designing more effective social media strategies.

REFERENCES
Information Engineering and Educational Technology, 2(1), 1.
https://doi.org/10.26740/jieet.v2n1.p1-5


https://doi.org/10.31294/bi.v8i2.8191

https://doi.org/10.31294/bi.v8i2.8191

https://doi.org/10.31294/bi.v8i2.8191

https://doi.org/10.24114/cess.v5i2.18186


https://ejournal.unsrat.ac.id/index.php/actadiurnakomunikasi/article/view/31979/30381

https://doi.org/10.30587/jurnalmanajerial.v10i02.5279


https://doi.org/10.34010/common.v3i1.1950

https://doi.org/10.21070/ups.2461


https://doi.org/10.30596/interaksi.v5i2.5899


https://doi.org/10.21460/inf.2012.81.111

https://doi.org/10.26623/themessenger.v3i2.270