ISSN 2227-524X

International Journal of Engineering and Technology (IJET)

Volume No. 17
Issue No. 2
May - August 2025

ENRICHED PUBLICATIONS PVT. LTD

JE-18, Gupta Colony, Khirki, Extn, Malviya Nagar, New Delhi-110017
PHONE : - + 91-8877340707
E-Mail : info@enrichedpublications.com

ISSN 2227-524X

International Journal of Engineering and Technology (IJET)

Aim & Scope

International Journal of Engineering and Technology (IJET) is a scholarly open access, peer-reviewed, interdisciplinary, quarterly and fully refereed journal focusing on theories, methods and applications in Engineering and Technology. IJET covers all areas of Engineering and Technology, publishing refereed original research articles and technical notes. IJET reviews papers approximately two to four months of submission date and publishes accepted articles on the forthcoming issue upon receiving the final versions.

previously unpublished research results, experimental or theoretical, and will be peer-reviewed. Articles submitted to the journal should meet these criteria and must not be under consideration for publication elsewhere. Manuscripts should follow the style of the journal and are subject to both review and editing.

All the papers in the journal are also available freely with online full-text content and permanent worldwide web link. The abstracts will be indexed and available at major academic databases.

IJET welcomes author submission of papers concerning any branch of the Engineering and Technology and their applications in business, industry and other subjects. The subjects covered by the journal includes but not limited to:

Aeronautical/ Aerospace Engineering, Agricultural Engineering, Automation Engineering, Automobile Engineering, Ceramic Engineering, Chemical Engineering, Civil Engineering, Collaborative Engineering, Communication Engineering, Complexity in Applied Science and Engineering, Computational Science and Engineering, Computer Aided Engineering and Technology, Computer Applications in Technology, Computer Engineering, Continuing Engineering Education and Life-Long Learning, Control Theory, Data Mining and Bioinformatics, Design Engineering, Electrical Engineering, Electronic and Electrical Engineering, Embedded Systems, Engineering Management and Economics, Environmental Engineering, Forensic Engineering, Industrial Engineering, Information Systems and Management, Information Technology, Instrumentation Engineering, Intelligent Engineering Informatics, Knowledge Engineering and Data Mining, Manufacturing Engineering, Marine Engineering, Materials Engineering, Mechanical Engineering, Mechatronics, Metallurgical Engineering, Microengineering, Mining Engineering, Nuclear Engineering, Petroleum Engineering, Process Systems Engineering, Production Engineering, Remote Monitoring, Sensor Network, Soft-computing and Engineering Education, Software Engineering, Strategic Engineering Asset Management, Structural Engineering, Textile Engineering, Transportation.

ISSN 2227-524X

IJET Editorial Board

ISSN 2227-524X

ISSN 2227-524X

ISSN 2227-524X

ISSN 2227-524X

International Journal of Engineering and Technology (IJET)

(Volume No. 17, Issue No. 2, May - August 2025)

Contents

ISSN 2227-524X

Visual Modeling: "Unlocking Ideas and Enhancing Understanding: The Power of Visual Modeling"

Saloni Kumari1*
1Software Engineer II at EY (Ernst & Young), Hyderabad, India

ABSTRACT

To make ideas, concepts, or systems easier to comprehend and explain, visual modelling is the act of putting them into visual form. A range of methods and tools, including mind maps, flowcharts, diagrams, and wireframes, can be used to do this. In a range of settings, including business, education, and software development, visual models can be used to organize and clarify information, uncover links and patterns, and improve communication and teamwork.
This paper discusses the SWOT analysis, the concept of use cases, the organization chart and its tiers, the stakeholder map, the scoring matrix, process flowcharts, and user stories.

Keywords: *Visual Modeling, SWOT Analysis, Use Cases, Organization Chart, Tiers, Stakeholder Map, Scoring Matrix, Process Flowcharts, User Stories, Strategic Planning, Decision-Making, Stakeholder Management, Process Optimization, User-Centric Design, Project Management.*

1. Introduction

There are numerous forms of visual models that can be employed for various tasks. One type of diagram that demonstrates the steps in a process or the flow of information is a flowchart. A mind map is a diagram that organizes concepts around a central notion, with branches designating ideas that are related to the center concept. The layout and functionality of a website or application are planned and designed using wireframes, which are low-fidelity models of a user interface.

To efficiently organize and comprehend complicated information as well as convey ideas to others, visual models can be created. Due to its ability to visually examine and evaluate various possibilities and scenarios, it can also be a useful tool for problem-solving.

2. Research Methodology

2.1. Swot Analysis

Evaluation of a company's or organization's SWOT—strengths, weaknesses, opportunities, and threats—is done using the SWOT analysis, a strategic planning technique. To design strategies to maximize opportunities and minimize threats, a SWOT analysis aims to identify the internal and external elements that can affect the success of the company or organization.

You must first identify the internal elements that are important to your company or organization to perform a SWOT analysis. These can include both your assets (such as a powerful brand, a skilled workforce, or an original product) and liabilities (such as limited financial resources, outdated technology, or poor customer service). The external influences that can affect your company or organization must therefore be considered. Opportunities (such as new market trends, alliances, or developing technology) and threats (such as competition, regulatory changes, or economic downturns) can be among them.

To develop a SWOT matrix, which is a graphic depiction of the strengths, weaknesses, opportunities, and threats, you first need to identi fy the internal and external elements. This can assist you in developing plans to take advantage of opportunities and counter threats, as well as in analyzing the strengths and weaknesses of your company or organization with respect to the opportunities and risks it confronts.

SWOT analysis is frequently used in organizational and company planning, but it may also be helpful in preparing for one's own life and career, as well as for identifying and solving issues in several situations.

2.1.1. SWOT Analysis Breakdown – S

The letter "S" stands for strengths in a SWOT analysis. Strengths are the inherent characteristics that set your company or organisation apart from rivals. A strong brand, an original product or service, a skilled staff, ample financial resources, or a prime location are a few examples of these. The SWOT analysis approach includes identifying your strengths since it enables you to understand what your company or organisation does well and where you have a competitive edge. By concentrating on your strengths, you may expand upon them and leverage them in your market. The following are some instances of strengths that could be found in a SWOT analysis: Strong customer relationships, strong brand recognition or reputation, high-quality goods or services, a talented and experienced team, proprietary technology, a strategic location or access to vital markets, and patents or other intellectual property are all desirable qualities. When performing a SWOT analysis, it's critical to be sincere and unbiased when defining your strengths. Recognize your skills without hesitation but be mindful of any prejudices or blind spots that can hinder you from seeing them properly.

2.1.2. SWOT Analysis Breakdown – W

The letter "W" in a SWOT analysis stands for weaknesses. The internal variables known as weaknesses prevent your company or organization from operating effectively or efficiently. These are places where your company or organization has an advantage over rivals. Finding your vulnerabilities is a crucial step in the SWOT analysis process since it clarifies your areas for improvement and shows you how to deal with any shortcomings that might be preventing you from moving forward. You may improve your competitiveness and position yourself to succeed in your market by being aware of and taking steps to solve your vulnerabilities. In a SWOT analysis, weaknesses like these might be discovered: Low financial resources; outdated technology or equipment; outdated employees; little brand awareness or market penetration; the absence of a distinctive or original product or service; subpar customer service or support; and inadequate processes or systems When performing a SWOT analysis, it's critical to be sincere and unbiased when assessing your weaknesses. Recognize your areas of weakness without being ashamed to admit them but be mindful of any prejudices or blind spots that can hinder you from recognizing your limitations clearly.

2.1.3. SWOT Analysis Breakdown – O

The "O" stands for opportunities in a SWOT analysis. Opportunities are outside forces that may be advantageous to your company or group. You can take advantage of these circumstances or trends to advance your objectives or strengthen your position in the market. The SWOT analysis approach includes identifying opportunities since it aids in your understanding of your market and the areas where

you might profit from trends or developments. Taking advantage of opportunities can help you increase your chances of success and expand your business or organization.

The following are some examples of opportunities that could be found in a SWOT analysis: new market trends or customer needs; changes in technology or the competitive environment; partnerships or collaborations with other companies or organizations; Changes in regulations or policies that open up new opportunities; growth into new markets or areas; Modifications in customer behavior or desires It's crucial to be proactive and search for opportunities that can help your company or group. Be open to new concepts and keep an eye out for trends or changes that can present you with opportunities. Consider opportunities that might be outside of your present focus or area of expertise without being scared to think outside the box.

2.1.4. SWOT Analysis Breakdown – T

The "T" in a SWOT analysis stands for threats. Threats are outside forces that have the ability to hurt your company or organization. To lessen their effects or completely avoid them, you need to be aware of certain situations or patterns and be ready for them. Finding threats is a crucial step in the SWOT analysis process because it enables you to comprehend the dangers and problems that may exist in your market and to create plans to reduce or eliminate them. You can lessen your susceptibility and improve your chances of success by dealing with threats. A SWOT analysis may identify threats like intense competition; modifications to laws or policies that introduce new difficulties; changes in customer behavior or preferences; economic downturns or market instability; disruptive technology or new market entrants; and political or social unrest. Being proactive and on the lookout for potential risks to your company or organization is crucial. Keep an eye on what is going on in your market and be ready to adjust as necessary to new situations or problems. It is better to be ready than to be caught off guard, so don't disregard prospective risks or believe they won't affect you.

2.2. Introduction to Use Cases

A use case is a description of how a user of a system or piece of software will carry out a certain task. In software development, use cases are frequently used to record the specifications and design of a system or application and to direct the development process. A use case often includes a main flow of events, which outlines the procedures necessary to achieve the objective, as well as any potential alternate flows or exceptions. A use case typically also contains information on the actors (the systems and people that interact with the system) and the prerequisites (the conditions that must be met for the use case to begin). Use cases offer a concise and in-depth description of what the system should do and how it should act in various scenarios, making them ideal for capturing and documenting the requirements of a system or application. Additionally, they are helpful for testing and confirming that the system or application is operating according to plan. Depending on the requirements of the project, several levels of information might be included in use cases. A detailed use case could outline the processes and interactions involved in a particular job or process, whereas a high-level use case might offer a comprehensive overview of the entire system. To build a thorough model of the system or application, use cases are frequently combined with additional modelling approaches like flowcharts, sequence diagrams, and class diagrams.

2.3. Introduction to the Organization Chart

A visual representation of the structure and relationships inside an organisation is called an organisation chart (also known as an organisational chart or org chart). It displays the organisational structure, the duties and responsibilities of each position, the channels of communication, and the relationships between those in charge of reporting. In an organisation chart, the various positions within the company are often represented by boxes or other forms, and the reporting links between the jobs are typically shown by lines or arrows. The highest degree of authority is typically represented by the position at the top of the chart, such as the CEO or president, while the lower levels of authority are typically represented by the vice presidents, managers, and team members. Organization charts can be used for many things, such as: communicating the structure and hierarchy of the organisation; defining roles and responsibilities; Identifying lines of communication and reporting relationships; facilitating decision-making and communication within the organisation; Analysing the organization's structure and identifying areas for improvement Organization charts can be made with several tools, such as spreadsheet programmes, diagramming software, and online resources. Both small and large organisations can benefit from them, and they can be modified as the organisation develops or changes.

2.4. Understanding Org Chart Tiers

The tiers in an organizational chart stand for the various levels of the organizational hierarchy. Depending on the size and structure of the organization, there may be more or fewer tiers and different positions within each tier. The greatest level of power is typically represented by the CEO or president at the top of an organizational chart. Vice presidents and other senior managers may make up the second tier, and managers and team leaders may make up the third tier. The lower echelons may consist of team members or individual contributors. An organizational chart might have a few tiers (such as in a small business with a flat structure) or several tiers (such as in a large corporation with a complex hierarchy). To demonstrate further degrees of hierarchy or specialty within the organization, the tiers can also be further broken down into sub-tiers. Understanding an organization chart's tiers will help you better grasp the functions and responsibilities of each position within the organization, as well as its hierarchy and reporting links. Additionally, it may assist you in locating chances for professional progress as well as in comprehending the organization's decision-making process and communication channels.

2.5. Introduction to Identify and Manage Stakeholders

Every project and commercial effort must identify and manage its stakeholders. Customers, employees, shareholders, suppliers, regulators, and other people or organizations can all be stakeholders if they have a stake in the project's or venture's success or conclusion. Identification of all the stakeholders who will be influenced by the project or enterprise, as well as awareness of their interests, worries, and expectations, are necessary for effective stakeholder management. It also entails creating plans for interacting with stakeholders, handling their issues, and controlling their expectations. You can take the following actions to locate and manage stakeholders: Determine stakeholders: Make a list of all the stakeholders who could be touched by the project or business before it gets started. Customers, staff members, stockholders, suppliers, regulators, and other individuals or groups may fall under this category. Examine the interests and worries of stakeholders: Think about each stakeholder's interests and worries regarding the project or enterprise. What do they anticipate getting out of the initiative, if anything? What do they anticipate? Set stakeholder priorities: Each stakeholder's significance and possible influence on the project or enterprise should be considered. Some stakeholders could be more important than others and call for additional resources. Develop a stakeholder engagement plan: Create

a plan for engaging with stakeholders and meeting their needs based on your research of their interests and concerns. This could include specific activities or projects to address stakeholder concerns as well as communication tactics like routine meetings or updates. Watch and evaluate: To make sure that it is effective and that stakeholders are being managed properly, periodically review and evaluate your stakeholder engagement plan. As the project or venture develops, make any necessary revisions.

2.6. The Stakeholder Map and its Purpose

Stakeholder analysis and prioritization in a project or commercial endeavor are done using a stakeholder map, which is a visual tool. A stakeholder map's goal is to assist project managers or corporate leaders in comprehending the interests, worries, and expectations of stakeholders and in formulating plans for productive interactions with them. A stakeholder map typically consists of a grid or matrix that plots stakeholders along two axes: the degree to which they are interested in the project or enterprise, and the degree to which they have control over it. The stakeholder positions are then plotted on the map, with high-interest, high-influence stakeholders in the top right quadrant and low-interest, low-influence stakeholders in the bottom left quadrant, based on their positions on these two axes. The following are some advantages of adopting a stakeholder map: It assists you in understanding the interests and concerns of all the stakeholders who will be influenced by the project or enterprise. It enables you to prioritize stakeholders and devote resources and attention in accordance with their level of interest and impact. The stakeholder group's dynamics and possible conflicts can be better understood by using the visual representation of stakeholder relationships that is provided by this method. It assists you in creating plans for interacting with stakeholders and controlling their expectations. To develop a stakeholder map, you must gather details about the stakeholders, such as their expectations, amount of influence and power, and interests and concerns. The important stakeholders may then be identified, and a strategy for engaging with them can be developed using this information, which you can then plot on the map.

2.7. Scoring Matrix Basics

A score matrix is an instrument for assessing and contrasting options or alternatives according to a set of criteria. It is frequently used in decision-making procedures to assist in selecting the best choice from a list of options. A score matrix typically consists of a table with columns reflecting the criteria and rows representing the options being evaluated. Depending on how well an option satisfies a certain requirement, it is assigned a score for that criterion. To compare the possibilities and choose the best one, the scores are then added up to provide an overall score for each choice. There are several steps to using a scoring matrix: Identify the options: Choose the alternatives that you want to compare and analyse. These alternatives could take the form of goods, services, plans, or other things. Identify the criteria: Establish the criteria you'll use to compare your selections. This ought to be precise, quantifiable, and pertinent to the choice you're trying to make. Assign weights to the criteria: Determine the importance of each criterion in relation to the others. To indicate each criterion's relative weight, you can give it a percentage or a number. Score the options: Assign a score to each choice for each criterion depending on how well it satisfies that criterion. The results may be based on a set of specified values or a scale (such as 1–5 or 1–10). (Such as "high," "medium," or "low"). Calculate the total scores: Determine the overall score of each choice by adding the points for each one. If you have given the criteria weights, you can also weight the scores according to the significance of each criterion. Compare the options: Compare the choices and choose the best one using the overall scores. You might also want to consider any other

elements or aspects that the scoring matrix missed. For organised, unbiased comparison and evaluation of possibilities, scoring matrices can be an effective tool. But it's crucial to remember that they are only one tool and ought to be used with other methods of making decisions.

2.8. Process Flowcharts

A flowchart is an illustration of the steps in a process or workflow. It is employed to symbolise the flow of data, items, or jobs as they pass through several procedures or phases. A typical process flowchart is made up of several boxes or shapes that stand in for the various steps in the process and arrows that depict how the process moves from one step to the next. The flowchart's boxes and shapes each represent a distinct stage or activity and may include information such as input, output, and any decision points. Using a process flowchart has numerous advantages, including: Clarifying and describing the stages of a process; Finding process bottlenecks or inefficiencies; communicating the process to others; finding chances for improvement or automation; Facilitating decision-making and problem-solving A range of settings, including business, manufacturing, healthcare, and software development, can benefit from the use of process flowcharts. They can be made with a variety of tools, including internet tools, spreadsheet programmes, and diagramming software. To design a process flowchart, you must first specify the goals and parameters of the process, then list and map out the many steps. To make sure that the flowchart accurately depicts the process, you might also need to solicit input and feedback from stakeholders. Once the flowchart is finished, you can use it to convey the process to others as well as to analyse and improve it.

2.9. User Stories

An account of a feature or functionality from the viewpoint of the user is known as a "user story." In agile software development, requirements for a system or product are gathered and used to guide the development process. User stories frequently have a straightforward structure, like this: As [kind of user], I desire [some goal] so that [some reason]. To focus on the wants and objectives of the user, user stories are used to record the requirements for a system or product. Typically created by the product owner or business analyst, they serve as a roadmap for the development team as they construct the product or system. The goal of user stories is to be finished in a single sprint or iteration; hence, they are frequently brief and narrowly focused. They are often prepared in straightforward language with the intention of being comprehended by both technical and non-technical stakeholders. User stories are an essential component of the agile development process because they ensure that the system or product being developed meets user needs and adds value. Additionally, they are used to prioritise development tasks and monitor advancement in relation to the overall product roadmap.

3. Results and Discussion

Visual modeling is a potent tool for arranging and communicating complicated information, assisting in problem-solving, and promoting efficient communication across a variety of areas, including business, education, and software development. SWOT analysis, use cases, organization charts, stakeholder mapping, scoring matrices, process flowcharts, and user stories are just a few of the visual modeling concepts and approaches that have been briefly discussed in this paper. We will talk about the importance and implications of these visual modeling tools in this part.

SWOT Analysis:

• The SWOT analysis is a crucial tool for strategic planning in both professional and private settings. It enables people and organizations to recognize their advantages, risks, weaknesses, and opportunities.
• Businesses may take advantage of opportunities, strengthen their weaknesses, manage threats, and capitalize on strengths by performing a SWOT analysis.
• The SWOT analysis is a flexible method that may be used for a variety of purposes, including corporate planning and individual career development.

Use Cases:

• Use cases are crucial for capturing and describing a system or application's needs throughout software development. They give a thorough explanation of the system's user interface.
• Use cases guarantee that the software development process is in line with user needs, resulting in products that are more user friendly and efficient.
• These models are especially useful in Agile approaches since they direct task-based development and prioritize features based on user stories.

Organization Charts and Tiers:

• Organizational charts are useful tools for displaying the relationships and hierarchical structure inside an organization.
• Understanding organizational levels is essential for understanding each position's functions and responsibilities, decision making procedures, and communication routes.
• Stakeholders can evaluate the reporting structure, spot areas for prospective adjustments, and spot opportunities for career progression by visualizing an organization's hierarchy.

Stakeholder Mapping:

• Project success depends on the efficient identification and management of stakeholders.
• Stakeholder maps give a clear visual picture of the influences and areas of interest of various stakeholders, assisting in prioritization and engagement strategies.
• Project managers and leaders can use this tool to proactively address issues and expectations, reducing conflicts and fostering teamwork.

Scoring Matrices:

• When several options need to be assessed against a set of criteria throughout the decision-making process, scoring matrices are helpful.
• They provide a methodical and objective means of evaluating options and choosing the best one.
• Even though scoring matrices offer a systematic approach, they should be utilized in conjunction with other decision-making techniques for a more thorough study.

Process Flowcharts:

• For visualizing and optimizing workflows in a variety of disciplines, process flowcharts are crucial.

Process Flowcharts:

• For visualizing and optimizing workflows in a variety of disciplines, process flowcharts are crucial.
• They support the discovery of obstacles, inefficiencies, and areas for improvement.
• Process flowcharts help with clear process communication and serve as a foundation for decision-making and problem-solving.

User Stories:

• Agile software development is not complete without user stories, which make sure that user needs, and expectations are taken into account.
• They assist with feature prioritization, sprint planning, and progress monitoring.
• By giving all stakeholders a shared knowledge of the project's objectives, user stories improve collaboration between technical and non-technical parties.

4. Conclusion

In a variety of industries, visual modeling is useful for both individuals and organizations. SWOT analysis, use cases, organization charts, stakeholder mapping, scoring matrices, process flowcharts, and user stories are just a few of the essential visual modeling techniques and
concepts that have been covered in this paper. With the help of these technologies, we are better able to communicate effectively, make educated judgments, and clarify difficult information.
Visual modeling offers several key takeaways:

Strategic Planning: Businesses and people can identify strengths, weaknesses, opportunities, and threats with the aid of tools like SWOT analysis, which offers a structured approach to strategic planning. This information serves as the basis for making wise decisions and employing powerful techniques.
Effective Software Development: To ensure that software satisfies user needs and expectations, use cases and user stories are crucial to software development. They support agile development, enabling incremental enhancements and user-centered products.
Organizational Clarity: Transparency, effective communication, and effective decision-making are made possible by organizational charts and a thorough understanding of organizational layers. These resources assist people in navigating their professional lives and assist businesses in structuring their operations.
Stakeholder Engagement: Organizations may identify, prioritize, and effectively engage with stakeholders by using stakeholder map ping. Collaboration is encouraged, disagreements are reduced, and project success is ensured.
Decision-Making Support: To make objective decisions, scoring matrices provide a methodical technique to compare solutions to predetermined criteria. They improve the ability to make decisions when combined with other strategies.
Process Optimization: Process flowcharts are essential for illustrating workflows, locating bottlenecks, and simplifying procedures.
They open the door for increased effectiveness and output.
User-Centric Development: Prioritizing user needs during product development results in solutions that are both user-friendly and effective. They promote cooperation amongst different stakeholders.

We can comprehend complicated systems more fully, express concepts clearly, and arrive at sound conclusions when we apply visual modeling tools to our work. Visual modeling helps us to manage obstacles and seize opportunities with clarity and assurance, whether it is used in business, education, or our personal lives. Visual modeling serves as a guiding light, assisting us in plotting our way toward success in a world where information is plentiful and complexity is the norm. We can more effectively innovate, attain our goals with more precision, and simplify the complex by utilizing the power of these technologies.

Acknowledgement

I would like to express our sincere gratitude to my organization EY (Ernst & Young) for unwavering guidance, invaluable insights, and constant encouragement throughout this journey. Their valuable input and feedback significantly improved the quality of this research. I would like to acknowledge the contributions of my research colleagues and friends who provided valuable feedback, engaging discussions, and constructive criticism. Their diverse perspectives enriched this work significantly.

References

[1] Beyer, H., & Holtzblatt, K. (1999). Contextual Design: Defining Customer-Centered Systems. Morgan Kaufmann.
[2] Jacobson, I., Christerson, M., Jonsson, P., & Övergaard, G. (1992). Object-Oriented Software Engineering: A Use Case Driven Approach. Addison-Wesley.
[3] Rosenberg, D., & Scott, H. (2007). Use Case Driven Object Modeling with UML: A Practical Approach. Apress.
[4] Ambler, S. W. (2004). Introduction to UML 2 Activity Diagrams.
[5] Freeman, R. E., & Reed, D. L. (1983). Stockholders and Stakeholders: A New Perspective on Corporate Governance. California Management Review, 25(3), 88-106.
[6] Axelrod, C. W. (2013). Organizational Structure: Six Key Elements of an Effective Organization. CreateSpace Independent Publishing Platform.
[7] Dixon, N. M., & Adamson, L. (2011). The Power of the 2x2 Matrix: Using 2x2 Thinking to Solve Business Problems and Make Better Decisions. Jossey-Bass.
[8] Linton, I. D. (2020). Flowchart Symbols and Meaning: A Comprehensive Guide. [Online Article] Available at: https://creately.com/blog/diagrams/flowchart-symbols-and-meaning/
[9] Cohn, M. (2004). User Stories Applied: For Agile Software Development. Addison-Wesley.

ISSN 2227-524X

Data integration: “Seamless data harmony: The art and science of effective data integration”

Saloni Kumari *
Software Engineer II at EY (Ernst & Young), Hyderabad, India

ABSTRACT

The idea of data integration has evolved as a key strategy in today's data-driven environment, as data is supplied from various and hetero geneous sources. This article explores the relevance, methodology, difficulties, and transformative possibilities of data integration, delving into its multidimensional world. Data integration serves as the cornerstone for well-informed decision-making by connecting heterogeneous datasets and fostering unified insights. This article gives readers a sneak preview of the in-depth investigation into data integration, illuminating its technical complexities and strategic ramifications for companies and organizations looking to maximize the value of their data as-sets.

Keywords: *Data Analytics; Data Processing; Data Storage; NoSQL Database; Distributed Computing; Scalability; Fault Tolerance; Data Warehousing; Data Ingestion; Workflow Scheduler; Coordination Service; Big Data Architecture; Hadoop Ecosystem.*

1. Introduction

It is obvious that our society is driven by data, and as we steadily move toward fully digitalized lives, data is becoming a valuable resource for the contemporary economy. We create important data whenever we use the internet to make purchases, view content, or share it on social media. Many of the biggest internet companies increasingly rely on data-driven business models to run their operations. However, without data integration, none of it is possible. Data integration is the glue that allows raw data to be transformed into an asset. Lack of data integration results in a host of business issues. Due to fragmented data silos between organizations or departments within companies, it becomes necessary for users to rekey data or duplicate their efforts. Making decisions may be difficult in the absence of uniform data views. When individuals or departments only have partial access to data, they frequently make judgments that don't consider the overall process and are therefore suboptimal. Businesses waste a lot of money due to inefficiencies and bad decisions that result from poor data integration. Techniques for data integration aid in minimizing these issues. The process of obtaining data from many sources and transforming it into a data store or business application so that it can be used more efficiently is known as data integration. Three different kinds of data integration exist: Business-to-business integration entails cross-organizational linkages to improve the efficiency of business transactions between trading partners. The goal of application integration is to connect different corporate applications to create an integrated process. The data store itself is the level at which database integration takes place. Building pipelines to transfer raw data between data stores falls under this category. When developing data warehouses and business intelligence systems, this kind of integration is used. In the contemporary workplace, all three forms of integration are regularly used and are very beneficial to comprehend. In the current business environment, businesses that excel at data integration will have a significant competitive advantage.

Let's imagine a producer of agricultural equipment wants to use the information gathered by its tractors to improve crop yields. Perhaps sensors on the tractors can assess the soil moisture. The producer may gather this information from all of their tractors and integrate it with information from other sources,

such as weather or commodity market prices. The ultimate result may be advice for farmers on how to improve the efficiency of their irrigation systems to maximize harvests, or the data could even be sold to outside parties like hedge funds to assist them in making smarter investment choices. This may develop into a new revenue stream for the corporation that would eventually complement or even outperform its current business strategy. That is an illustration of how digitization may be used to transform industries, and it all hinges on data fusion.

2. Research methodology

2.1. Business integration

Business to business, or B2B, integration permits the electronic interchange of commercial transactions between two or more trading partners, such as orders or payments. To accommodate certain scenarios, B2B messaging occasionally needs the extra flexibility provided by XML or API standards. As businesses rely more and more on alliances or intricate supply chains to hasten entrance into new markets and boost competitive advantages without B2B messaging, B2B integration is becoming more and more crucial. The players in the supply chain ultimately communicate manually via email or by sharing Excel attachments. B2B enables unconnected businesses to link their separate business systems into a cohesive workflow. A customer could send a B2B communication, including a purchase order, to the supplier's production system from their ERP system. A purchase order acknowledgement may be sent to the customer automatically by the supplier system after checking production schedules. Trading partners can handle large numbers of transactions with less labor-intensive manual labor and less error-prone automation. Since it must handle the extra challenges of delivering data across corporate boundaries, B2B integration differs from application-toapplication integration and intra-company database connections. B2B integrations must check trading partner communications for compliance with CDI requirements, acknowledge trade partners, monitor the progress of messages, and transmit data securely.

2.2. Application integration

Any type of software used to carry out tasks is referred to as an application. This includes corporate applications like CRM or ERP, iPhone or Android mobile apps, and cloud services like MailChimp or Google Analytics. Due to how essential software has become to our everyday lives, it is usually required to use many applications to execute activities. These programmes are connected by application integration, which creates an efficient workflow.

2.3. Database integration

Simply merging data from numerous sources into one unified perspective to produce insights might be referred to as database integration. It gathers information from many sources and changes it into something more worthwhile and beneficial. Most database integration strategies fit into one of these groups. Data from different storage locations is combined into one data repository through data consolidation. ETL, or extract, transform, and load, is a typical strategy for data consolidation. A specific type of data consolidation called data warehousing gathers data from numerous systems and merges it into a single storage engine that is intended to allow analytical queries. Moving data from one place to another is known as data propagation, and replication is a frequent type of data propagation. Replication

tools that can automatically sync data from an origin source to a destination source are available in the majority of relational databases. This is frequently used to help with disaster recovery or boost data access performance. In contrast to data consolidation, which transfers data into a single data source, data virtualization offers a single view of data across multiple data sources. Users can operate with a façade that is created through data virtualization. Behind the scenes, it retrieves data from the many data sources. In contrast to data virtualization, data federation enforces a single data model across all of the diverse data sources. Database workloads are frequently divided into one of two categories: LTP or OLAP. A database that handles common corporate transactions, such as an ERP or CRM system, is referred to as old TPE, or online transaction processing. A mix of database reads and writes involving recent business transactions, such as orders or leads, is typical of old-style workloads. Online analytical processing, or OLAP, is primarily concerned with reporting and analytics. OLAP tasks entail database reads across big data sets, such as queries on the daily average of orders over the last 12 months. Data warehouses are made to serve these OLAP workloads. For instance, getting data from the billing system would require a very timeconsuming query if we needed a report that compared the total revenue for this year to the previous year. It might be necessary to add up millions of bills from the previous two years. Or perhaps there are different billing systems in use in various locations, and the revenue figures need to be added together to provide a whole picture. All of these data sources would be combined into a database with a data warehouse, allowing for rapid queries on millions of entries. These queries can be executed far more quickly than source systems using certain technologies. They sometimes combine the outcomes. This implies that the billions of billing transaction rows are condensed in a batch process, and the annual revenue total is saved in the database as a single figure for easy retrieval. Star schemas, which describe the arrangement of tables and columns in the database, are frequently used by data warehouses for their data models. In a star schema, data is kept in two types of tables: facts and dimensions. Fact tables keep track of important financial data like income. This design results in a large number of rows for the fact table and a relatively smaller number of rows for the dimension tables, which makes for simpler queries, query performance gains, and faster aggregations, operations where you sum up rows, like in our example, where we needed the total revenue for a year. This design results in many rows for the fact table and a relatively smaller number of rows for the dimension tables. Data warehouses are frequently replaced by technologies like Hadoop or Spark. By splitting data over clusters of servers rather than transforming and loading data into a data warehouse, Hadoop enables rapid queries on extremely big datasets.

2.4. ETL extract, transform, load

ETL, which stands for Extract, Transform, and Load, is the procedure for loading a data warehouse and entails polling data from the source systems and putting it into a staging area. To transform data is to prepare it ready for loading into the target system. And loading refers to putting data into the intended system. The process of extracting involves taking data from multiple source systems and putting it into processing staging regions. On-site source systems are an example of a source system. ERP programmes such as SAP, cloud applications such as Salesforce, CSV files, or SQL databases. The business's operating data is contained in these source systems. For huge data sets, care must be taken not to negatively affect the performance of the source system. The extract process will read data from these systems using a variety of ways and write the data into a file system or database for the following stage in the processing pipeline. Typically, data is retrieved from these extracts in its original format and promptly stored in new staging storage in that format before any transformations are applied. This lowers the amount of computer power needed to extract the data from the source system. Although the

full data set may occasionally be recovered in one batch from the source system, it is typically preferable to use a change data capture process in the source system. to handle newly added, updated, or removed records. Data must be transformed before being put into the target system. Data transformations can take a variety of forms. First,data purification is required for this type of transformation to prevent the loading of faulty data into the target system. Eliminating faulty records, getting rid of duplicates, or resolving formatting issues are common cleansing chores. Enrichment of data is frequently essential to enrich the source data before it is loaded. This entails adding information to the data that was not included in the source system but is required to be loaded into the target system. For instance, the customer's address might be provided by the source system, but the target system might need the GPS coordinates, latitude, and longitude of the address to geocode the address data and collect GPS coordinates prior to loading data.

Large data sets can be put into the target database after the data has been extracted and converted. Even while it's common to write data into a relational database using SQL statements like insert, update, and delete, loading 300,000 insert statements to load 300,000 records for an ETL task will be slow and use resources on the target system that could have an influence on performance. Many databases have specialized bulk load capabilities that assist in the effective loading of massive data sets. Managing master data and foreign key relationships is one frequent problem. In a database, master data refers to reference information that is used across several tables. A foreign key is a referential integrity database constraint that makes sure a reference value from one table is present in another related table. When loading new orders in an e-mail process, for instance, a customer mentioned on an order must be present in the customer master table. It is important to manage the key correctly since it links the orders customer field to the customer master table. In data warehouses, certain kinds of data links are handled in a specific fashion. In fact, tables linked to the dimensions by a foreign key, master data is frequently kept in dimension tables. Master data changes over time, as well as their connections to business operations, are frequently captured using a technique termed "slowly evolving dimensions."

The most common way to enter data into data warehouses is through ETL. However, the procedure is frequently referred to as LTE, or extract, load, and transform, when working with data lakes. This is because the data is directly fed into the data system after being extracted from the source system. The lake is transformed at query time. The ETL process is normally done during a period of low activity that will not affect business users of the business information commonly held in a data warehouse. Data warehouse ETL processes are typically batch-focused, possibly once a day or once a week. Given that a data warehouse is typically used for operational or financial analyses, this makes sense. Technically speaking, real-time OLAP is far more difficult to construct than batch based typical OLAP systems. Real-time analysis can be done in many ways.

Apache Spark is a popular framework for implementing streaming analytics. Real-time streaming analytics uses Spark streaming to absorb small batches of data and transform them into a searchable data store. Most event-driven data stores, like Apache Kafka, have built-in handlers in tools like Sparke. ETL procedures can be designed and carried out using a wide range of ETL tools, some of which are incorporated into database systems. An example of this is SQL Server Integration Services, which executes a workflow comprising data sources, data targets, and data flow activities. It may be used to connect to a wide variety of databases and data sources despite being strongly integrated with SQL Server. An open source ETL tool called Taloned supports many different types of databases. Open Studio for Windows or Mac can be downloaded for free. It also features a graphic designer and allows

connectivity with SaaS providers, packaged software apps, and data sources like Dropbox. Although Apache NiFi is not explicitly an ETL tool, given its versatility, it generally automates data transfers across systems. It could be used for both database and application integration. The vast array of data sources that NiFi provides processors for includes on-premises databases, big data platforms, and cloud services. The two most well-known cloud providers, Amazon Web Services and Microsoft Azure, both include ETL tools. Its name is AWS Glue from Amazon. The fact that Glue is a cloud-native title tool means that it offers a visual designer that can be used in a Web browser. Python or Scala are two programming languages that can be used to create transformations.

Data formatting capabilities are one of AWS Glue's distinctive advantages. One of the most widely used methods for storing data such as files is the AWB cloud storage service, known as S03. These files can be crawled by AWS Glue and it can create a data catalogue that lists the data that is accessible in the data lake. AWS Glue makes it simple to transfer this data into different data warehousing services on our platform, such as Amazon Redshift or Amazon. Similar cloud ETL software on the Microsoft Azure cloud is called Azure Data Factory (ADF), which similarly offers a web based visual ETL builder. Building EDF ETL processes doesn't require programming, in contrast to AWS Glue. More than 90 data connectors are available through ADF, including sources from all the main cloud service providers, including Amazon and Google. ADF's ability to host SQL Server Integration Services packages makes it possible to run tasks created using SQL Server's standard ETL tool on Azure cloud infrastructure. Data propagation, which is the process of moving data from one place to another, is frequently used to transfer a database's entirety or a specific subset from one location to another. Users at the target site can now access the data more quickly, and the source and target sites may benefit from redundancy as a result. Data propagation and data warehousing are sometimes used in tandem. Even though an organization may have an enterprise data warehouse where a large global data set is stored, this data is frequently propagated to regional data marts where a smaller portion is made available to local business units.

Better response times for regional users and a more pertinent data set that makes business intelligence jobs easier are two benefits of employing data marts. Edge computing is another possibility for data dissemination. Although moving data and computation to the cloud has been the general trend, businesses are discovering an increasing number of use cases that demand computing at local sites like retail storefronts or warehouses. Edge computing can improve the performance and dependability of services at these outlying locations. An application that uses facial recognition to identify workers entering a warehouse is a typical example of edge computing. Data replication is a frequently used tool for carrying out data dissemination. Most database engines, including PostgreSQL, SQL Server, and Oracle, all have replication features built in. This makes setting up replication and transferring data from a source database to a target database quite simple. Replication's functional implementations differ greatly among these databases. The replication solution may, in some situations, be centered on replicating a database to a backup location for disaster recovery. In other situations, the technology aims to transfer a portion of a master database to a readonly copy to facilitate reporting and analytics.

3. Results and discussion

This paper provides a thorough review of the many facets of data integration, including business integration, application integration, data base integration, and the crucial ETL (Extract, Transform, Load) concept. Data integration is a crucial activity in the contemporary digital landscape. Based on the data in the paper, the following main findings and discussion points are listed:

• Importance of data integration:
In today's data-driven economy, the article emphasizes the importance of data integration. Organizations struggle with issues like data silos, redundant work, and ineffective decision-making without data integration.
• Types of data integration:
The paper discusses three distinct types of data integration: business integration, application integration, and database integration. These types are well-defined and essential in a variety of business contexts.
• Business integration:
B2B integration makes it easier for trading partners to communicate and exchange information about transactions. It emphasizes how crucial standards like XML and APIs are for facilitating these interactions.
• Application integration:
By connecting several software programs, application integration is said to build effective processes. The given example shows how this integration can increase productivity and streamline procedures.
• Database Integration:
Data from several sources are combined into one perspective through database integration, increasing its value and usability. The article addresses several database integration techniques, including data consolidation, propagation, virtualization, and federation.
• ETL process:
It involves removing data from source systems, converting it into a format that can be used, and putting it into a target system. Important steps in this process include data cleaning, enrichment, and bulk loading.
• Challenges in data integration:
In data integration procedures, issues with managing master data, foreign key relationships, and the dynamic nature of data are frequent. To ensure data consistency and correctness, these issues must be resolved.
• Technologies and tools:
Many ETL tools and cloud-based options, such as AWS Glue and Azure Data Factory tools, make data integration processes simpler and provide features like data formatting and broad data connectors.
• Real-time data integration:
It is known that real-time data integration is a trickier but more crucial component of data integration. In the study, real-time analytics solutions like Apache Spark are alluded to.
• Data dissemination:
Data replication and propagation are two ways for disseminating data that help move data to several sites for greater accessibility and redundancy.

Data integration is a vital procedure that enables businesses to fully utilize their data. The concept and its many elements are thoroughly discussed in this study, which also provides insightful information on the difficulties, methods, and tools involved in data integration. The capacity to convert and analyze data effectively can result in major competitive advantages in today's data-driven corporate environment, underscoring the crucial role data integration plays in this context.

4. Conclusion

This article concludes by delving deeply into the area of data integration and highlighting the crucial role it plays in our data-driven society. For organizations to succeed in the digital age, data integration acts as the keystone that turns raw data into an asset. Business integration, application integration, and database

integration are the three main types of data integration that are thoroughly examined in the article. This information gives readers a thorough grasp of how these three types of data integration interact to maximize the value of data. The paper emphasizes how crucial data integration is in the modern economy, where data drives innovation, efficiency, and competitive advantage. Without efficient data integration, businesses must contend with fragmented data silos, duplication of effort, and inadequate decision-making procedures. Ineffective data integration can lead to big monetary losses and lost opportunities. In-depth analyses of each sort of data integration are provided, highlighting the various uses and advantages of each. B2B communication, which is the emphasis of business integration, automates procedures and lowers mistake rates by streamlining transactions between businesses. Application integration links many software programs to ensure efficient operation and smooth operations. Database integration gathers data from several sources, making it easily accessible for reporting and analysis. A crucial step in data integration is the ETL (Extract, Transform, Load) process, which is covered in detail in the article. Data warehousing and analytics require ETL because it ensures data quality and gets it ready for analysis. The debate over data lakes and warehouses also emphasizes how the field of data integration is constantly changing. It emphasizes the move toward real-time analytics and the effective use of tools like Hadoop and Spark for handling large information. The essay goes on to detail several ETL tools and cloud-based solutions, giving readers insights into how data integration methods are really put into practice. It emphasizes how crucial it is to pick the appropriate tools to make the integration process simpler and more efficient. Further demonstrating the adaptability of data integration techniques is the discussion on data replication, propagation, and edge computing. These methods enable businesses to move data where it is needed, speed up responses, and increase service dependability, especially in distant contexts.

This article essentially emphasizes the critical role that data integration plays in contemporary corporate processes. For businesses looking to make the most of their data, take wise decisions, and gain a competitive edge in a world that is becoming more and more data-centric, it is a priceless resource. By navigating the complexities of data integration and selecting the appropriate tactics and resources, one can succeed in the digital age and keep data from being an unmanaged resource but rather an asset.

Acknowledgement

I would like to express our sincere gratitude to my organization EY (Ernst & Young) for unwavering guidance, invaluable insights, and constant encouragement throughout this journey. Their valuable input and feedback significantly improved the quality of this research. I would like to acknowledge the contributions of my research colleagues and friends who provided valuable feedback, engaging discussions, and constructive criticism. Their diverse perspectives enriched this work significantly.

References

[1] "Data Integration Blueprint and Modeling: Techniques for a Scalable and Sustainable Architecture" by Anthony David Giordano.
[2] "Data Integration in the Life Sciences: 5th International Workshop, DILS 2008, Evry, France, June 25-27, 2008, Proceedings" edited by Alfonso Valencia and Paolo Romano.
[3] Kimball, R., Ross, M., Thornthwaite, W., Mundy, J., & Becker, B. (2008). The Data Warehouse Lifecycle Toolkit. Wiley.
[4] Inmon, W. H., & Linstedt, D. (2015). Data Architecture: A Primer for the Data Scientist. Morgan

Kaufmann. https://doi.org/10.1016/B978-0-12802044-9.00001-5.
[5] Inmon, W. H., & Hackathorn, R. D. (2007). Using the Data Warehouse. Wiley.
[6] Vassiliadis, P., Simitsis, A., & Georgantas, N. (2002). Conceptual modeling for ETL processes. International Journal of Data Warehousing and Mining, 8(4), 1-23. https://doi.org/10.1145/583890.583893.
[7] Duan, S., & Cercone, N. (2008). Data integration: A theoretical perspective. Journal of Computer Science and Technology, 23(4), 615-626.
[8] Piplai, T., & Piplai, S. (2014). Data integration: A theoretical perspective. International Journal of Advanced Research in Computer Science and Software Engineering, 4(12).

ISSN 2227-524X

Do Trust based Social Recommendation Algorithms Work as Intended?

Chaitanya Krishna Kasaraneni1* and Mahima Agumbe Suresh2
1Egen, Inc. 2San Jose State University

ABSTRACT

Recommender systems are powerful tools that filter and recommend content/information relevant to a given user. Collaborative filtering is the most popular technique used in building recommender systems and it has been successfully incorporated in many applications. These conventional recommendation systems require a minimum number of users, items, and ratings in order to provide effective recommendations. This results in the infamous cold-start problem where the system is not able to produce effective recommendations for new users. Recently, there has been an escalation in the popularity and usage of social networks, which persuades people to share their experiences in the form of reviews and ratings on social media. The components of social media such as the influence of friends, interests, and enjoyment create the opportunities to develop solutions for sparsity and cold start problems of recommendation systems. This paper aims to observe these patterns and analyze three of the existing social recommendation systems, SocialMF, SocialFD, and GraphRec. SocialMF and SocialFD algorithms are based on matrix factorization and distance metric learning respectively whereas GraphRec is an attention based deep learning model. Through extensive experimentation with the datasets that these algorithms were tested on and one new dataset, we compared the results based on evaluation metrics including Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE). To investigate how trust impacts the performance of these models, we evaluated them by modifying the trust and social component. Experimental results show that there is no conclusive evidence that trust propagation plays a major part in these models. Moreover, these models show a slightly improved performance in the absence of trust statements.

Keywords: *Distance Metric Learning; Matrix Factorization; Neural Networks; Recommender Systems; Social Networks; Social Recommendations*

1. Introduction

With the proliferation of Internet usage, the amount of information available over World Wide Web (WWW) has increased enormously. It is becoming increasingly necessary to recommend relevant parts of online information to users based on their preferences. Recommendation systems fill this gap by predicting "ratings" or "preferences" that a user would give to an item [1]. These are used in variety of areas, mostly in playlist recommendations on Spotify, video recommendations on Netflix and YouTube, product recommendation on Amazon and e-bay, or content recommendations on social media such as Facebook and Twitter[1]. In a recommendation system/recommender system (RS), there are a set of users and a set of items, where each user gives ratings to a subset of items available. The task of the RS is to predict the rating r that user u would give to a non-rated item i or to recommend user u with some items based on the ratings that are already given by user to other items.

RS are generally classified into two types: memory-based recommender systems and model-based recommender systems. Memory-based algorithms, also known as collaborative filtering RS, explore the user-item rating matrix and recommend based on the ratings of item I by a set of users whose rating profiles are most similar to that of user u[2]. Model-based approaches learn and only store the parameters of a model. As a result, these algorithms have no need to explore the rating matrix. Model-

based approaches are very fast after the algorithms learn parameters of the model. The performance bottleneck for model-based approaches is the training phase, whereas memory-based approaches have no training phase. However, the prediction is slower as user-item matrix needs to be accessed several times.

In the present day, with the rapid increase in the popularity and usage of social networks, there is a dramatic growth in number of registered users and various products, which also leads to intractable increase of the cold start problem (new users into the system with less past social behavior) and the sparsity of datasets. Collaborative filtering works effectively when users have expressed a minimum number of ratings to have common ratings with other users in the dataset. For relatively new users, the performance suffers due to the cold start problem. In RS, cold start users are users who are either new to the platform or have given only a few ratings. Using similarity-based approaches, it is infeasible to find corresponding similar users since the cold start users only have a few ratings.

The interpersonal relationships, especially the friends' circles in social networks make it possible to solve the cold start and sparsity problem. The richness of social media give us some valuable insights to drive user recommendations, especially for items such as music, movies, news, brands, and travel. Many social network-based models for recommender systems have been developed to refine the performance but only a handful have considered social circles in their respective approaches. This gap motivates the development of an RS that considers the personal interests of users, interpersonal similarity [3] of interests with their friends, and influence of these interpersonal interests. Asocial rating network consists of a social network with ratings expressed by each user to some items apart from creating social relations to other users. A sample social rating network is depicted in Figure 1.

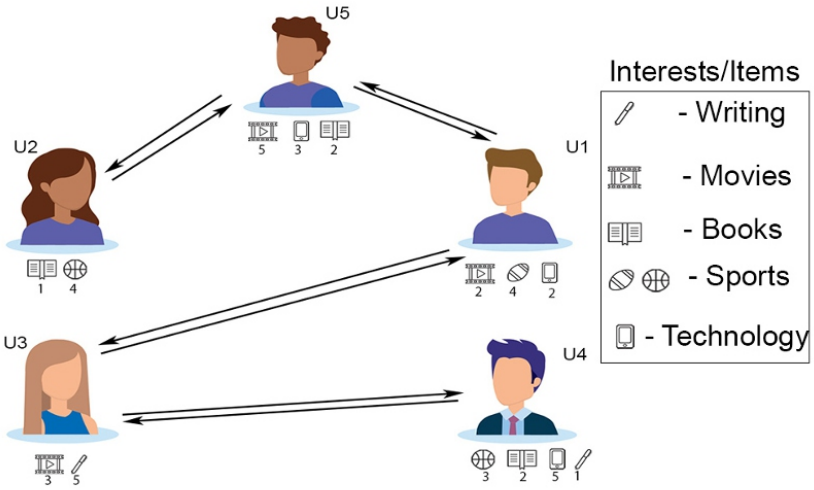

Figure 1: Sample Social Network indicating relations between users and their interests

Table 1 shows the matrix representation of the user-item ratings and Table 2 shows the user-user relationship. Here '1' indicates that user u trusts v. The terms "trust network" and "social network" are used synonymously in this paper.

Table 1: Sample User-Item Rating Matrix

	Sports	Phones	Movies	Writing	Reading
U1	4	2	2		
U2	4				1
U3			3	5	
U4	3	5		1	2
U5		3	5		2

Table 2: Sample Social Trust Matrix

	U1	U2	U3	U4	U5
U1			1		1
U2					1
U3	1			1	
U4			1		
U5	1	1			

Many RS approaches have been proposed using social rating networks [4],[5],[6],[7],[8],[9]. Of them [4],[5],[6],[7] are memory-based methods that explore a social network and find neighborhoods of direct or indirect trusted users and recommend to users by aggregating ratings. These approaches use transitive property to obtain trust to indirect neighbors. These memory-based algorithms are slower compared to model-based approaches in test phase since they have to traverse the entire social network. Model-based RS using social rating networks have been developed in [8][9]. These techniques utilize the matrix factorization to obtain latent features for each user and item from the ratings observed. Experiments show that these model-based approaches perform better compared to state-of-the-art memory-based algorithms. But the major setback is that these algorithms do not take account of trust propagation. To solve this issue, SocialMF [10], a matrix factorization technique based recommendation model was proposed. This model includes trust propagation to improve the quality of recommendations. Optimizing this, another method called SocialFD[11] that incorporates distance metric learning alongside matrix factorization was proposed.

With the recent increase in use of graph neural networks, attention, and deep learning, an attention based deep learning model known as GraphRec [12] was developed. This model contains two components. The first one is to learn user latent factors that contains two separate aggregations, one for learning interactions between users and items in the user-item graph and the other for social aggregation. The second component is extracting item latent factors which contains user aggregation. Finally, model parameters are learned via predictions by integrating both the components. This paper evaluates the performance of SocialMF [10], SocialFD [11] and GraphRec [12], both in presence and absence of social trust. More accurately, the contributions of this paper are:

• Create a new large dataset, extending a prior dataset called TwitterEgo [32], with high quality social circle information extracted from Twitter

• Perform extensive experiments on the SocialMF, SocialFD, and GraphRec algorithms with the datasets that these algorithms were tested on and the new dataset based on TwitterEgo

• Compare the results of these experiments based on the evaluation metrics including Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE)

• Investigate the performance of these models in the following cases:– Whenthere is no trust between any users i.e., the number of trust statements is 0,– Whenthe users are friends only with themselves i.e., users trust only themselves, and– Whenthe users are friends with everyone else excluding themselves The rest of this paper is arranged as follows: Some related works are discussed in section 2. Section 3 summarizes the SocialMF [10], SocialFD [11] and GraphRec [12] models. Experiment methodology and datasets are explored in section 4. Experimental results and comparisons are analyzed in section 5. Finally, the paper is concluded with some directions for future work in section 6.

2. Related Work

This section reviews some of the works in recommendations using social network. Trust propagation is widely considered in memory-based approaches whereas model-based recommendation approaches broadly use matrix factorization [13][14][15]. But the major setback is that these algorithms do not consider social network of users. Model-based recommendation approaches have been developed which utilize matrix factorization technique for recommendations in social networks [8][9], but these approaches do not examine trust propagation. In this section, some model-based works in social networks are discussed after reviewing memory-based models. Using a modified breadth first search technique on the trust network, a memory-based algorithm called TidalTrust [7] was proposed to determine a prediction. TidalTrust tries to find raters with the shortest distance from the user and combines their ratings weighted with trust between the user and these raters [7]. TidalTrust combines the trust value between user u's direct neighbors and v weighted by the trust values of u and its direct neighbors to compute the trust value between user u and v who are indirectly connected [7]. Another approach called MoleTrust which is similar to TidalTrust was introduced in [5]. The major difference is that MoleTrust [5] considers all raters till maximum-depth of the input irrespective of specific user or item. Backward exploration is used in MoleTrust, to compute the trust between users u and v. i.e., the calculated trust value is an aggregation of trust between user u and users who directly trust user v weighted by their direct trust values.

In [16], the authors proposed a maximum flow trust metric called Advogato. This approach helps in discovery of trusted users in an online community. Input for Advogato will be the total number of users to be trusted n. The algorithm needs to understand the whole network structure in order to transform the network to be able to edges of network with capacities. Furthermore, Advogato only calculates the nodes to trust but not different degrees of trust. This technique is not suitable for trust-based recommendations as the trusted users are independent of users and items in the network and the distinction between trusted users is negligible. To consider enough ratings and excluding the noisy data, a random walk approach called TrustWalker was proposed in [4]. This approach combines both item-based and trust-based recommendations. This method not only considers ratings of required item, but also the ratings of similar items. The likelihood of considering these similar items increases with the increase in walk length. Additionally, this framework contains both trust-based and item-based recommendations as special cases. Experiments show that this algorithm outperforms other existing memory-based techniques allowing them to calculate confidence of predictions.

In [8], authors proposed an approach called STE which is a matrix factorization based approach for social network based recommendations. This approach is a sequential combination of basic matrix factorization technique [15] and a social network based technique. Experimental results show that this approach excelled the existing basic matrix factorization based recommendation techniques. However,

the feature vectors of direct neighbors of user u affect the ratings of u instead of affecting the feature vector of u in this model [10]. And also, this model doesn't address trust propagation. Although social network is integrated, real world recommendations are not reflected in this model. Furthermore, this model's interoperability is difficult as two sets of dissimilar feature vectors is considered.

In recent years, there have been many developments in deep neural networks for graph data, especially the social network data [22]. These are known as Graph Neural Networks (GNNs). Works like [23][24][25] have been proposed to learn meaningful insights and representations for graph data. The main idea in these works is to use neural networks for aggregating features from local graph neighborhoods iteratively. Some of these models use graph neural networks. DANSER [17] is one of the most recent algorithms that uses dual graph attention networks to learn representations for two-fold social effects, where one is modeled by a user-specific attention weight and the other is modeled by a dynamic and context-aware attention weight[17]. There are some other social recommender models. Of them, SocialMF [10], SocialFD [11], and GraphRec [12] are focused in this paper.

3. Models used in this paper

Most of the conventional recommender system algorithms do not consider the social relations among the users in a network. With the increasing usage of social networking applications, incorporating this information into recommendation systems has also become increasingly important. We compared the performance of some baseline algorithms, GraphRec[12], RSTE[8], SoRec[9], SoReg[33], Singular value decomposition (SVD)[14] based RS, SocialFD[11], and SocialMF[10]. Figure 2 depicts the performance of these algorithms on standard recommendation datasets in terms of RMSE and MAE.

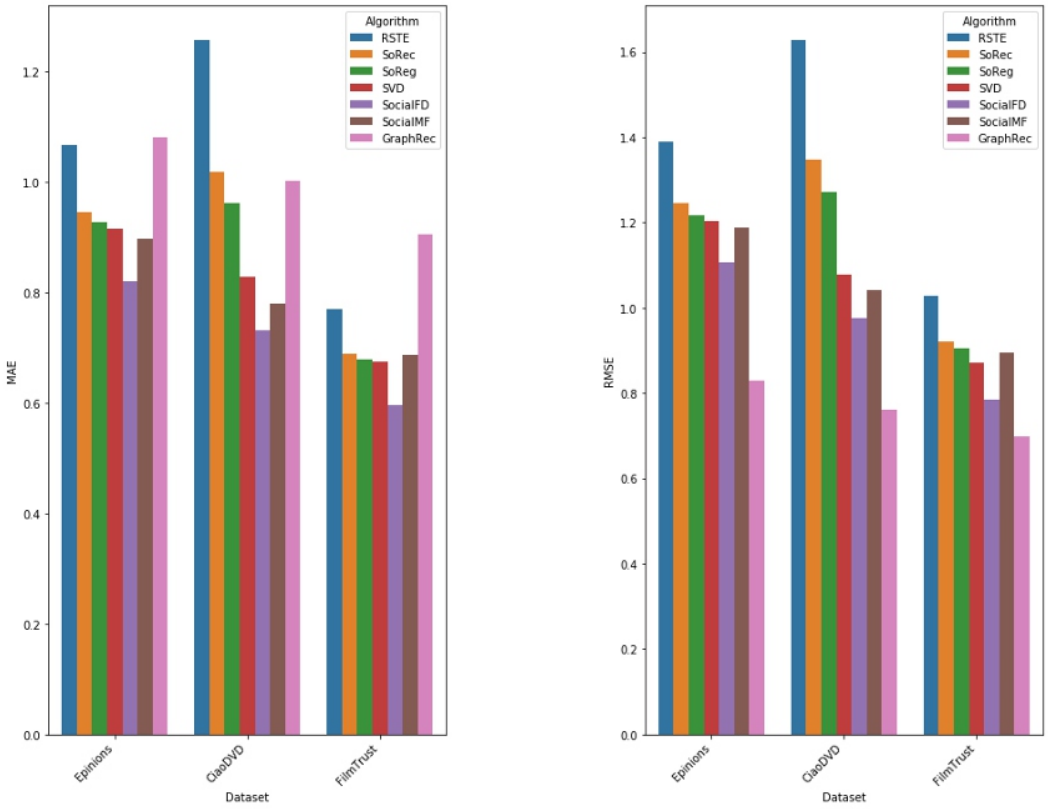

Figure 2: Performance comparison of various recommender algorithms.

These results show average RMSE and MAE from k-Fold cross validation. These algorithms show better performance with learning rates in the range 0.001 to 0.01. The performance of social algorithms is dependent on a social regularization parameter (which is given as a hyperparameter). Of these seven algorithms, SocialMF [10], SocialFD [11], and GraphRec [12] performed better in terms of RMSE and MAE. These algorithms also include social relations and use trust propagation in the recommendation process, and are usually used as baselines in new trust-based recommender systems.

3.1. SocialMF Model

Jamali and Ester proposed this method in [10]. This model incorporates propagation of trust into matrix factorization for recommending a product/an item in social networks and is closely related to STE model[8]. This model addresses trust transitivity in social networks i.e., this model considers propagation of trust. From the graphical representation of SocialMF model in Figure 3 [10], it is evident that feature vector of a user is dependent on feature vectors of user's direct neighbor. This is a recursive dependence, i.e., feature vector of direct neighbor depends on feature vectors of his/her direct neighbors. In baseline matrix factorization model [15] and STE model [8], features are learned from only observed ratings. However, in real world social networks, most of the users only participate in social network but do not express ratings to items. This makes it hard to learn feature vectors from observed ratings. SocialMF model handles these users by learning to tune the latent features of these users close to their neighbors. Hence, even if user does not express any ratings, the feature vectors are learned in a way that these are close to feature vectors of their neighbors. As the learned features are typically based on the retained observed ratings, the evaluation of these learned features for users who haven't expressed ratings is difficult. In a social network, some users actively participate in rating a product or writing a review, but most of the users express very few ratings. These users are called cold-start users. This algorithm has shown improved performance on cold-start users compared to the STE [8] model. However, the SocialMF model has higher cost in calculation of social factor and its gradients against user and item feature vectors.

3.2. SocialFD Model

In this sub-section, Social Recommender that combines Factorization and Distance metric learning, also called SocialFD [11] model is discussed. Yu et. al. proposed this model to make recommendations more reliable. This model is inspired by the concept "distance reflects likability." With the success of distance metric learning in classification tasks [18], [19] Yu et. al. integrated distance metric with matrix factorization in this model [11]. The main idea of distance metric learning is to "learn a desired distance metric that can make data points with the same class label closer and discriminate data points in different sets with larger distance" [11]. SocialFD model, on the contrary, tries to minimize the distance between each user and his/her friends and items that are rated positively. Also, this algorithm maximizes the distance between users and items rated negatively.

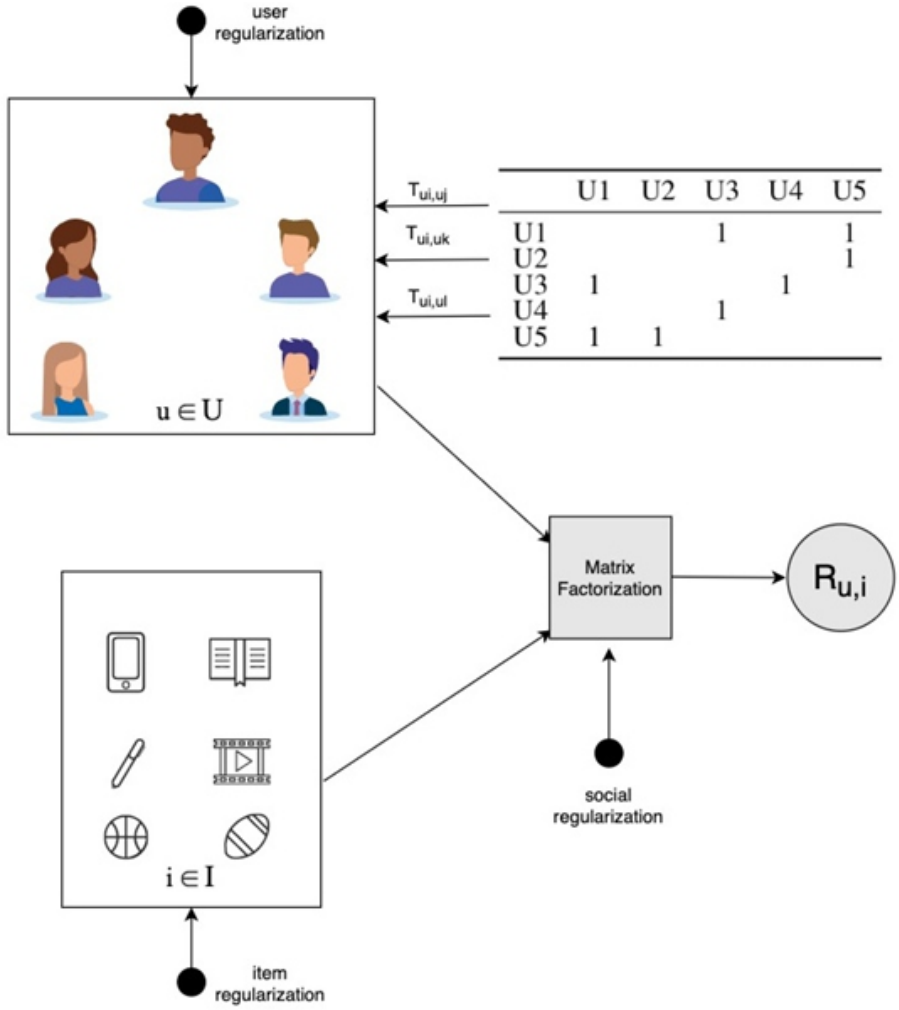

Figure 3: Graphical representation of SocialMF model[10].

The trust propagation in SocialFD model plays an important role and is similar to that of the collaborative filtering model [10]. Given the information of users' likes and friends where user is denoted by u, item by i, and friends by k, SocialFD also pulls user k and item I relatively closer in addition to pulling user u and item i closer. The sparsity problem of user k can be overcome by recommending user u's preferred items. Likewise, SocialFD model keeps user u away from disliked item j, pushing item j far from user k. Additionally, SocialFD algorithm decreases the distance between users with similar interests or indirect connections. One major drawback with the matrix factorization in general is that it is hard to combine a well-trained vector. On contrast, SocialFD model does flexible inclusion of the ready-made representation of additional knowledge. All the assumptions till now were ratings and social network connections. However, in real-time, user profiles contain huge amounts of texts. These texts can also be accumulated to enhance the quality of recommendations [20][21].

The graphical illustration of SocialFD model can be seen in Figure 4 [11]. In the figure, user and items are represented using circles and crosses respectively. These representations are represented in low-dimensional space in the figure. Closer the two symbols are, higher the probability that user prefers that item or trusts another user. Mahalanobis distance is used in this model and is calculated product of latent features difference and the distance metric. At training stage of the model, constraints are imposed such that users and preferred items/ friends are closer and distant from disliked items/users. The ratings and social connections help model to determine the positions of users and items. i.e., if user has expressed

only few ratings, his/her social connections/relations can help recommending items to user. These obtained latent features are interpreted as coordinates and the distance is used to generate meaningful recommendations.

3.3. GraphRec Model

The GraphRec Model consists of three components: user modeling, item modeling and rating prediction [12]. In user modeling, the latent factors users are learned by the model.There are two aggregations in this component, item aggregation and social aggregation. The item aggregation helps in learning item-space user latent factor from user-item ratings data. This is learned by considering the items that user u has interacted with and the opinions i.e. ratings that u has on these items. In social aggregation, social-space user latent factor is learned from the social data. According to social correlation theories[26][27], users opinions towards an item or preferences are either influenced or similar to their direct friends in social networks. In social aggregation of GraphRec,the authors proposed social-space user latent factors, which is to aggregate the item-space user latent factors of neighboring users from the social graph, to incorporate the social correlation theories. Combining the item-space latent factor and social-space latent factor, the total user latent factor is learned. The next component is item modeling. In this component, the item latent factor can be learned by user aggregation. User aggregation associates item i with users that interacted with i and their opinions. These opinions or ratings from different users help in capturing the features of same item in different ways provided by users. This helps in modeling item latent factors.

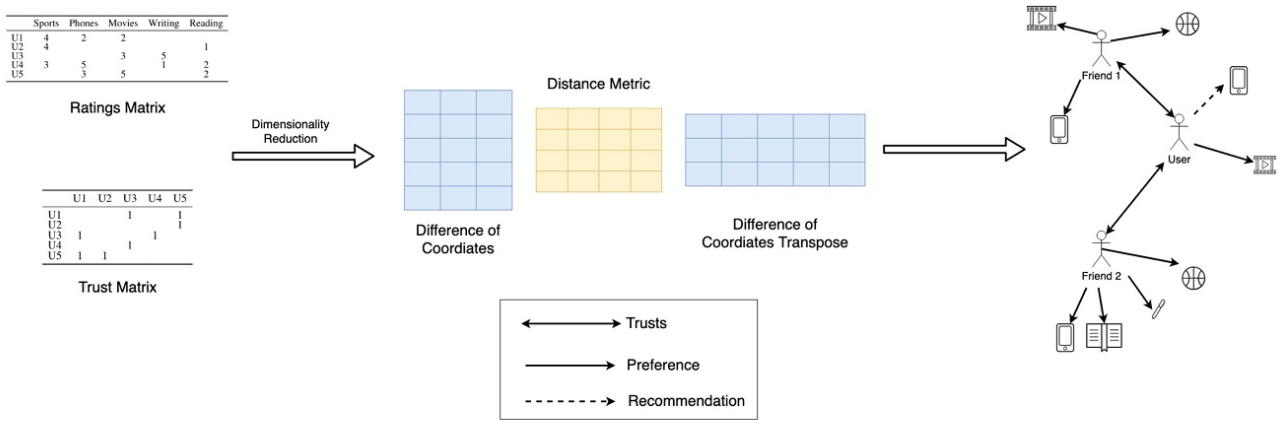

Figure 4: Graphical Representation of SocialFD model [11].

Finally in the third component, the model parameters are learned ratings are predicted using GraphRec model. To learn the model parameters, the authors utilized the most commonly used objective function which is formulated as:

$$Loss = \frac{1}{2|O|} \sum_{i,j \varepsilon O} (r'_{ij} - r_{ij})^2 \tag{1}$$

where O is the number of total observed ratings, r_{ij} is rating given by user i to item j.
For optimization of objective function, the authors used RMSprop defined in [28] rather than the vanilla Stochastic Gradient Descent (SGD). This RMSprop, each time,selects training instance randomly and updates each model parameter towards the direction of its negative gradient [12]. The three embedding item, user and opinions are initialized randomly and learned during the training stage.
The layers in the model are

$$g_1 = [h_i \oplus z_j] \tag{2}$$

$$g_2 = \sigma(W_2 \cdot g_1 + b_2) \tag{3}$$

$$g_l = \sigma(W_l \cdot g_{l-1} + b_l) \tag{4}$$

rating is predicted using:

$$r'_{ij} = w^T \cdot g_{l-1} \tag{5}$$

where l is the index of a hidden layer, and r'_{ij} is the predicted rating from u_i to v_j
To avoid overfitting, a persistent problem in optimization of deep neural networks, dropout [**?**] - a regularization technique for deep neural networks, is utilized. While testing, the dropout regularization is disabled which allows the usage of whole network.

4. Experiments

This section gives an overview of the experiments and datasets used in these experiments. Also included are the evaluation metrics used to evaluate these experiments done with SocialMF, SocialFD and GraphRec models.

4.1. Approach

In this paper, our goal is to study how the three representative social recommenders, i.e., SocialMF, SocialFD, and GraphRec, leverage social network information. Towards this goal, we make the hypothesis that "if the recommender system truly captures the social network information, making perturbations to the social network should have a significant impact on the performance of these systems." We will study these algorithms on four datasets and three ways to perturb the social network information, as described in the rest of the section.

4.2. Datasets

The major bottleneck in research of social network based recommender systems is the lack of publicly available social rating network datasets. These models were experimented with four datasets. Epinions.com is one of the popular publicly available social rating network datasets. For experimentation with these models, we used a version of Epinions dataset published by authors of [21]. On average, each user has 8 expressed ratings and have 7 direct neighbors. The next dataset we used is a version of CiaoDVD by authors of [30]. This is a smaller one compared to Epinions dataset. Another relevant dataset we used is FilmTrust [31]. FilmTrust is the smallest dataset used in experimentation crawled from FilmTrust website in 2011. Wehave created an additional dataset to experiment with these models. This dataset is an extension of TwitterEgo dataset by authors of [29]. The basic dataset consists of social circles from Twitter data which was crawled from public sources. Using Tweepy API, we extracted tweetstowhicheachoftheusersreacted.UsingthetweetIDs,thetweetstowhichusersreactedareratedas1andthetweetsfromotheruserstowhichtheusersdidnotreactwereratedas0.Finally,wecreatedaratingsdatasetfortheusersandtweets.Wegeneratedthesocialrelationsdatasetusingthesocialcirclesfromtheoriginaldataset.Wedividedthedataintotestandtrainusingstratifiedsplittechnique. Table 3showsthecompositionsofeachofthedatasetsusedforexperimentation.

Table 3: Data Statistics

Dataset	Users	Items	Ratings	Relations	Rating Scale
CiaoDVD	7,375	105,114	284,086	111,781	1.0 – 5.0
Epinions	40,163	139,738	664,823	487,183	1.0 – 5.0
FilmTrust	1,508	2,071	35,497	1,853	1.0 – 5.0
TwitterEgo	10,419	177,558	367,868	566,822	0.0 – 1.0

4.3.SocialNetworkModifications

Takingtheoriginaldatasets,wemodifiedthesocialtrustdataineachofthefourdatasetstofitforourexperiments.Thissub-sectiondiscusses themodificationswemadetothedatasetsandanexampleforeach.

4.3.1.Thereisnotrustbetweenanyusers

Forthisexperiment,wemodifiedthesocialdatabyremovingallthetruststatementsandprovidingnumberoftrust statements as 0. i.e. the socialnetworkpartofdataisnotconsidered.Thesamplesocialtrustmatrixwouldlookasintable 4.

Table 4: Sample Social Trust Matrix for Users have no trust

	U1	U2	U3	U4	U5
U1					
U2					
U3					
U4					
U5					

4.3.2. Users trust only themselves

In this experiment, the social data is modified in such a way that user u only trusts or friends with u and not anyone else. From the example from Figure 1, the modified social trust matrix would look as in Table 5.

Table 5: Sample Social Trust Matrix for Users trust only themselves

	U1	U2	U3	U4	U5
U1	1				
U2		1			
U3			1		
U4				1	
U5					1

4.3.3. Users trusts everyone else except themselves

In this experiment, the social data is modified in such a way that user u_1 only trusts or friends with all other users in the network U. From the example from Figure 1, the modified social trust matrix would look as in Table 6.

Table 6: Sample Social Trust Matrix for Users trusts everyone else except themselves

	U1	U2	U3	U4	U5
U1		1	1	1	1
U2	1		1	1	1
U3	1	1		1	1
U4	1	1	1		1
U5	1	1	1	1	

4.4. Evaluation Metrics

In these experiments, Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) are chosen to evaluate the quality of recommendations produced by these models.
RMSEis calculated using:

$$RMSE = \sqrt{\frac{1}{n}\Sigma_{u,i}(r_{ui} - r'_{ui})^2} \tag{6}$$

and MAE is defined by:

$$MAE = \frac{1}{n}\Sigma_{u,i}|r_{ui} - r'_{ui}| \tag{7}$$

where n denotes number of ratings in test set, r_{ui} is actual rating and r'_{ui} is the predicted rating. Lower MAE and lower RMSE indicate that the missing ratings are predicted more accurately.

4.5. Hyperparameter Tuning

Wetuned the hyperparameters such as learning rate, regularization factors (for users, items, and social circles), number of factors, batch size, and number of epochs. All three algorithms are highly dependent on learning rate. The performance of SocialMF and SocialFD algorithms is dependent on social regularization parameter. At lower values of social regularization parameter, the performance of SocialMF and SocialFD algorithms is similar to that of the baseline matrix factorization algorithms.

5. Results

In this section, the results of SocialMF, SocialFD, and GraphRec models are reported for each dataset and compared using the RMSE and MAEevaluation metrics. The SocialMF model results are shown in Table 7, SocialFD model results are in Table 8, and GraphRec results in Table 9. It is important to remark here that the code used for implementation of SocialMF algorithm is a modified version provided by the authors of SocialFD [11]. This version has an implementation difference and is mostly in terms of the training phase where the authors of SocialFD have used SGD to obtain the local minimum. The underlying graphical model is still the same. Results of the SocialMF model can be seen in Table 7

Table 7: SocialMF model results on 4 datasets with modified social trust information

Experiment	Metrics	CiaoDVD	Epinions	FilmTrust	TwitterEgo
User trusts friends	RMSE	1.0269	1.1044	0.8429	0.1015
	MAE	0.7718	0.8683	0.6389	0.0234
User doesn't trust anyone	RMSE	1.0248	1.1455	0.8419	0.0998
	MAE	0.7685	0.8425	0.6344	0.0160
Users trust only themselves	RMSE	1.0259	1.1037	0.8399	0.1129
	MAE	0.7817	0.8702	0.6389	0.0215
User trusts everyone except themselves	RMSE	1.0269	1.1549	0.8521	0.0973
	MAE	0.7785	0.8524	0.6449	0.0159

NOTE: Because of the usage of random seed and gradient descent in these experiments, up to two percent difference in RMSE and MAE is negligible. Also, RMSE and MAE of TwitterEgo dataset are low compared to other datasets because its rating scale is binary, i.e., 0 or 1. The difference of RMSE between experiments "users trust friends" and "user doesn't trust anyone" for Epinions increases to 4%. This increase is slightly higher than our threshold, but the MAE is lower. As graph density changes for each experiment, there might be some effect of this change on the SocialMF model.

Although both SVD-based RS [14] and SocialMF utilize matrix factorization in recommending content or products to users, there are some differences in implementation. In SocialMF, there is an additional step to update the user u's latent factors based on the user u's neighbors. This implies that only user matrix is updated in SocialMF and the social trust matrix remains the same. In case of "user doesn't trust anyone," the user matrix does not get updated and the original user matrix is retained. However, compared to SVD-based RS, the errors decrease significantly. This might be due to the use of user and item regularization terms. More exploration is needed in understanding how transitivity of trust is affected by changing the social trust information.

Based on how the inference is influenced by the trust matrix, it is hard to draw any formal conclusions about the impact of the trust matrix on the recommendation. This could be because of multiple reasons. First, based on the fact that social network metrics for these graphs are varied, it may not have anything to do with the social network structure itself. It is, however, possible that the dataset itself is not one in which social influence plays a role. We need further experimentation to verify this formally. Another reason could be that the modified trust matrices somehow "balanced" out the user latent vectors or the loss function. Details for each modified social graph are below. In SocialMF, the user latent feature vector is the weighted average of the latent feature vectors of adjacent users in the social graph. The implication of the changed social networks here is that the "user doesn't trust anyone" experiment maintains the original latent featurvectors and recommendations are solely based on users that are similar in terms of their ratings. The behavior of "user doesn't trust anyone" is therefore expected to be similar to SVD. However, the presence of the social factor (which is setup as a hyperparameter) might have scaled the recommendation scores.

Extending the argument to the "user trusts only themselves" social graph, the user latent vectors are weighted by their own ratings further. We expected that it might have reduced the influence of similar users in the traditional sense to have lesser influence on a user's recommendation. For the "user trusts everyone except themselves" social graph, the user latent vectors are influenced by the average of the latent vectors of other users. For "user trusts everyone except themselves" social graph, we expected that the average latent vectors might pollute the user similarity with respect to ratings. All of these social graphs were expected to harm the performance of SocialMF. However, the experimental results did not show a significant change in performance.

Akey reason for this could be the social trust parameter that tunes the influence of the social network on the recommendation. The authors of SocialMF [10] use a Gaussian prior to determine this factor, which plays a crucial role in their loss function. For our trivial social networks, the priors do not hold. The results, however, do highlight the need to explore the role of the social network further. In the future, we propose to run further experimentation with random trust matrices to draw more formal conclusions about the role of the social network in the SocialMF algorithm.

Table 8: SocialFD model results on 4 datasets with modified social trust information

Experiment	Metrics	CiaoDVD	Epinions	FilmTrust	TwitterEgo
User trusts friends	RMSE	0.9645	1.0458	0.7806	0.0221
	MAE	0.7261	0.7932	0.5948	0.0159
User doesn't trust anyone	RMSE	0.9641	1.0456	0.7803	0.0161
	MAE	0.7260	0.7906	0.5947	0.0069
Users trust only themselves	RMSE	0.9594	1.0485	0.7709	0.0198
	MAE	0.7368	0.7899	0.6008	0.0079
User trusts everyone except themselves	RMSE	0.9590	1.0548	0.7702	0.0159
	MAE	0.7160	0.7897	0.5893	0.0068

From the Tables 7 and 8, it can be inferred that the SocialFD algorithm in general performs better than the SocialMF algorithm consistently for these diverse datasets. This indicates that distance metrics may have an important role in social recommendations. However, when we compare these models between "user trusts friends" and "user doesn't trust anyone", there is no significant change in the RMSE and MAE. In some cases, there seems to be a marginal improvement without trust information. Looking back at our hypothesis in Section 4.1, the results indicate that there is a need for deeper exploration on these lines.

Table 9: GraphRec model results on 4 datasets with and without social trust information

Experiment	Metrics	CiaoDVD	Epinions	FilmTrust	TwitterEgo
User trusts friends	RMSE	1.0022	1.0818	0.9057	0.0209
	MAE	0.7611	0.8299	0.6970	0.0149
User doesn't trust anyone	RMSE	0.9989	1.0786	0.9051	0.0194
	MAE	0.7645	0.8255	0.6921	0.0140
Users trust only themselves	RMSE	1.0342	1.1008	0.8975	0.0198
	MAE	0.7903	0.8382	0.7108	0.0151
User trusts everyone except themselves	RMSE	0.9689	1.0983	0.9005	0.0182
	MAE	0.7945	0.8356	0.6891	0.0137

GraphRec showed improved performance while considering social trust information when compared to other models like SocialMF [10], SoRec[9], etc. Although GraphRec is a more complex deep learning algorithm, we did not observe much difference in RMSE and MAE as compared to SocialFD. In some cases, these errors increased significantly. For example, for our TwitterEgo dataset, the MAE increased significantly and for the FilmTrust dataset both metrics increased. We also observed that for the same model, performance improves in some scenarios and deteriorates in the others. Regardless, the difference in performance is very little. This highlights the fact that deep learning techniques cannot guarantee a better performance even with huge datasets such as Epinions and TwitterEgo. It is particularly interesting that in most cases, the trivial social trust datasets perform better than the default "user trusts friends" datasets.

All of the above experiments indicate the following key observations. Comparing social recommendation algorithm with trust information indicates that SocialFD is a superior algorithm for all the datasets and social trust data compared to SocialMF. However, from our experiments, social trust information does not significantly improve the performance of these representative models on diverse datasets. Also, we understand that superior performance of SocialFD algorithm has something to do with the utilization of distance metric. Further study is needed in the direction of incorporating distance metric learning into social recommender systems. The computational overhead and complexity of deep learning models may not make a significant difference to the performance. It also highlights the importance of studying the interpretability of deep learning social recommender systems in general.

6. Conclusion and future work

Recommender systems are powerful tools that filter and recommend content/information relevant to a given user. With the advent of social networks, it has become very important to utilize the data on social ties between in a social network to recommend a product to the users who expressed a few ratings. In this paper, we explored three such social recommender models, SocialMF, SocialFD, and GraphRec. SocialMF is a model that incorporates trust propagation into matrix factorization, SocialFD model uses distance metric learning in addition to matrix factorization, and GraphRec is a model that uses attenttion-based graph neural networks.

Experiments on 4 real life datasets, CiaoDVD, Epinions, FilmTrust, and TwitterEgo, show that these algorithms outperform the conventional collaborative filtering algorithms as well as the previously developed social recommender systems. Of these three, the SocialFD model performs better than the SocialMF model with inclusion of trust. At lower social parameter values, these models' performance is similar to the performance of collaborative filtering algorithms. However, when social trust factor is not given, these models show similar performance compared to the models that contain social trust parameter. From these experiments, it can be seen that there is less conclusive evidence that social recommender systems are influenced by social trust data. The experiments highlight the need to explore further to gain better understanding of the role of social networks in recommender systems.

This work suggests some interesting directions for future research. These models can be extended further to handle zero and negative trust relations. In general, negative trust, also called distrust, gives more information about a user than positive opinions. Also, currently, social regularization parameter is given as an input to these models. Future work can help in the development of a model that incorporates automatic tuning of social trust. In real-time, user profiles and item profiles contain huge amounts of text data and other features. These features can also be accumulated into the recommender system to enhance the quality of recommendations. In the future, we would like to explore a dataset where social network is explicitly synthesized to have an impact on recommendation and repeat these experiments on that synthetic dataset.

Acknowledgement

This research is partially supported by Charles W. Davidson College of Engineering at San Jose State University.

References

[1] Ricci, F., Rokach, L. & Shapira, B. Introduction to Recommender Systems Handbook. Recommender Systems Handbook. pp. 1-35 (2011)

[2] Goldberg, D., Nichols, D., Oki, B. & Terry, D. Using Collaborative Filtering to Weave an Information Tapestry. Commun. ACM. 35, 61-70 (1992,12)

[3] Qian, X., Feng, H., Zhao, G. & Mei, T. Personalized Recommendation Combining User Interest and Social Circle. IEEE Transactions On Knowledge And Data Engineering. 26, 1763-1777 (2014)

[4] Jamali, M. & Ester, M. TrustWalker: A Random Walk Model for Combining Trust-Based and Item-Based Recommendation. Proceedings Of The 15th ACMSIGKDDInternational Conference On Knowledge Discovery And Data Mining. pp. 397-406 (2009)

[5] Massa, P. & Avesani, P. Trust-Aware Recommender Systems. (Association for Computing Machinery,2007)

[6] Ziegler, C. Towards decentralized recommender systems. (University of Freiburg,2005)

[7] Golbeck, J. Computing and Applying Trust in Web-based Social Network. (University of Maryland,2005)

[8] Ma, H., King, I. & Lyu, M. Learning to Recommend with Social Trust Ensemble. (Association for Computing Machinery,2009)

[9] Ma, H., Yang, H., Lyu, M. & King, I. SoRec: Social Recommendation Using Probabilistic Matrix Factorization. Proceedings Of The 17th ACM Conference On Information And Knowledge Management. pp. 931-940 (2008)

[10] Jamali, M. & Ester, M. A Matrix Factorization Technique with Trust Propagation for Recommendation in Social Networks. Proceedings Of The Fourth ACMConference On Recommender Systems. pp. 135-142 (2010)

[11] Yu, J., Gao, M., Rong, W., Song, Y. & Xiong, Q. A Social Recommender Based on Factorization and Distance Metric Learning. IEEE Access. 5 pp. 21557-21566 (2017)

[12] Fan, W., Ma, Y., Li, Q., He, Y., Zhao, E., Tang, J. & Yin, D. Graph Neural Networks for Social Recommendation. (Association for Computing Machinery,2019)

[13] Koren, Y. Factorization Meets the Neighborhood: A Multifaceted Collaborative Filtering Model. (Association for Computing Machinery,2008)

[14] Koren, Y., Bell, R. & Volinsky, C. Matrix Factorization Techniques for Recommender Systems. Computer. 42, 30-37 (2009)

[15] Richardson, M. & Domingos, P. Mining Knowledge-Sharing Sites for Viral Marketing. (Association for Computing Machinery,2002)

[16] Levien, R. Attack-Resistant Trust Metrics. Computing With Social Trust. pp. 121-132 (2009)

[17] Wu, Q., Zhang, H., Gao, X., He, P., Weng, P., Gao, H. & Chen, G. Dual Graph Attention Networks for Deep Latent Representation of Multifaceted Social Effects in Recommender Systems. (Association for Computing Machinery,2019)

[18] Goldberger, J., Hinton, G., Roweis, S. & Salakhutdinov, R. Neighbourhood Components Analysis. Advances In Neural Information Processing Systems 17. pp. 513-520 (2005)

[19] Weinberger, K. & Saul, L. Distance Metric Learning for Large Margin Nearest Neighbor Classification. Journal Of Machine Learning Research. 10, 207-244 (2009), http://jmlr.org/papers/v10/weinberger09a.html

[20] Musto, C. Enhanced Vector Space Models for Content-Based Recommender Systems. (Association for Computing Machinery,2010)

[21] Semeraro, G., Lops, P., Basile, P. & Gemmis, M. Knowledge Infusion into Content-Based

Recommender Systems. (Association for Computing Machinery,2009)

[22] Kipf, T. & Welling, M. Semi-Supervised Classification with Graph Convolutional Networks. (2017)

[23] Defferrard, M., Bresson, X. & Vandergheynst, P. Convolutional Neural Networks on Graphs with Fast Localized Spectral Filtering. (Curran Associates Inc.,2016)

[24] Hamilton, W., Ying, R. & Leskovec, J. Inductive Representation Learning on Large Graphs. (Curran Associates Inc.,2017)

[25] Ma, Y., Wang, S., Aggarwal, C., Yin, D. & Tang, J. Multi-dimensional graph convolutional networks. SIAM International Conference On Data Mining, SDM2019. pp. 657-665 (2019), 19th SIAM International Conference on Data Mining, SDM 2019

[26] MARSDEN, P. & FRIEDKIN, N. Network Studies of Social Influence. Sociological Methods Research. 22, 127-151 (1993)

[27] McPherson, M., Smith-Lovin, L. & Cook, J. Birds of a Feather: Homophily in Social Networks. Review Of Sociology. 27 pp. 415-444 (2001)

[28] Tieleman, T. & Hinton, G. Lecture 6.5-rmsprop: Divide the gradient by a running average of its recent magnitude. COURSERA: Neural Networks For Machine Learning. 4, 26-31 (2012)

[29] Leskovec, J. & Mcauley, J. Learning to Discover Social Circles in Ego Networks. Advances In Neural Information Processing Systems 25. pp. 539-547 (2012)

[30] Zhao, G., Qian, X. & Xie, X. User-Service Rating Prediction by Exploring Social Users' Rating Behaviors. IEEE Transactions On Multimedia. 18, 496-506 (2016)

[31] Tang, J., Gao, H. & Liu, H. MTrust: Discerning Multi-Faceted Trust in a Connected World. Proceedings Of The Fifth ACM International Conference On Web Search And Data Mining. pp. 93-102 (2012)

[32] Guo, G., Zhang, J. & Yorke-Smith, N. A Novel Bayesian Similarity Measure for Recommender Systems. (AAAI Press,2013)

[33] Ma, H., Zhou, D., Liu, C., Lyu, M. & King, I. Recommender systems with social regularization. Proceedings Of The Fourth ACM International Conference On Web Search And Data Mining. pp. 287-296 (2011).

Instructions for Authors

Essentials for Publishing in this Journal

1 Submitted articles should not have been previously published or be currently under consideration for publication elsewhere.

2 Conference papers may only be submitted if the paper has been completely re-written (taken to mean more than 50%) and the author has cleared any necessary permission with the copyright owner if it has been previously copyrighted.

3 All our articles are refereed through a double-blind process.

4 All authors must declare they have read and agreed to the content of the submitted article and must sign a declaration correspond to the originality of the article.

Submission Process

All articles for this journal must be submitted using our online submissions system. http://enrichedpub.com/ . Please use the Submit Your Article link in the Author Service area.

Manuscript Guidelines

The instructions to authors about the article preparation for publication in the Manuscripts are submitted online, through the e-Ur (Electronic editing) system, developed by **Enriched Publications Pvt. Ltd**. The article should contain the abstract with keywords, introduction, body, conclusion, references and the summary in English language (without heading and subheading enumeration). The article length should not exceed 16 pages of A4 paper format.

Title

The title should be informative. It is in both Journal's and author's best interest to use terms suitable. For indexing and word search. If there are no such terms in the title, the author is strongly advised to add a subtitle. The title should be given in English as well. The titles precede the abstract and the summary in an appropriate language.

Letterhead Title

The letterhead title is given at a top of each page for easier identification of article copies in an Electronic form in particular. It contains the author's surname and first name initial .article title, journal title and collation (year, volume, and issue, first and last page). The journal and article titles can be given in a shortened form.

Author's Name

Full name(s) of author(s) should be used. It is advisable to give the middle initial. Names are given in their original form.

Contact Details

The postal address or the e-mail address of the author (usually of the first one if there are more Authors) is given in the footnote at the bottom of the first page.

Type of Articles

Classification of articles is a duty of the editorial staff and is of special importance. Referees and the members of the editorial staff, or section editors, can propose a category, but the editor-in-chief has the sole responsibility for their classification. Journal articles are classified as follows:

Scientific articles:

1. Original scientific paper (giving the previously unpublished results of the author's own research based on management methods).

2. Survey paper (giving an original, detailed and critical view of a research problem or an area to which the author has made a contribution visible through his self-citation);

3. Short or preliminary communication (original management paper of full format but of a smaller extent or of a preliminary character);

4. Scientific critique or forum (discussion on a particular scientific topic, based exclusively on management argumentation) and commentaries. Exceptionally, in particular areas, a scientific paper in the Journal can be in a form of a monograph or a critical edition of scientific data (historical, archival, lexicographic, bibliographic, data survey, etc.) which were unknown or hardly accessible for scientific research.

Professional articles:

1. Professional paper (contribution offering experience useful for improvement of professional practice but not necessarily based on scientific methods);
2. Informative contribution (editorial, commentary, etc.);
3. Review (of a book, software, case study, scientific event, etc.)

Language

The article should be in English. The grammar and style of the article should be of good quality. The systematized text should be without abbreviations (except standard ones). All measurements must be in SI units. The sequence of formulae is denoted in Arabic numerals in parentheses on the right-hand side.

Abstract and Summary

An abstract is a concise informative presentation of the article content for fast and accurate Evaluation of its relevance. It is both in the Editorial Office's and the author's best interest for an abstract to contain terms often used for indexing and article search. The abstract describes the purpose of the study and the methods, outlines the findings and state the conclusions. A 100- to 250-Word abstract should be placed between the title and the keywords with the body text to follow. Besides an abstract are advised to have a summary in English, at the end of the article, after the Reference list. The summary should be structured and long up to 1/10 of the article length (it is more extensive than the abstract).

Keywords

Keywords are terms or phrases showing adequately the article content for indexing and search purposes. They should be allocated heaving in mind widely accepted international sources (index, dictionary or thesaurus), such as the Web of Science keyword list for science in general. The higher their usage frequency is the better. Up to 10 keywords immediately follow the abstract and the summary, in respective languages.

Acknowledgements

The name and the number of the project or programmed within which the article was realized is given in a separate note at the bottom of the first page together with the name of the institution which financially supported the project or programmed.

Tables and Illustrations

All the captions should be in the original language as well as in English, together with the texts in illustrations if possible. Tables are typed in the same style as the text and are denoted by numerals at the top. Photographs and drawings, placed appropriately in the text, should be clear, precise and suitable for reproduction. Drawings should be created in Word or Corel.

Citation in the Text

Citation in the text must be uniform. When citing references in the text, use the reference number set in square brackets from the Reference list at the end of the article.

Footnotes

Footnotes are given at the bottom of the page with the text they refer to. They can contain less relevant details, additional explanations or used sources (e.g. scientific material, manuals). They cannot replace the cited literature.
The article should be accompanied with a cover letter with the information about the author(s): surname, middle initial, first name, and citizen personal number, rank, title, e-mail address, and affiliation address, home address including municipality, phone number in the office and at home (or a mobile phone number). The cover letter should state the type of the article and tell which illustrations are original and which are not.