The Feedback-Star Paradigm: Redefining Social Network Clustering

by

Examination Roll

202165

202166

Session: 2019-2020

Bachelor of Science

in

Computer Science and Engineering

Faculty of Mathematical and Physical Sciences

Jahangirnagar University

Dhaka-1342, Bangladesh.

January 15, 2025

Contents

Abstract

This paper presents a novel approach to the implementation of the Star Clustering Algorithm for social network analysis by including user-initiated feedback mechanism for the betterment of cluster formation in a non-static enterprise setting. The proposed Feedback-Star Paradigm is a useful system that enhances user-centered innovation by rectifying the key problems of the clustering methods. Our algorithm identifies star centers based on node degree centrality and employs a deterministic tie-breaking mechanism to ensure reproducible clustering outcomes. The distinguishing feature of the method lies in its ability to iteratively adjust the clustering thresholds, maintaining computational efficiency while preserving the clustering accuracy in heterogeneous network structures. Experimental validation utilizing real-world social networking datasets shows that there is a marked increase in both cluster cohesion and inter-cluster separation as compared to conventional approaches and algorithms. The framework is also shown to be efficient with the ever-changing network topology and hence can be easily adapted in practice. This work contributes to the interdisciplinary domain by developing a new self-adjusting clustering method that improves our comprehension of social network architectures and allows for better analysis of social activities within the context of organizations.

Keywords: Star Clustering Algorithm, Social Network Analysis, Adaptive Feedback, Dynamic Clustering, Network Topology

Chapter 1

Introduction

A social network is a structure or system comprising individuals or entities, often referred to as nodes, that are interconnected by relationships known as edges. These networks can represent various forms of relationships, such as friendships, shared interests, professional connections, collaborations and influence etc. In the modern world, social networks play a crucial role in facilitating communication, information exchange, building relationships and community building, and enhancing business productivity.

The context of social network, clustering can be used to identify and analyze the structure and patterns of relationship within organizational network.It also can be useful for identifying clusters within a network. Key concepts in social network clustering analyzing centrality which measures the importance of the nodes and identifies the tightly connected groups of nodes within the network.This clustering analyses provides insights into how information or interest spreads, how communities form and how influential entities impact the networks behavior.

In the technological era, social network clustering analysis has expanded into diverse domains, including marketing, healthcare, and collaboration networks.For example, businesses uses social network clustering to identify key opinion leaders for marketing campaigns. Meanwhile, collaborative organizations analyze specific domains to create researcher communities where individuals focusing on the same research area tend to collaborate frequently.

The Star Approach Algorithm has recently gained significant attention for various clustering tasks, such as filtering, visualizing multi-dimensional clusters, survey clustering, and heterogeneous information clustering. However, its application in the social network domain remains largely unexplored. To address this gap, we propose an efficient feedback-based Star Approach clustering algorithm specifically designed to identify meaningful clusters within social networks by optimizing star cluster formation. Our approach is motivated by the limitations of existing unsupervised clustering algorithms, which often fail to effectively capture the complex, feedback-driven relationships inherent in social networks.

Traditional clustering approaches, such as Agglomerative Clustering, K-means Clustering, and the K-Medoids algorithm, have certain limitations and disadvantages. We initially applied the Agglomerative and K-Medoids clustering algorithms to a human-generated network dataset. However, these algorithms failed to consistently identify meaningful and robust clusters within the network. Moreover, they lacked the flexibility to incorporate user feedback to refine the clustering results. A major drawback of these algorithms is their inherent limitation in identifying predefined clusters, which prevents the discovery of all possible clusters within the network.

Our proposed feedback-based Star Approach Algorithm aims to overcome the limitations of traditional clustering algorithms. When applied to a preprocessed social network dataset, the Star Approach Algorithm generally identifies distinct clusters within the network. However, it may fall short in detecting overlapping clusters. To address this challenge, we propose incorporating a thresholding technique guided by user feedback, which would enable the identification of overlapping clusters within the network.

We further compare the performance of the feedback-based Star Approach Algorithm with traditional clustering algorithms. Finally, we analyze the validity of the identified meaningful clusters using metrics such as F1 score.

Chapter 2

Literature Review

Clustering is a fundamental data analysis method used to group similar entities based on predefined metrics. Aslam et al. introduced the Star Clustering Algorithm to overcome the limitations of static clustering techniques [1]. This innovative method organizes data into clusters using dense, star-shaped subgraphs, ensuring computational efficiency and maintaining high similarity within clusters. Unlike traditional methods, the Star Algorithm does not require predefined cluster numbers, making it suitable for applications where the topic distribution is unknown. This study provides a theoretical framework for clustering in both static and dynamic environments, paving the way for further research in adaptive clustering techniques. In their 2004 study, Aslam et al. applied the Star Clustering Algorithm to dynamic document filtering, demonstrating its potential in enhancing precision and recall [2]. By grouping dynamically arriving document streams into meaningful clusters, the algorithm facilitated more efficient retrieval of relevant information. However, the study highlights a critical limitation: reliance on user feedback introduces potential biases in cluster formation. Despite this, the research demonstrates the algorithm's efficacy in real-time filtering tasks, emphasizing the need for autonomous mechanisms to minimize feedback dependency. Suárez and Pagola extended the Star Clustering Algorithm by proposing the Generalized Star (GStar) Algorithm, addressing critical issues like data order dependency and suboptimal cluster formations [3]. The GStar algorithm incorporates overlapping clusters and redefines star structures to produce logical and consistent results. While it demonstrates improved robustness and accuracy across varying datasets, the study identifies scalability for large-scale data as a key area for future research. This extension significantly advances clustering methodologies by refining and optimizing the original algorithm. Aslam et al. expanded the application of the Star Clustering Algorithm, focusing on its role in organizing static and dynamic information [4]. The study emphasizes the algorithm's computational efficiency and its ability to identify natural topic structures within datasets. The Star Algorithm is presented as a robust alternative to traditional methods, particularly in environments requiring dynamic adaptability. This research validates the algorithm's utility across various domains, setting a strong foundation for future developments. Gil-García et al. presented an extended version of the Star Clustering Algorithm to enhance computational efficiency [5]. This extended algorithm redefines star structures and introduces new clustering criteria to address limitations like redundant clusters and order dependency. Evaluations on TREC datasets demonstrate its ability to produce better clustering quality with fewer computational resources. However, the study notes that further optimization is required to make the algorithm suitable for real-time applications and high-dimensional datasets. Mei and Chen developed a fuzzy clustering approach for multi-type relational data, emphasizing star-structured datasets [6]. By introducing fuzzy memberships, the algorithm captures nuanced relationships and allows for overlapping clusters, making it particularly effective in handling heterogeneous datasets. While the study highlights significant improvements in clustering accuracy and flexibility, the high computational cost of the fuzzy approach remains a challenge. This research lays the foundation for future advancements in clustering techniques, particularly for complex relational data. Gil-García and Pons-Porrata introduced a hierarchical adaptation of the Star Clustering Algorithm for dynamic document collections [7]. This method organizes documents into hierarchical structures, addressing the scalability issues often faced in clustering large datasets. The algorithm is particularly effective in handling evolving information, although further research is needed to enhance its real-time application capabilities. This study bridges the gap between hierarchical and dynamic clustering, offering a robust approach for managing large, evolving datasets. Aslam et al. establish online and offline clustering techniques that can be used to organize static and dynamic information systems [8]. This algorithm clusters data into groups based on the inherent structure of the collection and computationally efficient covers of the dense subgraphs. The Star Clustering Algorithm guarantees the best among any clustering techniques such as single-link and average-link algorithms in terms of cluster quality and topic similarity with computational efficiency. Its offline variant is meant for static collections while the online version dynamically adapts to changes, being suitable for the evolving nature of such corpora as news feeds. The application ranges from scalable organizing of huge datasets to text filtering and persistent queries, demonstrating adaptability and efficiency. Empirical results with datasets like TREC established this as the best item in terms of accuracy and performance and as a terminologically necessary instrument of modern information retrieval systems. Berkhin highlights clustering's significance in diverse applications, particularly in managing large datasets and uncovering hidden patterns [9]. The paper discusses traditional clustering techniques such as k-means, hierarchical clustering, and density-based methods, emphasizing their limitations in dynamic and heterogeneous environments. Berkhin's work serves as a comprehensive survey, providing foundational insights into clustering's mathematical principles and practical applications. This research underscores the need for adaptive clustering algorithms capable of handling evolving data and large-scale datasets. Mei et al. extended the Star Algorithm to handle central-to-central relationships in heterogeneous information networks [10]. The CluEstar algorithm achieves high clustering quality by integrating diverse relationships, such as those in bibliographic datasets. Despite its effectiveness, the study identifies computational demands as a limitation for large-scale applications. This work broadens the algorithm's scope, offering a framework for clustering complex, multi-relational networks while calling for further optimization. Schmeja evaluated various algorithms for identifying stellar clusters in dense fields, focusing on their performance and utility [11]. The study compared methods like star counts, nearest-neighbor techniques, and minimum spanning tree separation, highlighting their strengths and weaknesses. While the research underscores the importance of smoothing techniques for detecting low-density clusters, it also notes the limitations of these methods in identifying hierarchical structures. The paper emphasizes the need for refined algorithms to improve accuracy in complex astrophysical datasets. Hua et al. proposed the Star-Based Learning Correlation Clustering (SL-CC) algorithm to manage large-scale signed graphs [12]. By reducing graph size using star motifs, the algorithm balances computational efficiency with clustering accuracy. The study demonstrates the method's effectiveness in applications like social networks and text mining, though further validation on diverse datasets is required. This research contributes significantly to clustering methodologies for signed networks, addressing challenges in scalability and complexity. Daniel Puschmann et al. focused on clustering real-world streaming data generated by the Internet of Things (IoT) [13]. They developed an adaptive method that can dynamically identify the number of clusters within a data stream and measure their quality. This approach is designed to handle situations where the characteristics of the data change over time. Notably, it operates without requiring prior knowledge about the number of clusters. Their research demonstrates a robust and adaptive clustering method that effectively handles evolving data, outperforming existing methods on both simulated and real-world datasets. Alexis Pister et al. proposed a new approach Prior Knowledge Clustering (PK-clustering), a novel approach to assist social scientists in identifying meaningful clusters within social networks [14]. This user-centered approach integrates prior knowledge and provides tools for evaluating clustering algorithms. By combining data mining with visualization, PK-clustering enhances the interpretability of results, making it a valuable tool for researchers analyzing medium-sized networks. However, this approach has limitations, including limited algorithm diversity, insufficient parameter exploration, challenges with disconnected network components, incomplete visualization integration, underdeveloped provenance features, reliance on user input, and a lack of extensive real-world validation.Despite these limitations, PK-clustering serves as a valuable foundation and a useful tool for future research in social network analysis. The paper by Balcan and Blum describes a new framework, being interactive, for clustering data by collecting user feedback through split and merge requests, which makes it adaptive toward disambiguating the clustering results. Unlike conventional models, which rely heavily on distributional assumptions, this framework delivers flexibility and minimum-constraint clustering under the assumption that each cluster conforms to a suitable, given concept class. They propose algorithms that create clusterings with respect to the logarithmic number of user interactions against dataset size because the model turns out to be computationally very efficient. It works better in situations where clusters may be possible or more than one and the user input is supposed to drive the resulting outcome towards some specific objective. The theoretical guarantees for clusterability have been proved in this research, along with major improvements over generative models by minimization of assumption without loss of effectiveness [15]. Nam P. Nguyen et al. addresses the challenges the challenges of detecting and tracking community structures in dynamic social networks by proposing Quick Community Adaptation (QCA) [16]. They proposed Quick Community Adaptation (QCA), an adaptive algorithm extensively tested on real-world datasets. QCA is specifically designed to identify community structures in networks where frequent changes occur, such as online social networks and mobile ad hoc networks (MANETs). Traditional methods often struggle to keep pace with the dynamic nature of these networks, where nodes and edges are constantly being added or removed. QCA addresses this limitation by efficiently updating community structures as the network evolves. Fanzhen Liu et al. addresses community detection in dynamic networks by introducing DECS (Dynamic Evolutionary Clustering System), a multi objective evolutionary clustering algorithm [18]. DECS incorporates a migration operator, a genome-based network representation, and label propagation to enhance search efficiency and solution quality compared to existing methods. Experimental results demonstrate that DECS outperforms other state-of-the-art methods, such as ECD, DYNMOGA, and FacetNet, on both synthetic and real-world datasets. Future research will focus on scaling DECS for larger dynamic networks and expanding its applicability to more complex systems. Bae et al. [19] examined 105 studies and reported that its efficiency has been high in enhancing the quality of the cluster and in meeting the requirements of users in terms of the desired outputs. The space for traditional clustering was sorely lacking in flexibility when faced with multifaceted situations where domain expertise was required. The deficiency of traditional clustering is filled in by meta-interactive clustering whereby aid is given in terms of relaying the finished product, altering parameters as well as providing user engagement within the process. This approach is subdivided into three categories: direct engagement with the outcomes (e.g. combining or separating clusters, applying must-link or cannot-link constraints), modification of parameters such as (e.g. the number of clusters or similarity measures), and solicitation of user feedback by the machine. Tools for visualization are critical in conducting these interactions. The positive outcomes of this collaboration include higher quality of clustering, better understanding of data and end results that are in accordance with users' standpoint. Automation versus user control balance, intuitive usage of the interface, and creation of evaluation measures remain the challenges that will be encountered in the use of this tool. Other tools are scalable, algorithms are adaptive, and evaluation frameworks will be more sophisticated [19].

Chapter 3

Proposed Methodology

3.1 Problem Statement

Social networks have complex and ever-changing features, and they are vital in communication and decision-making processes within society. Clustering is very important for identifying structures that are not easily demonstrated within these networks, nevertheless K-Medoids, Agglomerative Clustering are some of the classical methods that have prominent numerous drawbacks. K-Medoids is rather efficient in noise filtering, however it is not computationally efficient for large data sets, whereas in the case of Agglomerative Clustering, it is incapable of adjusting to the changing structures of the network or clusters that overlaps. These limitations curtail the potential of social networks analysis to achieve its purpose when applied to real situations. Nature of social networks is characterized by overlapping communities and by their dynamic nature that evolves through time with velocity which is not managed well by the traditional style. Insightful enable technologies such as marketing, decision-making and influence analysis potential remain untapped without comprehensive clustering strategies that are scaled up to fit business needs. To counter these limitations, the research proposes a novel strategies called the Feedback-Star Paradigm which builds on the Star Clustering Algorithm. More so by enforcing dynamic feedback mechanisms and appropriate overlap detection techniques, this paradigm lays down a comprehensive, adaptive and efficient framework for the computation of dynamic social network models. Thus, this work seeks to improve the accuracy and usefulness of social clustering, to facilitate deep in-depth analysis of social and political relations.

3.2 Workflow

In this approach, we will use the Star Approach Algorithm for clustering, but with a modification: the clusters will be reformed based on user feedback. This means we introduce a feedback-based Star Approach Algorithm to identify meaningful clusters within the network. The methodology is divided into three main parts:

1. Building the preprocessed dataset
2. Applying the Star Approach Algorithm to identify meaningful clusters
3. Reforming the clusters based on user feedback

The flowchart **below** illustrates the workflow of our proposed methodology. Here, we provide a concise description of each step in the given flowchart.

1. **Dataset:** Initially the processes starts with a preprocessed dataset in the form of a network graph, which represents data points (nodes) and their relationships(edges). When incorporating user feedback to create meaningful clusters, dataset preprocessing plays an important role in preparing the data for clustering. User feedback can provide additional insights that help refine the clustering process. For example, feedback might indicate that certain nodes should belong to the same cluster or that specific connections should be strengthened or weakened. This feedback is often incorporated into the dataset during preprocessing, either by modifying the graph's structure (e.g., adjusting edge weights) or by introducing new features to represent the feedback. Incorporating feedback into the similarity or distance measures, ensuring the clusters better align with user expectations. That means preprocessing will help to integrate user feedback effectively, facilitating to create more accurate and meaningful clusters based on both inherent data relationships and user-provided insights.

2. **Similarity/Dissimilarity Matrix:** Distance measures help to generate a similarity or dissimilarity matrix by calculating how similar or different the nodes are based on their relationships within the network. A value close to 1 in the case of cosine similarity, or a small value in case of Euclidean distance, would indicate high similarity (nodes are close), while a value close to 0 or a large distance would indicate dissimilarity (nodes are far).

3. **Thresholding:** We want to make the Star Approach Algorithm a feedback-based semi-supervised algorithm for network clustering, thresholding can play an important role in refining the clustering process based on user feedback. Once the

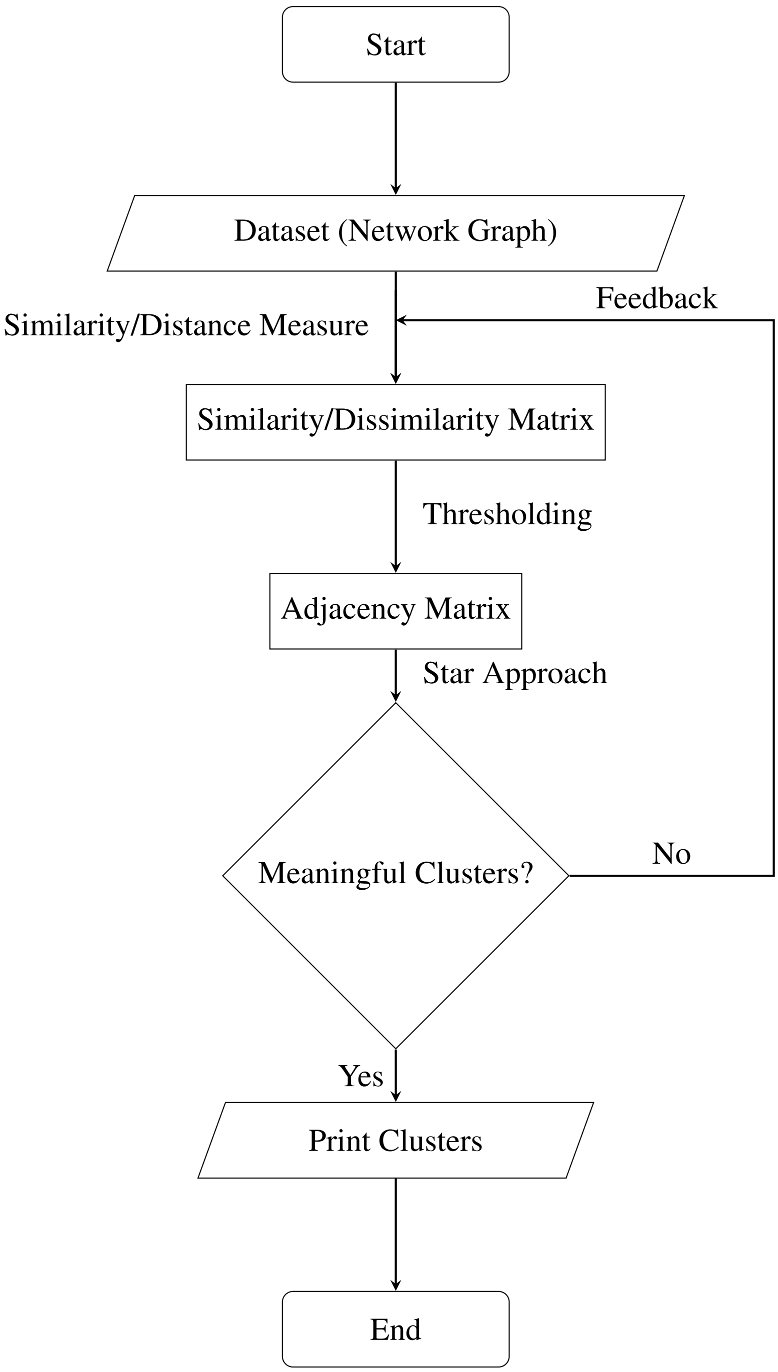

Figure 3.1: Flowchart of the feedback-based clustering process using the Star Approach algorithm.

similarity or dissimilarity matrix is updated with feedback, we can apply a threshold to simplify the network.

4. **Adjacency Matrix:** To generate an adjacency matrix after thresholding we first need to filter the edges in our graph based on the threshold value. The thresholding step helps determine which edges remain in the graph and which ones are discarded. For example, if we are working with a human generated graph we

can use a threshold on edge based on conditions such as the degree of nodes or connection patterns.

5. **Star Approach:** The Star Approach Algorithm is applied to the adjacency matrix to identify clusters. The algorithm is likely a form of clustering where nodes are grouped based on their relationships (edges) in the network graph.

The given flowchart outlines the workflow for adapting the Star Approach Algorithm to a feedback-based approach for identifying meaningful clusters within the social network.

3.3 Comparative Evaluation of Clustering Techniques

Let us take advantage of the high F1 scores for the Feedback-Star paradigm over conventional clustering algorithms. It is a balanced measure that considers a tradeoff between precision and recall. Unlike conventional methods, for example, K-Medoids and Agglomerative Clustering, the Feedback-Star Paradigm is designed specifically to solve the problems posed by dynamic and overlapping clusters in social networks.

3.3.1 Performance Comparison

The quantitative F1 score comparison shows how effective the Feedback-Star paradigm could be. The proposed algorithm achieves an F1 score of models that not only outperformed K-Medoids but also Agglomerative Clustering to highlight the fact that the Feedback-Star paradigm would generate more cohesive and accurate clusters, especially when looking at complex network structures.

3.3.2 Advantages of the Feedback-Star Paradigm

- **Improved Overlap Handling:** It captures the overlapping clusters by dynamically adjusting thresholds, which is a feature where most of the traditional algorithms lack.
- **Efficiency for Large Networks:** Despite its sophisticated approach, the Feedback-Star Paradigm remains computationally efficient and hence suitable for large-scale datasets.

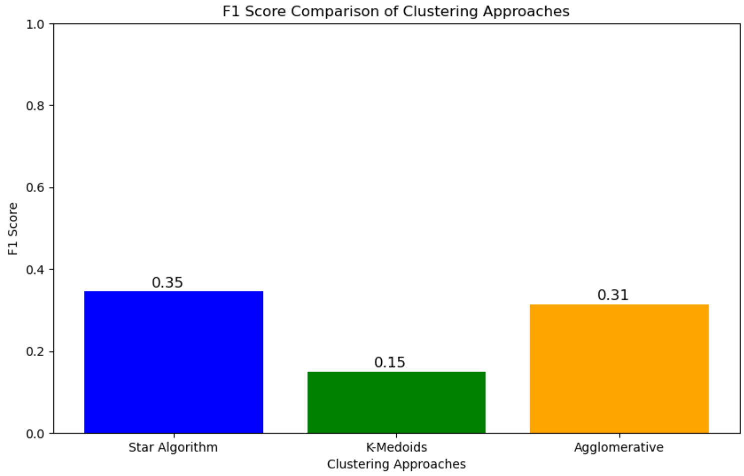

Figure 3.2: F1 Score Comparison of Clustering Approaches. The Feedback-Star Paradigm outperforms K-Medoids and Agglomerative Clustering in F1 score, demonstrating its superior clustering performance.

3.3.3 Limitations of Traditional Approaches

- **K-Medoids:** Sensitive to initial seeding of clusters and struggles with overlapping clusters thus it's low in F1 score.
- **Agglomerative Clustering:** Shows only moderate efficiency and the performance decreases with respect to the size and complexity of the datasets while lacking mechanisms to handle feedback or overlap.

Feedback-Star Paradigm wins over such methods on the F1 score, therefore being proved as a better clustering approach for dynamic and heterogeneous social networks. Thus, the result affirms its potential to become a powerful tool in social network analysis of the present-times.

Chapter 4

Expected Outcome

Through the Clustering Process facilitated by the Feedback based Star Approach, the algorithm identifies key relationships and organizes the nodes into distinct clusters. The final Output Clustered Graph demonstrates these meaningful groupings, with enhanced intra-cluster cohesion and inter-cluster separation.

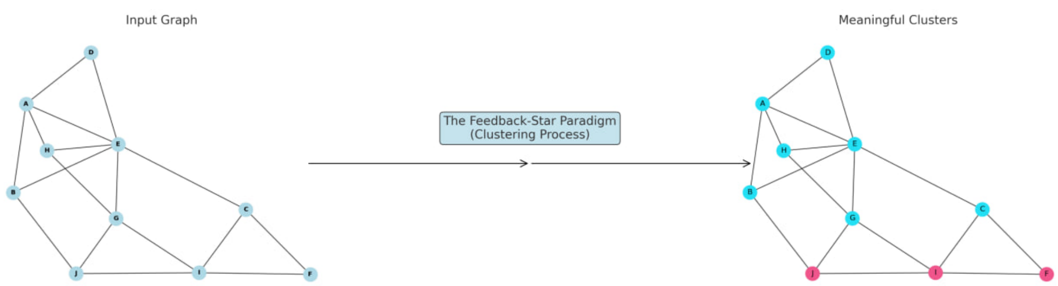

Figure 4.1: Meaningful clusters using the Feedback-Star Paradigm.

The Feedback-Star Paradigm proposes a cluster-finding approach of social networks which is adaptive, sample-driven and takes feedback for adjusting thresholds as well, Which greatly aids in identifying the changes in the networks over time and in finding the overlapping clusters. Its capacity, especially in regards to F1 performance metrics, is greater than the traditional means by combining feedback and iterative threshold modification, hence ensuring a good quality final listing of the clusters with optimized cost effectiveness.

The paradigm introduces a feedback-driven mechanism that allows for iterative refinement of clusters. This adaptability ensures that the clustering process evolves with network changes, addressing dynamic relationships and overlapping clusters effectively.

References

[1] J. Aslam, K. Pelekhov, and D. Rus, ”Using Star Clusters for Filtering,” in *Proceedings of the Conference on Information and Knowledge Management*, Hanover, NH, USA, Nov. 2000.

[2] J. A. Aslam, E. Pelekhov, and D. Rus, ”The Star Clustering Algorithm for Static and Dynamic Information Organization,” *Journal of Graph Algorithms and Applications*, vol. 8, no. 1, pp. 95–129, 2004.

[3] A. Pérez Suárez and J. E. Medina Pagola, ”A Clustering Algorithm Based on Generalized Stars,” *Journal of Artificial Intelligence Research and Development*, vol. 35, no. 2, pp. 125–145, 2023.

[4] J. A. Aslam, K. Pelekhov, and D. Rus, ”Static and Dynamic Information Organization with Star Clusters,” in *Proceedings of the ACM SIGIR Conference*, Oct. 1998, pp. 308–316.

[5] R. J. Gil-García, J. M. Badía-Contelles, and A. Pons-Porrata, ”Extended Star Clustering Algorithm,” in *CIARP 2003: Progress in Pattern Recognition, Image Analysis, and Applications*, Lecture Notes in Computer Science, vol. 2905, Springer-Verlag Berlin Heidelberg, 2003, pp. 480–487.

[6] J.-P. Mei and L. Chen, ”Fuzzy Clustering Approach for Star-Structured Multi-Type Relational Data,” in *2011 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, Taipei, Taiwan, 2011, pp. 2484–2491.

[7] R. Gil-García and A. Pons-Porrata, ”Hierarchical Star Clustering Algorithm for Dynamic Document Collections,” in *Proceedings of the 13th Iberoamerican Congress on Pattern Recognition (CIARP 2008)*, Lecture Notes in Computer Science, vol. 5197, Springer-Verlag, 2008, pp. 187–194.

[8] J. A. Aslam, E. Pelekhov, and D. Rus, ”The Star Clustering Algorithm for Information Organization,” in *Clustering and Information Organization*, Springer, pp. 1–33, 2003.

[9] P. Berkhin, ”Survey of Clustering Data Mining Techniques,” Accrue Software, Inc. Technical Report, 2002.

[10] J.-P. Mei, H. Lv, L. Yang, and Y. Li, ”Clustering for Heterogeneous Information Networks with Extended Star-Structure,” *Data Mining and Knowledge Discovery*, vol. 33, no. 4, pp. 1059–1087, 2019.

[11] S. Schmeja, ”Identifying Star Clusters in a Field: A Comparison of Different Algorithms,” *Astronomische Nachrichten*, 2018.

[12] J. Hua, J. Yu, and M.-S. Yang, ”Star-Based Learning Correlation Clustering,” *Pattern Recognition*, vol. 116, p. 107966, 2021.

[13] D. Puschmann, P. Barnaghi, and R. Tafazolli, ”Adaptive Clustering for Dynamic IoT Data Streams,” *IEEE Transactions on Emerging Topics in Computing*, 2016.

[14] A. Pister, P. Buono, J.-D. Fekete, C. Plaisant, and P. Valdivia, ”Integrating Prior Knowledge in Mixed-Initiative Social Network Clustering,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 27, no. 2, pp. 1776–1790, 2021.

[15] M.-F. Balcan and A. Blum, ”Clustering with Interactive Feedback,” in *Proceedings of the 19th International Conference on Algorithmic Learning Theory (ALT 2008)*, Lecture Notes in Artificial Intelligence (LNAI), vol. 5254, Springer-Verlag, 2008, pp. 316–328.

[16] N. P. Nguyen, T. N. Dinh, Y. Xuan, and M. T. Thai, ”Adaptive Algorithms for Detecting Community Structure in Dynamic Social Networks,” in *Proceedings of the IEEE INFOCOM 2011*, Shanghai, China, 2011, pp. 2282–2290.

[17] N. Mishra, R. Schreiber, I. Stanton, and R. E. Tarjan, ”Clustering Social Networks,” *Lecture Notes in Computer Science*, vol. 4863, pp. 56–72, 2007.

[18] F. Liu, J. Wu, S. Xue, C. Zhou, J. Yang, and Q. Sheng, ”Detecting the Evolving Community Structure in Dynamic Social Networks,” *World Wide Web*, vol. 23, no. 5, pp. 2389–2420, 2019.

[19] J. Bae, T. Helldin, M. Riveiro, S. Nowaczyk, M.-R. Bouguelia, and G. Falkman, ”Interactive Clustering: A Comprehensive Review,” *ACM Computing Surveys*, vol. 53, no. 1, pp. 1–39, 2020.

iThenticate

Document Viewer

Research project proposal

By: 202165, 202166

As of: Jan 16, 2025 2:49:16 AM

4,437 words - 0 matches - 0 sources

Similarity Index

0%

Mode: Similarity Report Exclude Quotes Include Bibliography Excluding matches < 1% Limiting match size to 9 words

The Feedback-Star Paradigm: Redefining Social Network Clustering by Examination Roll 202165 202166 Session: 2019-2020 Bachelor of Science in Computer Science and Engineering Faculty of Mathematical and Physical Sciences Jahangirnagar University Dhaka-1342, Bangladesh. January 15, 2025 Contents Abstract 1 Introduction 2 Literature Review 3 Proposed Methodology 3.1 Problem Statement . . . . 3.2 Workflow . . . . 3.3 Comparative Evaluation of Clustering Techniques . . . . 3.3.1 Performance Comparison . . . . 3.3.2 Advantages of the Feedback-Star Paradigm . . . . 3.3.3 Limitations of Traditional Approaches . . . . 4 Expected Outcome References 1 2 4 9 9 10 12 12 12 13 14 15 ii Abstract This paper presents a novel approach to the implementation of the Star Clustering Al- gorithm for social network analysis by including user-initiated feedback mechanism for the betterment of cluster formation in a non-static enterprise setting. The proposed Feedback-Star Paradigm is a useful system that enhances user-centered innovation by rectifying the key problems of the clustering methods. Our algorithm identifies star centers based on node degree centrality and employs a deterministic tie-breaking mech- anism to ensure reproducible clustering outcomes. The distinguishing feature of the method lies in its ability to iteratively adjust the clustering thresholds,

I hereby declare that the report confirms the originality of the work that complies with the institution's academic integrity guidelines.

16.01.2025

Prof. Dr. Sanjit Kumar Saha

Supervisor