Pioneering Data-Driven Decisions: The Future of Predictive Modeling in Data Science

Kartal DERİN

Smartpro Technology

To Cite This Chapter

Derin, K. (2024). Pioneering Data-Driven Decisions: The Future of Predictive Modeling in Data Science. In M. Hanefi Calp & R. Bütüner (Eds.), *Current Studies in Data Science and Analytics* (pp. 28-37). ISRES Publishing

Introduction

In today's rapidly evolving digital era, businesses across numerous sectors are increasingly turning to data-driven decision-making to enhance operational efficiency and achieve strategic objectives. At the core of this shift lies predictive modeling, a powerful analytical tool that leverages historical data to anticipate future outcomes. By employing machine learning algorithms and statistical techniques, predictive models continuously improve, providing organizations with unparalleled insights into consumer behavior, market trends, and operational risks. This paper examines the profound impact of predictive modeling across various industries, including healthcare, education, business, and technology. From AI-driven diagnostic tools to real- time gesture recognition systems and drones, predictive analytics is driving innovation in diverse fields. The study also discusses the evolution of predictive modeling, tracing its roots from traditional statistical methods to advanced machine learning techniques that form the foundation of future data science.

As emerging technologies like quantum computing, federated learning, and edge AI come into play, predictive modeling is poised to have an even greater impact. This paper offers a comprehensive overview of the current state of predictive modeling and its future potential, including an analysis of its applications and ethical considerations.

Figure 1
Predictive Modeling (original)

The Development and Progression of Predictive Modeling

Predictive modeling began with traditional statistical approaches, which offered a foundation for identifying patterns in data and predicting future events. Early methods, such as regression analysis and decision trees, were effective for smaller, less complex datasets. However, as computational power grew and data collection methods advanced, the demand for more sophisticated models emerged, leading to significant breakthroughs in the field.

By the late 20th century, machine learning began to replace traditional rule-based systems with more flexible models. Algorithms like support vector machines, decision trees, and neural networks enhanced the accuracy and complexity of predictive modeling. These algorithms could learn from vast datasets and uncover intricate patterns that were previously undetectable. The combination of big data, improved computational power, and advanced algorithms helped predictive modeling become widely used across industries during the early 21st century.

In recent times, predictive modeling has integrated deep learning techniques, allowing models to autonomously enhance their performance. Deep learning architectures, including convolutional neural networks (CNNs) and recurrent neural networks (RNNs), have enabled predictive analytics to excel in tasks such as image recognition, natural language processing, and behavioral prediction.

Figure 2

Machine Learning and Deep Learning (original)

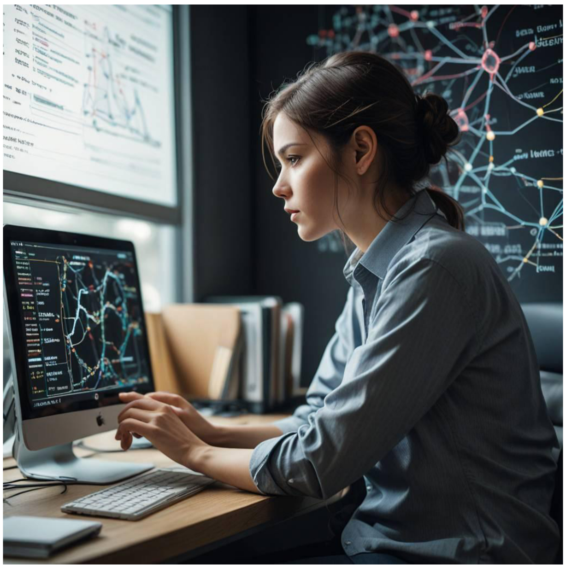
Machine Learning and Deep Learning in Predictive Modeling

The widespread application of machine learning (ML) and deep learning (DL) has fundamentally reshaped the landscape of predictive modeling. Where traditional models faced challenges with smaller datasets, ML algorithms now allow for the sophisticated analysis of much larger and more complex datasets. The ability of ML to refine models and improve prediction accuracy has introduced significant advancements in predictive analytics.

Machine learning uses both supervised and unsupervised techniques to reveal hidden patterns within data, leading to more accurate forecasts. Foundational ML methods, such as regression, decision trees, and support vector machines, have proven successful on large-scale datasets (Ghosh, 2021). Brown and Smith (2018) highlight that advanced machine learning techniques have significantly improved prediction accuracy, especially for large and complex datasets. On the other hand, deep learning is especially useful for

tackling unstructured or highly intricate data.
Deep learning models are inspired by neural architectures akin to the human brain, using layers of neural networks to process complex datasets. These models excel in areas like image recognition, natural language processing, and time-series analysis. Convolutional neural networks (CNNs) are frequently employed for tasks involving image recognition, while recurrent neural networks (RNNs) are better suited for sequential data processing.

Figure 3
Challenges and Opportunities (original



Data Sources for Predictive Modeling: Challenges and Opportunities

As predictive modeling advances, the diversity and complexity of data sources have significantly expanded. Traditionally, models relied on structured data from databases and historical records. However, with the advent of the Internet of Things (IoT), wireless sensing technologies, and social media, new dynamic data streams have emerged. While these sources present opportunities for deeper insights, they also introduce challenges related to data quality, consistency, and real-time processing.
The Internet of Things (IoT) has become a significant contributor to the vast amount of data available for predictive modeling. IoT-connected devices generate data that can predict everything from equipment malfunctions in industrial settings to consumer behaviors in smart homes.

However, IoT data often comes in unstructured or semi-structured forms, requiring advanced preprocessing to be used effectively.

Wireless sensing technologies, such as Channel State Information (CSI), add another dimension to predictive modeling. CSI data captures real-time wireless signals, which can be leveraged for human behavior detection systems, including gesture recognition. However, due to the large volume and variability of this data, sophisticated algorithms are needed to filter out noise and detect relevant patterns .

Figure 4

Human Gesture Recognition (original)

The Role of Neural Networks in Human Gesture Recognition

Neural networks have dramatically transformed human gesture recognition, providing powerful tools to interpret and predict human movements with remarkable accuracy. By leveraging deep learning architectures like CNNs and RNNs, gesture recognition systems can process and analyze real-time data to decode complex hand movements, body gestures, and facial expressions. These systems are applied in a variety of fields, ranging from human-computer interaction to advanced robotics and healthcare.

Figure 5

Real-World Applications (original)

Real-World Applications: Predictive Modeling in Industry

Predictive modeling has reshaped various industries by streamlining processes and enhancing decision-making. By utilizing large datasets with advanced algorithms, industries such as finance, manufacturing, supply chain management, and aerospace have significantly improved their performance, reduced costs, and forecasted future trends more accurately. The ability to offer actionable insights has made predictive modeling essential in today's highly competitive market landscape.

One notable application is financial forecasting. Financial institutions utilize machine learning models to analyze historical data, market trends, and economic indicators to predict future market movements. These models allow businesses to mitigate risks by detecting early signs of economic downturns, stock price fluctuations, and potential credit defaults. Additionally, banks rely on predictive models to assess the creditworthiness of their customers, offering more personalized and accurate loan offerings.

In supply chain management, predictive modeling helps companies optimize their operations by forecasting demand, managing inventory, and identifying potential disruptions. For instance, predictive models can anticipate changes in product demand by analyzing historical sales data, seasonal patterns, and real-time market fluctuations, allowing businesses to adjust production schedules and resource allocation efficiently.

Predictive modeling is also invaluable in project management, where it is used to forecast project timelines, predict budget overruns, and estimate resource requirements. By examining historical project data, managers can anticipate potential risks and inefficiencies, leading to more precise resource planning and timely project completion. Predictive models are particularly useful in large- scale software development projects, where delays can be costly and client satisfaction is paramount.

Figure 6

Real-World Applications2 (original)

In the aerospace industry, predictive modeling has become crucial for optimizing maintenance and improving safety protocols. Airlines rely on predictive maintenance models to forecast potential equipment failures, thereby minimizing costly downtime and ensuring passenger safety. By analyzing data from aircraft sensors and maintenance logs, these models predict when components are likely to fail, allowing preventive maintenance to be scheduled before an issue becomes critical.

Predictive modeling also plays a role in AI-powered drones. These systems use machine learning to predict flight paths, optimize navigation, and anticipate environmental conditions such as wind speed or obstacles. In sectors like agriculture, construction, and logistics, AI-powered drones equipped with predictive models survey large areas, monitor crop health, and deliver goods with precision. These drones can operate autonomously, adapting to dynamic environments in real-time (Zhang et al., 2021).

Finally, predictive modeling has found a place in the racing drone industry, where real-time data like speed, altitude, and trajectory are analyzed to optimize drone performance during races.

Predictive analytics helps identify potential risks or mechanical failures before they affect race outcomes, enhancing both performance and safety.

Figure 7
Healthcare and Education (original)

Predictive Analytics in Healthcare and Education

Predictive analytics has revolutionized healthcare and education by improving how these sectors utilize data to make informed decisions, enhance outcomes, and streamline operations. In healthcare, predictive models are used to diagnose diseases, forecast patient outcomes, and create personalized treatment plans. In education, predictive analytics powers personalized learning platforms, helping educators identify student needs and forecast academic performance.

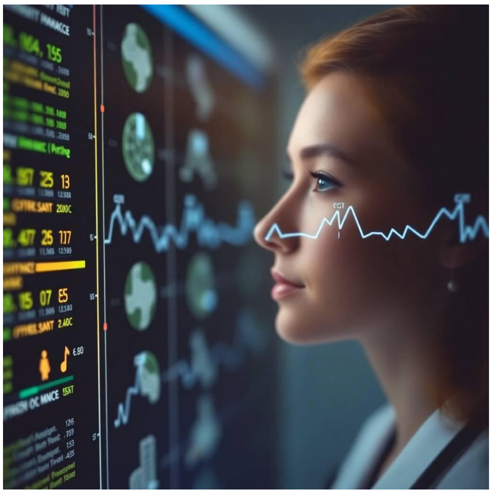
In healthcare, predictive analytics plays a key role in the early detection of diseases. By analyzing large datasets, including medical histories and imaging results, predictive models can identify patterns that signal the onset of diseases. For example, these models are increasingly being used to detect cancer at earlier stages by analyzing radiological images. Machine learning algorithms pick up on subtle changes in tissue that may not be visible to the naked eye, leading to earlier and more accurate diagnoses.

Additionally, predictive models can help anticipate outcomes for chronic conditions such as diabetes or heart disease. Data from wearable devices and real-time monitoring systems allow healthcare providers to intervene early, reducing hospitalizations. Predictive analytics also improves healthcare efficiency by predicting patient admissions and helping hospitals allocate resources better during peak periods .

In education, predictive analytics is transforming traditional models by providing personalized learning experiences. By analyzing student performance and engagement data, predictive models identify areas where students may need additional support, allowing teachers to tailor their approach. This personalized method improves learning outcomes. Predictive models also assist institutions in identifying at-risk students, enabling timely interventions to increase retention (Johnson & Johnston, 2019).

Figure 8

Real-Time Decision-Making (original)



Real-Time Decision-Making: Predictive Models in Action

One of the most valuable outcomes of advances in machine learning and AI is real-time decision- making, made possible by predictive models. Real-time analytics enables industries to respond quickly to changing environments, thereby improving decision-making efficiency and accuracy. In manufacturing, real-time object-counting systems powered by predictive models detect and track items on conveyor belts, ensuring precise counting and quality control. For example, in the food processing industry, predictive models count items such as oranges in real time, optimizing operations and reducing errors (Gonzalez & Peters, 2021). Similarly, in human gesture recognition, systems that integrate machine learning algorithms with computer vision allow users to control devices through hand or body movements. These systems are increasingly used in consumer electronics and healthcare for contactless interfaces. Real-time predictive models are also essential in healthcare for monitoring critical patients, allowing providers to intervene when necessary .

In autonomous systems, such as drones and self-driving cars, predictive models process real-time sensor data to navigate environments, avoid obstacles, and make quick decisions. This increases safety and operational efficiency, allowing drones to adjust to changing conditions such as weather or terrain (Zhang et al., 2020).

Figure 9

Emerging Trends (original)

The Future of Predictive Modeling: Emerging Trends

The future of predictive modeling is being driven by several important trends, including quantum computing, federated learning, and edge AI. These emerging technologies are enhancing the speed, precision, and scalability of predictive models, making them more effective and accessible for various industries.

Quantum computing holds significant potential to transform predictive modeling by processing enormous amounts of data at unprecedented speeds. In contrast to classical computers, which use binary bits, quantum systems rely on qubits, which can handle more complex calculations in less time. This improvement is especially advantageous for predictive analytics, where large data sets with numerous variables need to be processed efficiently. Quantum computing is expected to significantly advance machine learning algorithms, unlocking solutions to previously unsolvable challenges.

Federated learning is another innovative approach, addressing privacy issues by training models on decentralized data sources. This technique is particularly valuable in fields like healthcare and finance, where regulatory compliance and privacy are paramount. Federated learning allows predictive models to harness diverse data sets while safeguarding sensitive information.

Edge AI, which involves deploying predictive models directly on devices such as smartphones and sensors, eliminates the need for cloud-based processing. This minimizes latency and supports real- time decision-making, which is essential for applications such as self-driving vehicles. As edge computing technologies progress, predictive models will be able to process data locally, boosting the speed and efficiency of various operations .

Finally, Explainable AI (XAI) is gaining momentum as it addresses the issue of transparency in machine learning models. XAI aims to make models more understandable, providing insights into decision-making processes. This is crucial for industries like healthcare and finance, where transparency in decision-making builds trust and accountability (Miller, 2019).

Figure 10
Responsible AI(original)

Ethical Considerations and Responsible AI in Predictive Modeling

As predictive modeling becomes an essential part of decision-making across various industries, it raises critical ethical concerns like fairness, transparency, bias, and accountability. While predictive models provide significant insights, they also present challenges, particularly in maintaining fairness.

One of the primary concerns is bias. Predictive models, when built on biased data, may inadvertently reinforce or even magnify existing inequalities, resulting in unfair outcomes. This issue has been particularly highlighted in algorithms used in areas such as criminal justice and recruitment, where minority groups may be disproportionately disadvantaged. To combat this, it is crucial to utilize fairness-aware algorithms and ensure the inclusion of diverse, representative datasets that accurately reflect the population being analyzed.

Transparency is another major issue, especially with complex machine learning models that often function as “black boxes,” making it difficult to interpret how decisions are made. Garcia and Lopez (2020) emphasize that addressing ethical challenges, such as transparency and bias, is crucial for building trust and ensuring fair outcomes in AI-driven predictive analytics. In sensitive areas like healthcare, it’s vital for stakeholders to understand the rationale behind model decisions to foster trust. Explainable AI (XAI) provides a solution by clarifying how models arrive at their conclusions, thereby improving both accountability and trustworthiness.

Accountability is equally important as predictive models become embedded in the decision-making frameworks of organizations. Establishing proper oversight and audit mechanisms ensures that the ethical implications of these models are continuously evaluated. Regular assessments can help guarantee that predictive models remain fair and accurate.

Finally, privacy concerns become more pressing as predictive models increasingly rely on personal data. Federated learning offers a promising approach by allowing decentralized data processing, which helps safeguard privacy while still enabling robust predictive analysis.

Figure 12
Ethical (original)

Conclusion

Predictive modeling has firmly established itself as a key tool in modern data science, empowering industries to make informed, data-driven decisions. Technological advancements in machine learning, deep learning, and artificial intelligence have brought transformative improvements across fields like healthcare, finance, education, and manufacturing, enhancing operational efficiency, reducing risks, and optimizing decision-making processes.

The adoption of emerging technologies like quantum computing, federated learning, and edge AI will further boost the capabilities of predictive models, making them faster, more scalable, and more accurate. These advancements will enable predictive analytics to expand into new industries while addressing ethical issues like bias, transparency, and accountability.

To support the continued development of predictive modeling, it's essential to prioritize responsible and ethical usage. Focusing on fairness, transparency, and accountability will allow organizations to maximize the advantages of predictive models while minimizing potential risks, ensuring that these tools benefit all stakeholders.

References

Brown, J., & Smith, L. (2018). Advanced machine learning techniques in predictive modeling. Journal of Data Science and Applications, 6(4), 340–356. https://doi.org/10.1016/j.jdsapp.2018.06.003

Chen, P., & Zhang, Z. (2014).Title: Big data applications, opportunities, and challenges. Journal: Artificial Intelligence Review.DOI: https://doi.org/10.1007/s10462-014-9445-3

Garcia, M., & Lopez, A. (2020). Ethical challenges in AI-driven predictive analytics. AI Ethics and Society, 12(1), 45–60. https://doi.org/10.1080/AI.2020.0001

Montanaro, A. (2021).Title: Quantum computing and its impact on predictive analytics. Journal: Computing Futures, 5(3),67–79.DOI:https://doi.org/10.1016/j.cofut.2020.100051

Toma, M., & Ong Chi Wei, O. (2023).Title: Predictive modeling in medicine: A review. Journal: MDPI Journal of Medicine, 3(2), 590–601. DOI:https://doi.org/10.3390/encyclopedia3020042

About the Author

Kartal DERİN is a Data Scientist and the Head of the AI Department at SmartPro Teknoloji in İstanbul. With certifications from IBM, HP, and Google in data science and artificial intelligence, his expertise includes artificial intelligence, machine learning, and deep learning applications. Kartal has led and contributed to numerous global projects, particularly in healthcare and drone technology. His proficiency in Python programming has allowed him to develop advanced AI-powered solutions across various sectors. He is an active participant in AI research and frequently shares his insights and innovations on LinkedIn.
Email:kartalderinmail@yahoo.com ,**ORCID:** 0009-0009-7104-5955

Similarity Index

The similarity index obtained from the plagiarism software for this book chapter is 6%.

An Overview of Social Network Analysis: Metrics, Tools and Applications

Akça Okan YÜKSEL

Middle East Technical University

To Cite This Chapter

Yüksel, A. O. (2024). An Overview of Social Network Analysis: Metrics, Tools and Applications. In M. Hanefi Calp & R. Bütüner (Eds.), *Current Studies in Data Science and Analytics* (pp. 38-56). ISRES Publishing

Introduction

A social network is defined as a collection of social entities, such as individuals, groups, and organizations, connected by relational data with some interactions or relationships between them. Examples of such networks include friendship networks, follower networks, interaction networks, co-authorship networks, and spread networks (Tabassum et al., 2018). The two main components of any social network are entities and relationships (Scott, 2000). The combination of these two elements creates a social network. Entities may be individual people or collective actors, such as groups and organizations. Common examples of individual actors include students in a school, employees in a corporate firm, or members of a political organization. Collective actors could be companies, foundations, or political parties. Sometimes, networks consist of different types of entities, such as a healthcare system or an education system. A relationship is generally defined as a specific type of contact, connection, or bond between a pair of entities or a dyad (Wasserman & Faust, 2004). Relationships can be directed, where one actor initiates and the other receives (e.g., giving advice, selling), or undirected, where reciprocity occurs (e.g., chatting, collaborating). A relationship is not a characteristic of a single entity but is a common dyadic property that exists as long as both participants maintain it. The diverse relationships between individual and collective entities can represent network structures and explain their impacts. The specific type of relationship a researcher should measure depends on the research objectives. For example, a study on community networks will likely examine various neighborhood activities, whereas a study on banking networks will focus on financial transactions. Borgatti et al. (2009) classified the types of relationships in social networks. These classifications and examples related to them are presented in Table 1.

Table 1
Types of Relationships

Relationship Type	Example
Similarities	Being on the same team, attending the same school, same gender, similar hobbies
Relationships	Kinship, marriage, friendship

Interactions	Help, advice, recommendation
Flows	Information flow, personnel changes, international trade

Social networks are suitable subjects for both quantitative and qualitative studies as they contain both the structure and the content of social relationships (Coviello, 2005). The analytical approach used in studying social networks is Social Network Analysis. Social Network Analysis (SNA) is a powerful analytical method aimed at examining connections and interactions between individuals, groups, institutions, or devices to make inferences from these relationships (Edwards, 2010). SNA finds application not only in personal relationships and social circles but also in business, healthcare, education, biology, and many other fields. This analytical method provides data scientists, researchers, and analysts with a broad and versatile toolkit, enabling the understanding of complex network structures and the extraction of information from them.

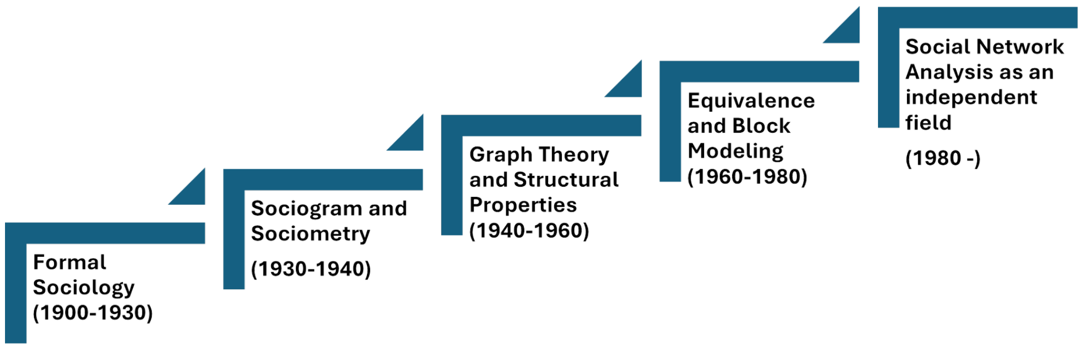
The rapidly growing subset of social networks is social media. Applications like Instagram, Twitter, LinkedIn, and WeChat facilitate daily information exchange. A significant concern is how the vast, complex data generated by online social network users can be searched, retrieved, stored, shared, processed, and visualized. SNA has become a widely applied method in research to investigate networks of relationships at individual, organizational, and societal levels. With the popularization of social networking sites like Facebook, Twitter, and Instagram, and the development of automated data collection techniques, the demand for SNA has recently increased significantly.

Social network analysis, with its interdisciplinary approach, is used in various fields and is of considerable importance. It offers the opportunity to analyze the relationships between social entities and the significance of these relationships (Oliveira & Gama, 2012). It allows the identification of similarities and differences in relationships between entities within social networks (Somyürek & Güyer, 2020). It also enables the integration of relationships and attributes in data structures (DeJordy & Halgin, 2008). In social network analysis, the visual presentation of data allows the entire structure to be seen as a whole and provides insight into the dynamics and effectiveness of any network (Lewis, 2011).

Historical Development of Social Network Analysis (SNA)

Social network analysis is a method that has evolved over time with the integration of various disciplines and has become an important tool in modern social sciences today. Social network analysis (SNA) has been used for a long time to represent complex relationships between participants in social systems at all scales. It is summarized the historical development process of SNA in five main stages (Somyürek & Güyer, 2020) (Figure 1).

Figure 1
Historical Development of SNA



Formal Sociology (1900-1930)

- George Simmel laid the theoretical foundation of modern social network analysis by focusing on the formal analysis of social interactions.
- Simmel examined different patterns of relationships, such as dyadic and triadic interactions, and argued that sociology should focus on the forms of social relationships.

Sociogram and Sociometry (1930-1940)

- Jacob Levy Moreno developed the techniques of sociogram and sociometry to study interpersonal relationships in small groups.
- Sociograms provided a visual dimension to network analysis by graphically representing social ties within groups.

Graph Theory and Development of Structural Features (1940-1960)

- Heider's (1946) work on group dynamics and balance made significant contributions to SNA.
- Cartwright and Harary (1956) mathematically analyzed social relationships in terms of balance and group dynamics.
- Graph theory, with its structures of nodes and edges, enabled a better understanding of social relationships.
- Modeling positive and negative relationships allowed for the identification of structural features like density and clustering in social networks.

Equivalence and Block Modeling (1960-1980)

- Harrison White and his students defined social structures through roles and relationships.
- Granovetter (1983) demonstrated that weak ties could be more effective than strong ties, highlighting the functionality of social relationship networks.
- Block modeling was used to group structurally similar nodes and to determine the fundamental features of the network.

Social Network Analysis as an Independent Field (1980-Present)

- In the 1980s, SNA emerged as a distinct research field within social sciences.
- INSNA (International Network for Social Network Analysis) was established, the Sunbelt Conference was organized, and the journal *Social Networks* began publication.

SNA has since taken a more analytical approach, developing its methodologies, theoretical expressions, and software. Software like UCINET, Gephi, PAJEK, and R packages has been developed, broadening the application of SNA.

Graph and Metrics

Social networks are generally presented in 2 ways. The first one is graphs. Graphs are structures that visually share information about social networks. Mathematical operations cannot be processed indirectly with analysis. The other is matrices. Since matrices allow for computational operations, detailed information is shared by conducting butchered analyses (Streeter & Gillespie, 1992).
Social networks are typologically classified as directed-undirected and binary-valued. If there are arrows between the links in the representation of a network, it is defined as a directed network; if there are no arrows, it is defined as a non-directional social network

(Tunalı, 2016). The other is classified as valued and binary according to the value of the link. The first one is the type that expresses the presence or absence of the links between nodes as 0-1 expressed as binary. The other is the type where the numerical value of a link indicates the density, strength, frequency, or volume of connections between pairs of nodes. This type is called valued (Tunalı, 2016).

Centrality

Centrality is an essential metric that indicates which node has a critical position within the network. If an actor has a high centrality value, it shows that this actor holds a central position in the network (Bloch, Jackson & Tabaldi, 2023).

In calculating centrality, the nature of the relationship is considered. Centrality is calculated based on whether the relationship is directed, undirected, weighted, or unweighted. In undirected networks, the degree of a node is the number of connections that node has. In directed networks, incoming connections to the node are referred to as in-degree centrality, while outgoing connections are referred to as out-degree centrality. The sum of the in-degree and out-degree is the total degree of that node.

Degree Centrality

In a network graph, degree centrality is measured by the total amount of direct connections to other nodes. It indicates the level of outward connectivity. Higher values suggest greater connectivity, indicating how central the node is relative to other nodes in the network (Laghridat & Essalih, 2023). In-degree centrality is based on relationships initiated by other users towards a user, while out-degree centrality is based on relationships initiated by a user towards others. Degree centrality assumes that all neighbors in the network are equally important. What matters is the number of neighbors. However, in many cases, if a node is connected to powerful nodes, its importance increases.

For undirected networks:

- Degree Centrality = Node's degree (number of connections) / (N-1)

For directed networks, it is divided into in-degree and out-degree centrality:

- In-degree Centrality = Number of incoming connections to the node / (N-1)
- Out-degree Centrality = Number of outgoing connections from the node / (N-1)

Betweenness Centrality

Betweenness centrality is used to measure a node that plays an 'intermediary' role in a network (Marin & Wellman, 2011). If a node is located on the only path that other nodes need to traverse, such as communication, connection, transportation, or transaction, then this node must be important and is likely to have a high betweenness centrality.

The betweenness centrality CB(v) for a node in a non-directional network is calculated by the formula.

$$c_{betweeneness}(v) = \sum_{s \neq t \neq v} \frac{\sigma_{st}(v)}{\sigma_{st}}$$

$$c_{betweeneness}(v) = \frac{c_{betweeneness}(v)}{(N-1)x(N-2)/2}$$

- s and t: Other nodes (node pairs) in the network,

- σ_{st} : The total number of shortest paths between nodes s and t,
- $\sigma_{st}(i)$: The number of shortest paths between nodes s and t that pass through node i
- N: The total number of nodes in the network

Closeness centrality

Closeness centrality is a measure of the total distance of a node from other nodes, if the length of the shortest paths of node N to other nodes in the network is small, then node N has a high closeness centrality (Zamanitajeddin et al., 2024). It refers to the convenience and ease of connections between the node of focus and other nodes.

Closeness centrality is calculated according to the following formula.
N: Total number of nodes in the network,
d (v,u): The length of the shortest path between node v and the other node u.

$$c_{closeness}(v) = \frac{(N-1)}{\sum_{u \neq v} d(v,u)}$$

The normalized closeness centrality for an undirected network can be expressed as follows.

$$c'_{closeness}(v) = \frac{c_{closeness}(v)}{(N-1)}$$

Eigenvector Centrality

This metric is based on assigning a relative score to each node and measures how well-connected a given actor is with other well-connected actors (Codal & Coskun, 2016). The main focus of eigenvector centrality is that the power and status of an actor is recursively defined by the power and status of its alters. Alters is a term often used in the egocentric approach of social network analysis and refers to actors who are directly connected to a particular actor, called the ego. What is noteworthy in this centrality measure is that the centrality of an individual depends on the centrality of all its neighbors with a positive constant. An individual with a high centrality of neighbors will also have a high centrality.

Eigenvector centrality is a more detailed version of degree centrality. It assumes that not all links have the same importance, taking into account not only the quantity but especially the quality of these links.

The eigenvector centrality x_i for node i in a network is calculated by the following formula.

Eigenvector centrality: $x_i = \frac{1}{\lambda} \sum_{j\epsilon\, N(i)} x_j$

x_i: Eigenvector centrality for node i.

λ: A constant scaling factor, eigenvalue

N(i): Neighbors of node i

x_j: The centrality of node j

This formula states that the centrality of each node is a function of the centralities of its neighbors. The eigenvalue λ is usually chosen to be the largest eigenvalue and the eigenvector centrality values are solved to find the eigenvector corresponding to the largest eigenvalue of the network.

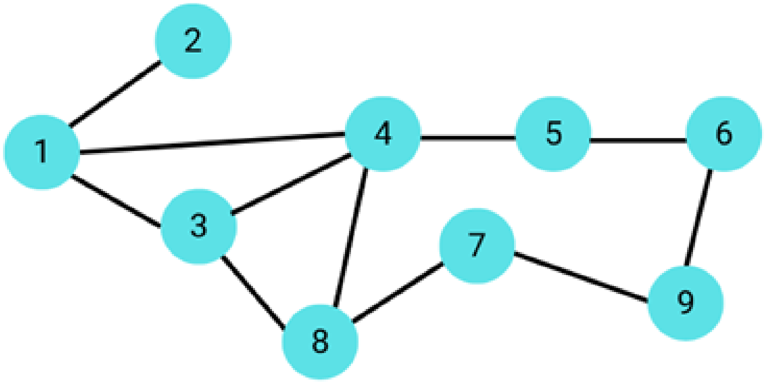
Pagerank Centrality

PageRank centrality is a special type of eigenvector centrality and is the ranking criterion of the popular search engine Google (Tunalı, 2016). The three different factors that determine the PageRank of a node are the number of incoming links, the link propensity of anchors, and the centrality of anchors (Gençer, 2023).

An Example of a Social Network on Centrality Values

Figure 2

A Sample Social Network Graph



The edges representing the connections between nodes in the social network shown are listed as follows.

Edges: (1, 2); (1, 3); (1, 4); (3, 4); (3, 8); (4, 5); (4, 8); (5, 6); (6, 9); (7, 8); (7, 9)

Average Degree of the Network

First, it is calculated the degrees of the nodes in this network to calculate the average degree.

Degree of node 1: 3; Degree of node 2: 1; Degree of node 3: 3; Degree of node 4: 4; Degree of node 5: 2; Degree of node 6: 2; Degree of node 7: 3; Degree of node 8: 2; Degree of node 9: 2

Total degree of the nodes: 3 + 1 + 3 + 4 + 2 + 2 + 3 + 2 + 2 = 22

Since the total number of nodes is 9, the average degree is: 22 / 9 = 2.44

Table 2

Centrality Calculations for the Social Network in Figure 1

Node/ Actor	Degree Centrality	Closeness Centrality	Betweenness Centrality	Eigenvector Centrality	PageRank Centrality
1	0.375	0.471	0.250	0.408	0.135
2	0.125	0.333	0.000	0.143	0.055
3	0.375	0.533	0.095	0.480	0.127
4	**0.500**	**0.615**	**0.405**	**0.540**	**0.167**
5	0.250	0.500	0.202	0.231	0.094

6	0.250	0.421	0.095	0.118	0.099
7	0.250	0.444	0.155	0.185	0.095
8	0.375	0.533	0.262	0.422	0.129
9	0.250	0.381	0.071	0.106	0.099

Table 2 shows centrality calculations regarding Figure 1.

In terms of degree centrality, node 4 is in the most central position (0.500).

The node with the highest closeness centrality is node 4 (0.615), meaning that node 4 can reach other nodes in the network at shorter distances. Node A is positioned closest to the center, allowing it to connect with other nodes either directly or through short paths, which indicates that node 4 plays a key role in the information flow within the network. The node with the lowest closeness centrality is node 2 (0.000). This node is located at the most peripheral position in the network and must take longer paths to reach other nodes. The low closeness centrality of node 2 suggests that it is distant from the center of the network and participates less in information flow.

Betweenness centrality shows how much a node acts as a “bridge” between other nodes in the network. Calculations reveal that node 4 has the highest betweenness centrality value, at 0.405. According to these results, node 4 stands out as the most critical node on the shortest paths in the network. Node 2 has a betweenness centrality of zero, indicating that it does not act as a bridge between other nodes.

Eigenvector centrality is a measure based on how central a node’s neighbors are. A node with high eigenvector centrality is more connected to central neighbors. Node 4, with the highest value, stands out as the most central node in the network, indicating that it is associated with neighbors who also hold highly central positions.

PageRank centrality calculates the importance of a node based on the centralities of the nodes that link to it. Originally used to rank web pages, this method is also widely used to identify influential nodes in networks. Node 4 has the highest PageRank value, indicating that it is in an important central position in the network. Node 2 has the lowest PageRank value, which suggests that its influence within the network is relatively low.

Local Clustering Coefficient

It is the ratio of the number of links between a node and its neighbors to the number of possible links they could have. In other words, it is a measure of the degree to which a node clusters with its neighbors.

The Local Clustering Coefficient indicates the probability that a node’s neighbors will connect with each other. The value ranges from 0 to 1, with 1 indicating that all neighbors are connected.

The clustering coefficient C_i for a node i is calculated by the formula.

$$c_{cc(i)} = \frac{2x(numberoftriangles)}{dx(d-1)}$$

C_{CC}(i): local clustering coefficient for node i

d= Number of neighbors of the node

triangle= Number of available triangles

The local clustering coefficient of each node in an example social network of Figure 1 is calculated and presented in Table 3.

Table 3
Local Clustering Coefficient

Node	1	2	3	4	5	6	7	8	9
Local Clustering Coefficient	0.33	0.00	0.67	0.33	0.00	0.00	0.00	0.33	0.00

Node 3 has the highest clustering coefficient with a value of 0.667, indicating that its neighbors have strong ties to each other.

For nodes 2, 5, 6, 7, and 9, the clustering coefficient is 0, meaning that there are no connections between the neighbors of these nodes.

Whole Network Metrics

Size

The size of a network is determined by the number of actors in that network. When we consider a school as a social network, if both students and teachers play the role of actors, then all individuals will constitute the size of the network. When considering politics, the network size represents the number of people with whom an individual discusses political topics.

Average Degree

One of the metrics that defines the overall structure of a network based on degree is the average degree of all nodes in the network. This is calculated differently depending on whether the network is directed or undirected. Letting k_i be the degree of node i, the average degree of an undirected network with N nodes and E edges is calculated as follows:

$$\text{Average Degree} = \frac{1}{N}\sum_{i}^{n} k_i = \frac{2E}{N}$$

Density

The density of a network is equal to the total number of connections divided by the number of possible connections. The number of possible connections assumes that each person can have a connection with every other person. The normalized range varies from 0-1. It represents the extent of communication within the network. Higher numbers (above .03) indicate faster information diffusion and greater group cohesion (Aboelela et al., 2007).
The density of the network shown in Figure 1 is calculated as the number of available edges divided by the maximum number of possible edges.

The density of the network according to the formula density = $\frac{2xE}{Nx(N-1)}$

E= Number of edges; N = Number of nodes

In this network

Total number of nodes N=9

Number of sides E=11

The density is calculated as = (2x11) / 9 x (9-1) = 22 / 72 = 0.306. The density of this network corresponds to approximately 31%.

Centralization

Centralization is based on the extent to which the majority of links are connected to a small set of nodes (Scott, 2000). It indicates whether there is an asymmetry in the distribution of connections. It indicates the degree to which communication is centralized around a single agent or a small group. More centralized groups tend to be more hierarchical in nature.

To calculate the centralization value of an example social network plotted in Figure 1, the following steps are followed:

1.Degree Centrality is calculated.
2.Determine the Maximum Degree of Centrality.
3.Centralization Value: The sum of the differences between the degree centrality of all nodes and the highest degree centrality.

This sum is divided by the theoretical sum of differences that could have the highest degree centrality in the network.

- C_{maks}: The highest degree centrality (centrality of the node with the highest degree).
- C_i: Degree centralities of other nodes.
- The value in the denominator is used as the theoretical maximum centralization value.

The degrees of each node in the network were calculated earlier.

- Node degrees: 3,1,3,4,2,2,3,2,2
- Highest degree: $_{Cmaks}$=4

Calculation of centralization

1.Find the differences between C_{max} - C_i for each node.

Table 4
Calculation of Centralization

c_{maks}- c_i	c_{maks}- c_1	c_{maks}- c_2	c_{maks}- c_3	c_{maks}- c_4	c_{maks}- c_5	c_{maks}- c_6	c_{maks}- c_7	c_{maks}- c_8	c_{maks}- c_9
Difference	4-3	4-1	4-3	4-4	4-2	4-2	4-3	4-2	4-2
Result	1	3	1	0	2	2	1	2	2
Total					14				

2.Calculate the theoretical maximum difference sum. In this case, the theoretical maximum difference occurs when the degree centrality of all nodes except the most central node is zero.

$$maks\sum(C_{maks} - C_i) = (N-1)xC_{maks} = 8\times4=32$$

3.Centralization=14/32=0.438

The centralization value of this network is 0.438.

Reciprocity

It is defined as the calculation of whether the connections between nodes are reciprocal, i.e. bidirectional. This metric, calculated in directed networks, is the ratio of the number of node pairs in the network to the number of all possible node pairs. The expression "follow to follow", which is frequently used in social media, refers to the reciprocity metric (Cheng et al., 2011).

SNA Software Tools

Tools for social network analysis are predominantly used for constructing networks, visualizing and manipulating network structures, conducting qualitative and quantitative/statistical analyses, detecting communities, and performing predictive analysis. Despite the availability of many tools, the most widely used ones include Pajek, Gephi, UCInet, NodeXL, R libraries, and NetworkX (Oliveria & Gama, 2011):

- **Pajek**: A free tool designed for analyzing and visualizing large-scale networks.
- **Gephi**: An open-source platform for network manipulation and exploration, featuring a three-dimensional render engine for displaying networks that evolve in real-time.
- **UCInet**: A commercial tool for social network analysis, which uses Pajek and NETDRAW for visualization. It is particularly well-suited for statistical and matrix-based analyses.
- **NodeXL**: A free add-in for Microsoft Excel, providing an accessible, user-friendly way to explore and visualize networks without requiring programming knowledge. However, it is not ideal for analyzing large networks.
- **R libraries** (e.g., igraph, sna, tnet, statnet): Free packages within the R environment, offering a comprehensive set of tools, including a large array of algorithms for community detection, longitudinal network analysis, and two-mode network analysis, with effective two- and three-dimensional visualization options.

NodeXL

NodeXL is mainly used for analyzing networks. It is mostly implemented as an add-in to Microsoft Excel. With the collection of network data, NodeXL provides quick statistics and reporting for people who can use the basic features of Microsoft Excel to analyze network data. NodeXL is a highly effective tool for analyzing and visualizing a social network. In addition to visualizing the entire network in the form of a graph, it can also draw graphs of different social network properties such as Proximity Centrality, Betweenness Centrality, Vertex Degree, Vertex PageRank, etc. Together with Nodexl, it enables network analysis by collecting data from social media platforms such as Twitter, Facebook, YouTube, and Flickr. In addition, topics that are on the agenda on Twitter can be analyzed and analyzed.
The analysis process with NodeXL generally consists of the following steps.

- Importing data
- Data preparation
- Grouping with clustering
- Calculating metrics
- Time series analysis
- Text analysis
- Identifying the Most Important Elements of the Network

- Visualizing the network

GEPHI

GEPHI is an open-source, independent software for visual and network analysis. The primary benefit of utilizing GEPHI for network research is its capability to handle extensive data sets or networks. The GEPHI program possesses certain drawbacks. Occasionally, the response time for a little task or procedure is excessively prolonged. For instance, accessing a file requires considerable time. Given that GEPHI is independent open-source software, it offers numerous functionalities. GEPHI is capable of importing data from text files (TXT), spreadsheets (CSV), and databases. GEPHI is capable of receiving information from various other social network analysis tools. GEPHI facilitates the straightforward graphical representation of a network. GEPHI is capable of producing network graphs and visual representations.

UCINet and Netdraw

UCINet is a menu-based application for social network analysis (Wu & Duan, 2015). UCINet is independent software. In UCINet, all data are represented as matrices. UCINet accepts two categories of input and produces two categories of output. The input comprises input parameters and datasets, whilst the output consists of output text and datasets. The spreadsheet editor is utilized for modifying, inputting new data, and converting UCINet data to Excel or SPSS formats. The UCINet spreadsheet accommodates tiny networks. For extensive datasets, multiple data formats are provided, accessible through an editor known as the dl editor.

UCINet employs Netdraw to facilitate network visualization. Netdraw facilitates various layouts for visualization objectives. These encompass isolates, components, subgroups, and centrality metrics. It also offers functionalities such as node restoration, color scheme adjustment, property visibility toggling, and node size modification, among others.

PAJEK

PAJEK is a software application designed for the visualization and analysis of extensive social networks. It is adequate to calculate most centrality measurements. Furthermore, functions that require many applications can be stored for subsequent re-analysis. PAJEK supports fundamental operations such as subnetwork extraction, identification of linked components (strong, weak, connected), determination of shortest pathways, calculation of maximum flow, centrality assessment (closeness, betweenness, degree, etc.), fragment search, and community detection. The findings generated by PAJEK can be further analyzed with R programming and SPSS. PAJEK accommodates bimodal networks, temporal networks (networks that evolve over time), acyclic and multi-associative networks (many interactions established among the same vertices), and signed networks (networks with both negative and positive connections). PAJEK additionally facilitates text-mining algorithms for the investigation of social networks (Majeed et al, 2020).

NetworkX

NetworkX is a powerful Python library for graph and network analysis. It is used to model and analyze graph structures such as social networks, biological networks, route optimization, connectivity analysis and many more. It allows you to analyze nodes and edges in network or graph structures. It is advantageous for research and applications where new metrics or algorithms need to be developed. Various types of graphs (directed, undirected, weighted, etc.) can be easily calculated with NetworkX. It supports reading and writing data from various formats (e.g. GraphML, GML, Pajek). Together with Matplotlib and other visualization libraries, graph structures can be visualized. In summary, NetworkX is a popular tool for both simple and complex network analysis.

R Programming and Packages

The R programming language has many libraries for analyzing social networks. One of these libraries is the igraph package. igraph is a fast, efficient, constantly updated, and highly preferred package. Other packages such as sna, tidygraph, and network have a large user base, especially among those interested in statistical network modeling (Gençer, 2023).

Social Network Visualizer

Social Network Visualizer (SocNetV) is a cross-platform, user-friendly free software application for social network analysis and visualization. With SocNetV, the following operations can be performed.

- Draw social networks with a few clicks on a virtual canvas, upload domain data from a file in a supported format (GraphML, GraphViz, Adjacency, EdgeList, GML, Pajek, UCINET, etc.), or browse the internet to create a social network of connected web pages.
- Organize actors and ties through point-and-click, graphs and social network properties that can be analyzed.
- Generate HTML reports.
- Standard graph and network fit metrics such as density, diameter, geodesics and distances, connectivity, eccentricity, clustering coefficient, reciprocity, etc.

VOSviewer and Bibliometrix

The 2 tools used for bibliometric analysis are VOSviwer and Bibliometrix. They are used to process and analyze data extracted from databases such as Scopus or Web of Science. They allow for analysis based on keywords in articles or on the relationships established through authors. Bibliometrix is an R package with a programmable structure (Gencer, 2023).

SNA in Higher Education

Social Network Analysis (SNA) in higher education enables the analysis of relationships and interactions between students, academics, and other stakeholders through network structures. With this method, information flow, collaboration, interaction patterns, and internal dynamics in higher education institutions are analyzed and patterns are revealed. Examples of social network analysis in higher education are listed below.

Examining Student Interactions

Academic achievement and social relationships: How students build social networks inside and outside the classroom can have an impact on their success. The SNA can be used to identify which groups students are more active in and the relationship between academic performance and social interaction.

Improving learning environments: Network structures of interactions in group work, project teams or online platforms (e.g. LMS) are analyzed. It can be revealed which students remain isolated or which groups collaborate more. For example, by analyzing the frequency with which students help each other or share information in the classroom, academic support mechanisms can be better shaped.

Examining Academic and Interdisciplinary Collaboration

Publication and project collaborations: Networks of articles and projects produced by academics together are analyzed. It is revealed which academics play central roles and how interdisciplinary collaborations are shaped.

Strengthening research networks: Areas of intra- or inter-institutional collaboration can be identified and incentives can be given to units that do not have strong ties. For example, by mapping the collaboration network of faculty members in a faculty, joint research projects can be proposed for academics who remain disconnected or isolated.

Analyzing Online Learning Environments

Learning analytics: Students' interactions with each other or with instructors through the LMS (Learning Management System) are analyzed. Student engagement can be increased by analyzing forums, discussion groups, and messaging networks.

Identify students at risk of failure: Isolated students can be identified in advance and counseling and support services can be provided.

Analysis of Management, Leadership, and Organization Networks

Internal decision-making processes: More effective governance can be achieved by analyzing the collaboration and communication structures between administrators, and academic and administrative units.

Leadership and information flow: It can be determined which administrators or academic units are central in information and decision flow, thus optimizing institutional processes.

Alumni tracking systems: Examine how alumni interact through job and career networks. Collaborating with alumni can contribute to the university's career network.

Mentoring networks: By establishing links between alumni and students, students can receive career guidance from alumni.

Policy Development and Performance Measurement

Department and program performance: Strategic plans can be created by examining inter-departmental collaboration networks. In addition, according to the network structures, it is determined which departments need more collaboration.

Innovation and entrepreneurship ecosystems: By analyzing the university's innovation centers and entrepreneurship networks, the ecosystem can be made to work more effectively.

As a result, social network analysis provides a better understanding of the relationships between students, academics, and management structures in higher education. In this way, it is possible to optimize information flow, increase collaboration, and support academic success.

Conclusion and Future Trends

This paper provides a comprehensive overview of the basic principles and objectives of Social Network Analysis (SNA) methods and their applicability to different types of networks. Various SNA metrics and related tasks are described with respect to the different structures of networks. Nowadays, the discovery of network structures in the data generated by many applications and the quality of the information extracted from these networks is increasing the popularity of network analysis. One of the main reasons for this increased interest is the analysis requirements arising from the ever-growing and more complex amount of data. At this point, effectively processing, managing, and extracting meaningful results from large volumes of data flowing at high speed is one of today's most important challenges. Especially with the proliferation of new technologies such as Web 2.0, Internet of Things (IoT), and Industry 4.0, the need for network analysis increases and analysis processes become more demanding and challenging. Therefore, the development of innovative methods that can cope with high volumes of data in network analysis stands out as one of the most current and critical requirements in this field.

Current trends in social network analysis are changing rapidly in line with technological and social developments. Interest and application areas of social network analysis are expanding in areas such as artificial intelligence, data privacy, community analysis, sentiment analysis and language processing, ethical issues, location-based social network analysis, micro-impact analysis, Multilayer Network analysis, disinformation detection, and decentralized networks. These trends indicate that in the future, SNA will become more effective and have more diverse use cases in both academic and commercial applications.

It is also expected that the popularity of SNA will continue to grow, attracting more researchers to the field and pushing an increasing number of companies to incorporate SNA methods into their business processes and expand their use as strategic tools.

Sample Social Network Analyses

Example 1

Scenario: Let's examine how metrics (degree centrality, betweenness centrality, closeness centrality, eigenvector centrality, local clustering coefficient, density and centralization) are calculated based on students' relationships in a social network of 10 students.

Students: A, B, C, D, E, F, G, H, I, J

Assume that friendship relationships are formed as follows.

- A: B, C, D
- B: A, E
- C: A, D, F
- D: A, C, G
- E: B, H
- F: C, I
- G: D, J
- H: E
- I: F, J
- J: G, I

In the sociomatrix formed according to this network of relationships, each row and column represents an individual. The corresponding cell indicates whether the individuals are connected to each other. It is coded 1 if there is a connection and 0 if there is no connection. The sociomatrix created for individuals whose names are coded as A, B, C, D, E, F, G, H, I, and J is given in Table 5.

Table 5
Relationship Matrix

	A	B	C	D	E	F	G	H	I	J
A	0	1	1	1	0	0	0	0	0	0
B	1	0	0	0	1	0	0	0	0	0
C	1	0	0	1	0	1	0	0	0	0
D	1	0	1	0	0	0	1	0	0	0
E	0	1	0	0	0	0	0	1	0	0
F	0	0	1	0	0	0	0	0	1	0
G	0	0	0	1	0	0	0	0	0	1
H	0	0	0	0	1	0	0	0	0	0
I	0	0	0	0	0	1	0	0	0	1
J	0	0	0	0	0	0	1	0	1	0

The following table shows the formulas for the degrees and the calculated centrality values of each node and the local clustering coefficient.

Table 6
Metric Calculations for Example Network

	Degree Centrality	Betweenness Centrality	Closeness Centrality	Eigenvector Centrality	Clustering Coefficient
Formula Node	$C_{degree} = \frac{DegreeofNode}{N}$	$c_{betweeness}(v) = \sum_{s \neq t \neq v} \frac{\sigma_{st}(v)}{\sigma_{st}}$ $c_{betweeness}(v) = \frac{c_{betweeness}(v)}{(N-1)x(N-2)/2}$	$c_{closeness}(v) = \frac{(N-1)}{\sum_{u \neq v} d(v,u)}$ $c'_{closeness}(v) = \frac{c_{closeness}(v)}{(N-1)}$	$x_i = \frac{1}{\lambda} \sum_{j \in N(i)} x_j$	$c_{cc(i)} = \frac{2x(numberoftriangles)}{dx(d-1)}$
A	0.33	0.5	0.50	0.49	0.33
B	0.22	0.39	0.41	0.24	0.00
C	0.33	0.28	0.47	0.49	0.33
D	0.33	0.28	0.47	0.49	0.33
E	0.22	0.22	0.32	0.11	0.00
F	0.22	0.17	0.39	0.27	0.00
G	0.22	0.17	0.39	0.27	0.00
H	0.11	0.00	0.25	0.044	0.00
I	0.22	0.06	0.33	0.17	0.00
J	0.22	0.06	0.33	0.17	0.00

The results for the 3 metrics that interpret the whole network (average degree, density and centralization) are as follows.

Average Degree

Average Degree = $\frac{(2xE)}{N}$ = (2x11)/10 =2.2

The average degree of this network is calculated as 2.2. This means that each node in the network has 2.2 links on average.

Density

The density of a network is calculated as the ratio of the number of available edges to the maximum possible number of edges between all nodes in the network.

$$D = \frac{2E}{(Nx(N-1))}$$

E: Number of available edges in the network => E=11

N: Number of nodes in the network => N=10

D= (2*11) / (10*9) = 0.244

The density of this network is calculated to be approximately 0.244. This means that about 24% of the node pairs in the network are directly connected by an edge.

Centralization

The centralization value of the network is a measure of the centrality differences of the nodes in the network and is calculated by the formula.

Centralization = $C = \dfrac{\sum_{i=1}^{N}(c_{maks} - c_i)}{(N-1)x(N-2)}$

C: Network centralization value

C_{maks}: Degree centrality of the node with the highest degree

C_i: Degree centrality of each node

N: Number of nodes (N=10)

C_{maks}= 3

Table 7

(C_{maks} - C_i) Values

	A	B	C	D	E	F	G	H	I	J
C_i	3	2	3	3	2	2	2	1	2	2
C_{maks} - C_i	0	1	0	0	1	1	1	2	1	1

Sum of C_{maks} and C_i differences = 0+1+0+0+1+1+1+2+1+1 = 8

Centralization = C = 8/ (9X8) = 0.11

The centralization value of this network is calculated to be approximately 0.111 or 11.1%. This value indicates that the degree of centralization of the network is quite low and that a central structure between nodes is not very obvious.

Example 2

Let's create a graph of the image for the nodes whose connections are given in Table 8.

Now let us examine the neighborhood matrix (sociomatrix) to show the relationships of the nodes. The number 1 in the cells where the rows and columns intersect indicates that there is a relationship for the intersecting nodes, while the number 0 indicates that there is no relationship.

Table 8

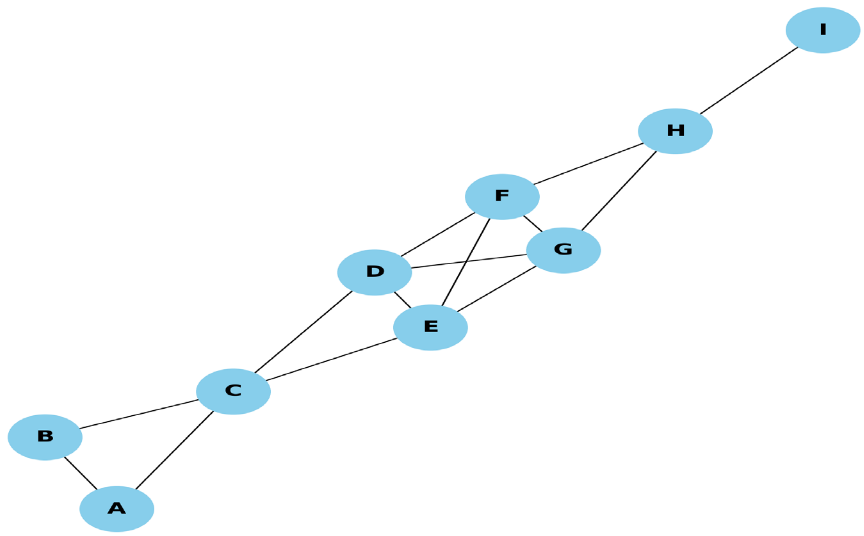
Sociomatrix

	A	B	C	D	E	F	G	H	I
A	0	1	1	0	0	0	0	0	0
B	1	0	1	0	0	0	0	0	0
C	1	1	0	1	1	0	0	0	0
D	0	0	1	0	1	1	1	0	0
E	0	0	1	1	0	1	1	0	0
F	0	0	0	1	1	0	1	1	0
G	0	0	0	1	1	1	0	1	0
H	0	0	0	0	0	1	1	0	1
I	0	0	0	0	0	0	0	1	0

For example, when node C is examined, it will be seen that it is related to nodes A, B, D and E.

In order to create this matrix in UCINET program, click on Data>Data Editors>Excel Matrix Editor and in the editor opened by clicking on Data>Data Editors>Excel Matrix Editor, the values seen above should be entered in the rows, columns and cells and saved as Ucinet files (*.##h) as foldername. To create the social network in UCINET program using this matrix, click on Visualize>NetDraw menu and select the previously saved foldername.##h file by clicking the Open button in the NetDraw window. Below is the graph of the network formed according to the connections in the given matrix.

Figure 3
Graph Structure for the Sociomatrix is Given in Table 8



References

Aboelela, S. W., Merrill, J. A., Carley, K. M., & Larson, E. (2007). Social network analysis to evaluate an interdisciplinary research center. *Journal of research administration, 38*(1), 61-75.

Acar, N (2024). Social network analysis in new media research: "Concepts, criteria, approaches". *Journal of Communication Science Research, 4*(2), 167-179. https://doi.org/10.5281/zenodo.10985614

Bloch, F., Jackson, M.O. & Tebaldi, P (2023). Centrality measures in networks. *Social Choice and Welfare, 61*, 413–453. https://doi.org/10.1007/s00355-023-01456-4

Borgatti, S. P., Mehra, A., Brass, D. J., & Labianca, G. (2009). Network analysis in the social sciences. *Science, 323*(5916), 892-895. DOI: 10.1126/science.1165821

Cartwright, D., & Harary, F. (1956). Structural balance: a generalization of Heider's theory. *Psychological review, 63*(5), 277. https://doi.org/10.1037/h0046049

Cheng, J., Romero, D. M., Meeder, B., & Kleinberg, J. (2011, October). *Predicting reciprocity in social networks*. In 2011 IEEE Third International Conference on Privacy, Security, Risk and Trust and 2011 IEEE Third International Conference on Social Computing (pp. 49-56). IEEE.

Codal, K. S., & Coşkun, E. (2016). A network analysis related to the comparison of social network types. *Abant Social Sciences Journal, 16*(1), 143-158. https://doi.org/10.11616/basbed.vi.455794

Coviello, N. E. (2005). Integrating qualitative and quantitative techniques in network analysis. *Qualitative Market Research, 8*(1), 39-60. https://doi.org/10.1108/13522750510575435

DeJordy, R., & Halgin, D. (2008). *Introduction to ego network analysis*. Boston MA: Boston College and the Winston Center for Leadership & Ethics.

Edwards, G. (2010). *Mixed-method approaches to social network analysis.* Discussion Paper. NCRM. Retrived from https://eprints.ncrm.ac.uk/id/eprint/842/

Gençer, M. (2023). *Applied social network analysis*. İzmir Economy University.

Granovetter, M. (1983). The strength of weak ties: A network theory revisited. *Sociological theory, Vol 1*, 201-233. https://doi.org/10.2307/202051

Laghridat, C., & Essalih, M. (2023). A set of measures of centrality by level for social network analysis. *Procedia computer science, 219*, 751-758. https://doi.org/10.1016/j.procs.2023.01.348

Lewis, J. M. (2011). The future of network governance research: Strength in diversity and synthesis. *Public Administration, 89*(4), 1221-1234. https://doi.org/10.1111/j.1467-9299.2010.01876.x

Majeed, S., Uzair, M., Qamar, U., & Farooq, A. (2020*). Social Network Analysis Visualization Tools: A Comparative Review.* Proceedings-2020 23rd IEEE International Multi-Topic Conference, INMIC 2020.

Marin, A. and Wellman, B. (2011) Social Network Analysis: An Introduction. In Scott, J. and Carrington, P.J., Eds.(pp. 11-25), *The Sage Handbook of Social Network Analysis*. Sage Publications, Thousand Oaks.

Moreno, J. L. (1937). Sociometry in relation to other social sciences. *Sociometry, 1*(1/2), 206-219. https://doi.org/10.2307/2785266

Oliveira, M., & Gama, J. (2012). An overview of social network analysis. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 2*(2), 99-115. https://doi.org/10.1002/widm.1048

Scoot, J. (1992). *Social network analysis theory and application*. Newberry Park CA: Sage.

Somyürek, S. & Güyer, T. (2020). Social network analysis. In Güyer, T., Yurdugül, H., & Yıldırım, S. (Eds*), Educational Data Mining and Learning Analytics (*pp.329-375). Nobel.

Streeter, C. L., & Gillespie, D. F. (1992). Social network analysis. *Journal of Social Service Research, 16*(1-2), 201-222. DOI: 10.12691/ajap-1-2-2

Tunalı, V. (2016). *Introduction to social network analysis*. Nobel Publishing.

Tabassum, S., Pereira, F. S., Fernandes, S., & Gama, J. (2018). Social network analysis: An overview. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 8*(5), e1256. DOI: 10.1002/widm.1256

Wasserman, S. & Faust, K. (2004). *Social network analysis: Methods and applications*. University of Cambridge University Press.

Wu, Y., & Duan, Z. (2015). Social network analysis of international scientific collaboration on psychiatry research. *International Journal of Mental Health Systems, 9*, 1-10.

Zamanitajeddin, N., Jahanifar, M., Bilal, M., Eastwood, M., & Rajpoot, N. (2024). Social network analysis of cell networks improves deep learning for prediction of molecular pathways and key mutations in colorectal cancer. *Medical Image Analysis, 93*, 103071. https://doi.org/10.1016/j.media.2023.103071

About the Author

Akça Okan YÜKSEL received a Ph.D. degree from the Department of Computer Education and Instructional Technologies at Gazi University, one of the most prestigious universities in Türkiye. He works as a Lecturer Doctor in Information System Coordination at Middle East Technical University. His research interests include Social Network Analysis, Learning Analytics, Educational Robotics Applications, Electronic Portfolios, and Learning Management Systems.
E-mail: akca@metu.edu.tr , **ORCID:** 0000-0002-5430-0821

Similarity Index

The similarity index obtained from the plagiarism software for this book chapter is 17%.