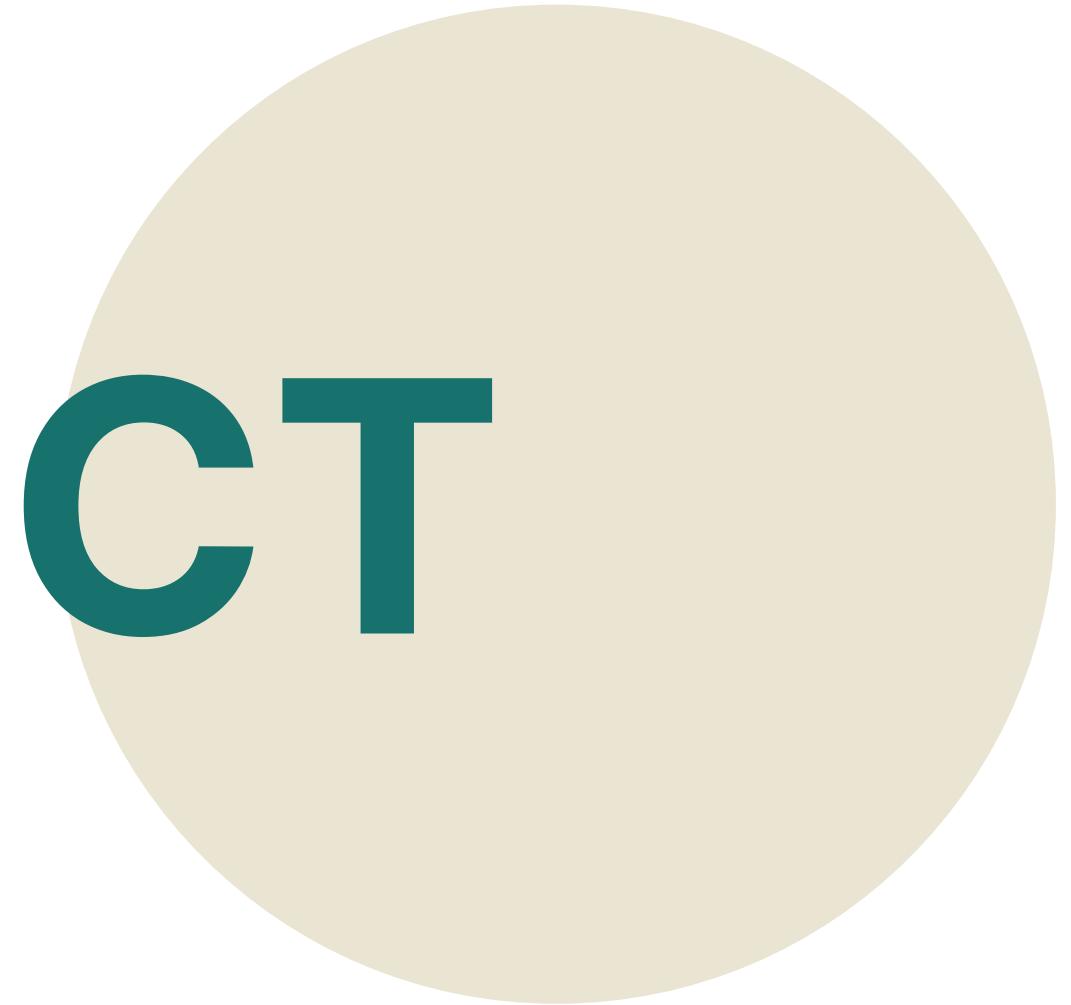


# MINI PROJECT

● PROJECT IDEA PRESENTATION



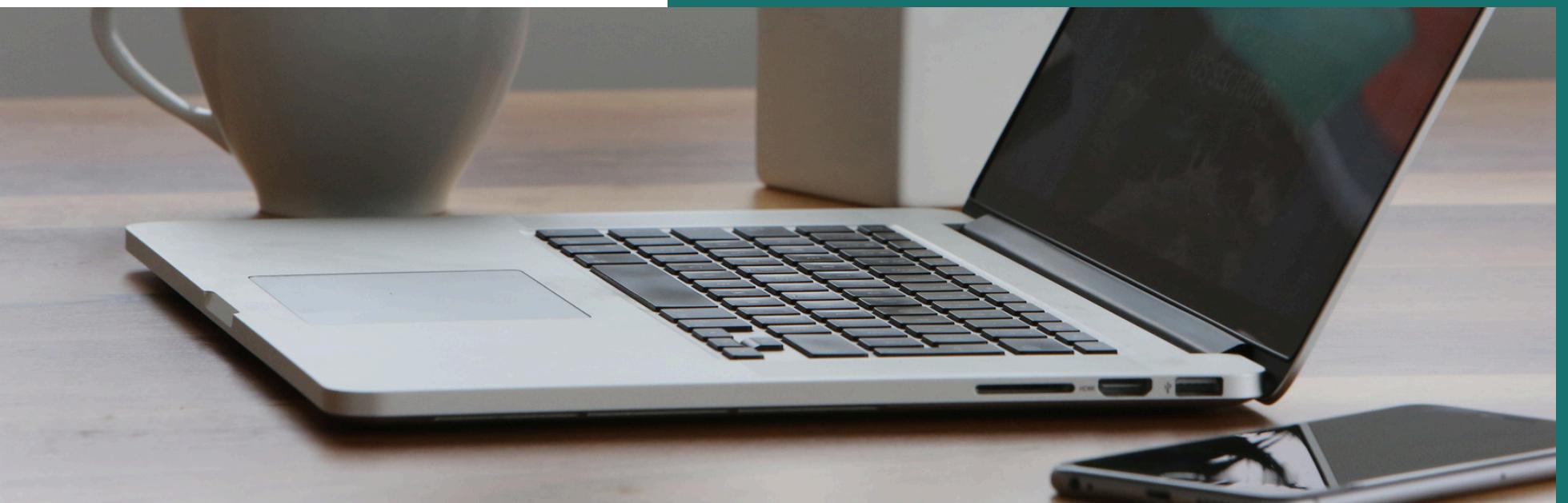
---

**Project Guide:** Dr.R.Malavika

AUGUST 2024

# TABLE OF CONTENT

---



## PROJECT IDEA PRESENTATION

**01** Project Team

**02** Problem Statement

**03** Existing Systems

**04** Proposed System

**05** Algorithm Flow

**06** Thank You!



# TEAM MEMBERS

## GREETING FROM US

---

- Jeyakeerthana M-71772118118
- Kalvikarasi R-71772118119
- Merlin Deborah S-71772118127
- Priyadharshini K-71772118134

DOMAIN:SOCIAL NETWORKS

GUIDED BY DR.R.MALAVIKA

# PROBLEM STATEMENT:

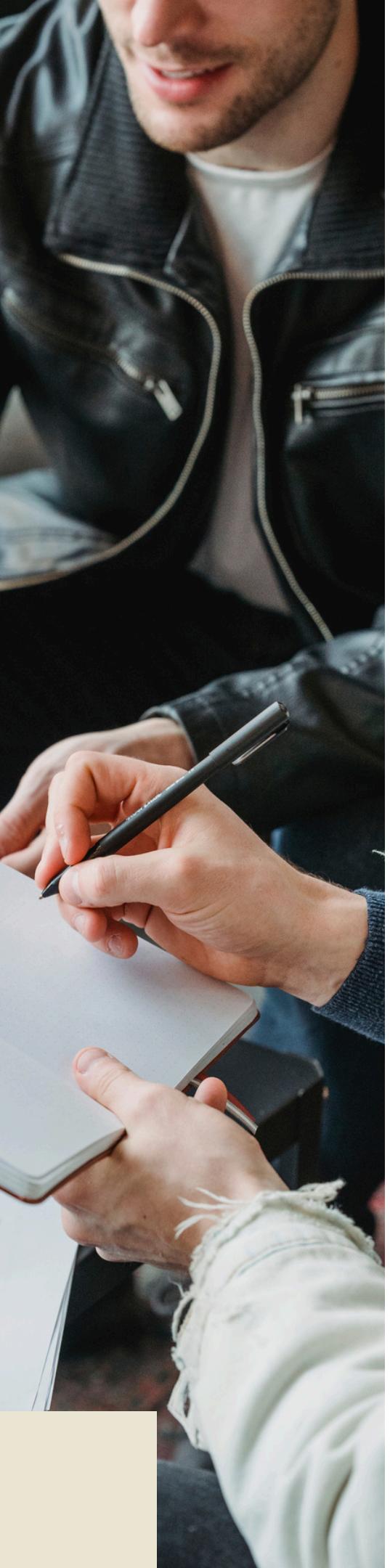
## **STATEMENT:**

Community Detection in Social Networks Using a Local Approach Based on Node Ranking

## **DESCRIPTION:**

Detecting communities in social networks is a complex, NP-hard problem. Existing algorithms either suffer from high computational costs or lack accuracy and determinism. There's a need for a more efficient and precise local algorithm that doesn't rely on initial seed nodes.



A vertical photograph on the left side of the slide shows a close-up of a man's hands. He is wearing a black leather jacket over a white t-shirt. He is holding a black pen and writing in a white notebook. His left hand is resting on the notebook. The background is dark and out of focus.

# EXISTING COMMUNITY DETECTION SYSTEMS:

## Global Algorithms:

- High accuracy but not scalable for large networks.
- Examples:
  - Graph Partitioning: Divides the network, minimizing inter-community edges.
  - Girvan-Newman: Removes high betweenness edges to identify communities.
  - Louvain Method: Optimizes modularity through iterative node merging.

## Local Algorithms:

- Efficient and scalable, but less stable.
- Examples:
  - Label Propagation (LPA): Fast, but can be uncertain.
  - WLPA: Enhances LPA by considering communication intensity.
  - RTLCD: Tackles seed-dependency issues and core validity.

## Challenges:

- Global: Not feasible for large networks due to high time complexity.
- Local: Dependence on predefined parameters, instability, lower accuracy.

# LITERATURE SURVEY:

PAPER NAME	YEAR	AUTHOR	DESCRIPTION	ADVANTAGE	DISADVANTAGE	COMPARISON WITH LCD-SN
Community Detection in Complex Networks Using Genetic Algorithms	2008	Mursel Tasgin, Amac Herdagdelen, and Haluk Bingol	This paper presents a genetic algorithm for community detection in complex networks using modularity Q as the fitness function. The algorithm employs an initial population setup, crossover, and mutation operations. Parameters like mutation rate, population size, and number of generations are tuned through trial and error.	<ul style="list-style-type: none"> <li>- Flexible approach with the ability to explore a wide range of potential solutions.</li> <li>- Allows automatic partitioning of communities through evolutionary mechanisms.</li> </ul>	<ul style="list-style-type: none"> <li>- Requires extensive parameter tuning (mutation rate, crossover rate, population size), which is done arbitrarily in this paper.</li> <li>- Time-consuming due to repeated generations and the iterative process.</li> </ul>	<ul style="list-style-type: none"> <li>- LCD-SN is faster with linear time complexity, making it better suited for large networks, whereas the genetic algorithm can be computationally expensive.</li> <li>- LCD-SN doesn't require arbitrary parameter tuning, unlike the genetic algorithm which requires parameters such as mutation and crossover rates</li> </ul>

# LITERATURE SURVEY:

PAPER NAME	YEAR	AUTHOR	DESCRIPTION	ADVANTAGE	DISADVANTAGE	COMPARISON WITH LCD-SN
LabelRank: A Stabilized Label Propagation Algorithm for Community Detection in Networks	2013	Jierui Xie; Boleslaw K. Szymanski	Proposes LabelRank, an efficient algorithm that improves upon traditional Label Propagation Algorithm (LPA) by introducing operators to control and stabilize the propagation process. The goal is to reduce the randomness and improve community detection in large-scale networks.	<ul style="list-style-type: none"><li>- Stabilizes label propagation to ensure consistent results in all runs.</li><li>- Efficient for large-scale networks.</li><li>- Reduces the randomness associated with traditional LPA.</li></ul>	<ul style="list-style-type: none"><li>- Still relies on label propagation, which may struggle with overlapping communities.</li><li>- Quality of results is dependent on the initial label assignment.</li><li>- Does not address the resolution limit issue.</li></ul>	<ul style="list-style-type: none"><li>- LCD-SN uses a more deterministic node ranking process, providing stable and consistent community detection without the need for stabilizing operators.</li><li>- LCD-SN resolves overlapping nodes more effectively, while LabelRank may struggle in networks with overlapping structures.</li></ul>

# LITERATURE SURVEY:

PAPER NAME	YEAR	AUTHOR	DESCRIPTION	ADVANTAGE	DISADVANTAGE	COMPARISON WITH LCD-SN
An Improved Community Detection Algorithm Based on Local Information in Social Networks	2013	Xia Lei; Zhang Lejun; Zhang Jianpei; Yang Jing; Guo Lin	The LPA-SNA algorithm is an enhanced version of the Label Propagation Algorithm (LPA), incorporating similarity of nodes' attributes to improve community detection and overcome the random partitioning issue in standard LPA.	<ul style="list-style-type: none"><li>- Enhances the traditional LPA by using node attribute similarity, improving the quality of detected communities.</li><li>- Reduces the randomness associated with LPA, leading to more stable results.</li></ul>	<ul style="list-style-type: none"><li>- Heavily reliant on node attribute similarity, which may not always be available or relevant in all types of networks.</li><li>- Still inherits some drawbacks of LPA, such as difficulty handling overlapping communities.</li></ul>	<ul style="list-style-type: none"><li>- LCD-SN does not rely on node attributes, making it more versatile across different types of networks, especially when node attributes are unavailable or incomplete.</li><li>- LCD-SN handles overlapping communities more effectively through structured merging processes, while LPA-SNA may struggle with overlap due to its label propagation nature.</li></ul>

# LITERATURE SURVEY:

PAPER NAME	YEAR	AUTHOR	DESCRIPTION	ADVANTAGE	DISADVANTAGE	COMPARISON WITH LCD-SN
Overlapping community detection in social networks	2014	Zeineb Dhouioui; Jalel Akaichi	A method for overlapping community detection that uses the edge betweenness principle to identify communities in complex networks. It aims to address the challenge of high complexity in detecting overlapping communities as network size increases.	<ul style="list-style-type: none"> <li>- Effectively detects overlapping communities.</li> <li>- Utilizes edge betweenness, a well-known and intuitive measure for community detection.</li> <li>- Suitable for small- to medium-sized networks.</li> </ul>	<ul style="list-style-type: none"> <li>- High computational complexity as network size increases.</li> <li>- Edge betweenness-based methods are less efficient for large networks due to the costly calculation of betweenness centrality.</li> </ul>	<ul style="list-style-type: none"> <li>- LCD-SN uses a deterministic ranking of nodes based on importance and similarity, making it more scalable and efficient in large networks, whereas the edge betweenness method struggles with scalability.</li> <li>- LCD-SN's approach to handling overlapping nodes is more structured and avoids the high computational cost of calculating betweenness centrality, making it more efficient for large-scale networks.</li> </ul>

# LITERATURE SURVEY:

PAPER NAME	YEAR	AUTHOR	DESCRIPTION	ADVANTAGE	DISADVANTAGE	COMPARISON WITH LCD-SN
Overlapping Local Community Detection in Directed Weighted Networks	2015	Shidong Li; Sheng Ge	A distributed algorithm for detecting overlapping local communities in directed, weighted networks. It defines Local Community Modularity and Tightness to guide the process. The algorithm has two phases: initial local community detection and merging of similar communities based on tightness.	<ul style="list-style-type: none"><li>- Tailored for directed, weighted networks, providing a more accurate reflection of real-world network dynamics.</li><li>- Uses local modularity and tightness metrics to improve the quality of detected communities.</li></ul>	<ul style="list-style-type: none"><li>- The algorithm's effectiveness depends on how well local modularity and tightness are defined, which may not generalize well to all network types.</li><li>- May struggle with scalability in very large networks or highly overlapping structures.</li></ul>	<ul style="list-style-type: none"><li>- LCD-SN uses a deterministic node importance and ranking approach, which ensures more consistent and reliable results compared to the modularity and tightness metrics used in this method.</li><li>- LCD-SN handles overlapping nodes through a more structured merging process, whereas this algorithm relies heavily on the tightness metric, which may not always capture nuanced community structures.</li></ul>

# LITERATURE SURVEY:

PAPER NAME	YEAR	AUTHOR	DESCRIPTION	ADVANTAGE	DISADVANTAGE	COMPARISON WITH LCD-SN
Local community detection algorithm based on links and content	2017	Cuijuan Wang; We nzhong Tang; Yan yang Wang; Jing Fang; Sha n Yao	A local community detection method based on links and content, focusing on specific users' communities rather than the global network. Tested on the Enron email dataset, it aims to improve accuracy despite missing links or low communication frequency.	<ul style="list-style-type: none"> <li>- Integrates both link density and content information, leading to more accurate community detection.</li> <li>- Effective for networks where connections alone don't represent relationships.</li> </ul>	<ul style="list-style-type: none"> <li>- May not perform well if content information is sparse or unreliable.</li> <li>- Content-based analysis can introduce complexity and dependency on data quality.</li> <li>- Focused on specific users, so not optimized for global detection.</li> </ul>	<ul style="list-style-type: none"> <li>- LCD-SN uses a deterministic node importance ranking based on network structure alone, making it less dependent on content, which can be unreliable in sparse datasets.</li> <li>- LCD-SN is better suited for networks without rich content information, while the content-based approach in this algorithm may falter without strong communication patterns.</li> </ul>

# LITERATURE SURVEY:

PAPER NAME	YEAR	AUTHOR	DESCRIPTION	ADVANTAGE	DISADVANTAGE	COMPARISON WITH LCD-SN
LPANNI: Overlapping Community Detection Using Label Propagation in Large-Scale Complex Networks	2018	Meilian Lu ; Zhenglin Zhang; Zhihe Qu; Yu Kang	LPANNI (Label Propagation Algorithm with Neighbor Node Influence) improves overlapping community detection by using a fixed label propagation sequence and label update strategies based on node importance and neighbor influence. It operates with linear time complexity.	<ul style="list-style-type: none"> <li>- Linear time complexity, suitable for large-scale networks.</li> <li>- Detects overlapping communities more effectively than traditional label propagation methods.</li> </ul>	<ul style="list-style-type: none"> <li>- May still face challenges with accuracy in networks with dense overlaps or where community boundaries are not well-defined.</li> <li>- Relies on the effectiveness of the neighbor influence and label preference strategies, which may not always capture the true network dynamics.</li> </ul>	<ul style="list-style-type: none"> <li>- LCD-SN handles overlapping nodes through a more deterministic process of resolving overlapping memberships, which is not as dependent on propagation sequence or neighbor influence, ensuring more consistent results.</li> <li>- Both algorithms are linear in time complexity, but LCD-SN may offer more flexibility and stability across different network types, while LPANNI excels in large-scale, highly overlapping networks.</li> </ul>

# LITERATURE SURVEY:

PAPER NAME	YEAR	AUTHOR	DESCRIPTION	ADVANTAGE	DISADVANTAGE	COMPARISON WITH LCD-SN
Community Detection Based on Node Influence and Similarity of Nodes	2022	Yanjie Xu; Tao Ren ; Shixiang Sun	An agglomerative community detection algorithm based on node influence and node similarity. It includes steps for identifying central nodes, expanding communities with similar nodes, and merging small communities. Evaluation is done using modularity and NMI.	<ul style="list-style-type: none"> <li>- Identifies central nodes, enhancing the detection of important communities.</li> <li>- Incorporates both node influence and similarity to form and merge communities.</li> <li>- Effective on real and synthetic networks.</li> <li>- Performs well in terms of modularity and NMI.</li> </ul>	<ul style="list-style-type: none"> <li>- The algorithm may struggle with highly overlapping communities due to the reliance on central nodes.</li> <li>- Computational complexity may increase in large-scale networks with many candidate neighbors and community merging steps.</li> </ul>	<ul style="list-style-type: none"> <li>- LCD-SN uses a more structured and deterministic approach for node importance, which ensures a consistent and reliable ranking without relying solely on central nodes.</li> <li>- LCD-SN has a more systematic process for handling overlapping communities through similarity measures and merging, whereas this algorithm may not handle overlaps as effectively.</li> </ul>

# LITERATURE SURVEY:

PAPER NAME	YEAR	AUTHOR	DESCRIPTION	ADVANTAGE	DISADVANTAGE	COMPARISON WITH LCD-SN
A Fast Local Balanced Label Diffusion Algorithm for Community Detection in Social Networks (LBLD)	2022	Hamid Roghani, Asgarali Bouyer	Proposes a local algorithm that uses balanced label diffusion to detect communities. The algorithm excludes nodes with degree 1 initially to improve efficiency and iteratively diffuses labels based on node similarity and importance. Nodes with higher similarity propagate their labels to neighbors.	- Efficient for large-scale networks due to exclusion of degree-1 nodes initially. - Balanced label diffusion ensures faster convergence and reduces randomness.	- Struggles with overlapping community detection. - The quality of results heavily depends on the node similarity metric, which may not work well for all network types.	- LCD-SN handles overlapping nodes more effectively, while LBLD struggles with overlap due to its reliance on label diffusion. - LCD-SN does not rely on excluding nodes with degree 1, ensuring a more comprehensive analysis of sparse networks, while LBLD may lose critical information in these networks.

# LITERATURE SURVEY:

PAPER NAME	YEAR	AUTHOR	DESCRIPTION	ADVANTAGE	DISADVANTAGE	COMPARISON WITH LCD-SN
Dynamic Local Community Detection Algorithms	2022	Sahar Bakhtar; Hovhannes A. Harutyunyan	The paper investigates local community detection algorithms in dynamic social networks, where complete network information is often unavailable. Algorithm P is found to outperform others in terms of speed within dynamic structures	<ul style="list-style-type: none"><li>- Handles dynamic networks, tracking community changes over time.</li><li>- Algorithm P offers significantly faster community detection compared to other local algorithms.</li><li>- Suitable for large networks where complete network information is unavailable.</li></ul>	<ul style="list-style-type: none"><li>- Focuses on local detection, which may not be as effective for global community structures.</li><li>- Effectiveness may vary depending on the network dynamics.</li><li>- May struggle with overlapping community structures.</li></ul>	<ul style="list-style-type: none"><li>- LCD-SN is designed for static networks, whereas this paper focuses on dynamic networks. However, LCD-SN provides a more structured and reliable approach for handling overlapping and weak communities.</li><li>- LCD-SN's deterministic nature ensures more consistent and stable results, while dynamic algorithms may face challenges in tracking rapidly evolving communities</li></ul>

# THE PROPOSED ALGORITHM: LCD-SN:

## **Introduction:**

The algorithm introduces an IMP index to assess node importance based on their neighbors, inspired by electric current continuity in circuits. This index drives the LCD-SN algorithm for community detection in social networks.

## **IMP Index:**

- Concept: Node importance is calculated using both first- and second-degree neighbors.

## **LCD-SN Algorithm Phases:**

### 1. Initial Community Formation:

- Rank nodes with the IMP index.
- Form communities by grouping the highest-ranked node with its neighbors.

### 2. Overlapping Node Resolution:

- Assign overlapping nodes to the community with the highest similarity using the GLHN index.

### 3. Community Integration:

- Merge small and weak communities with larger, stronger ones based on similarity.

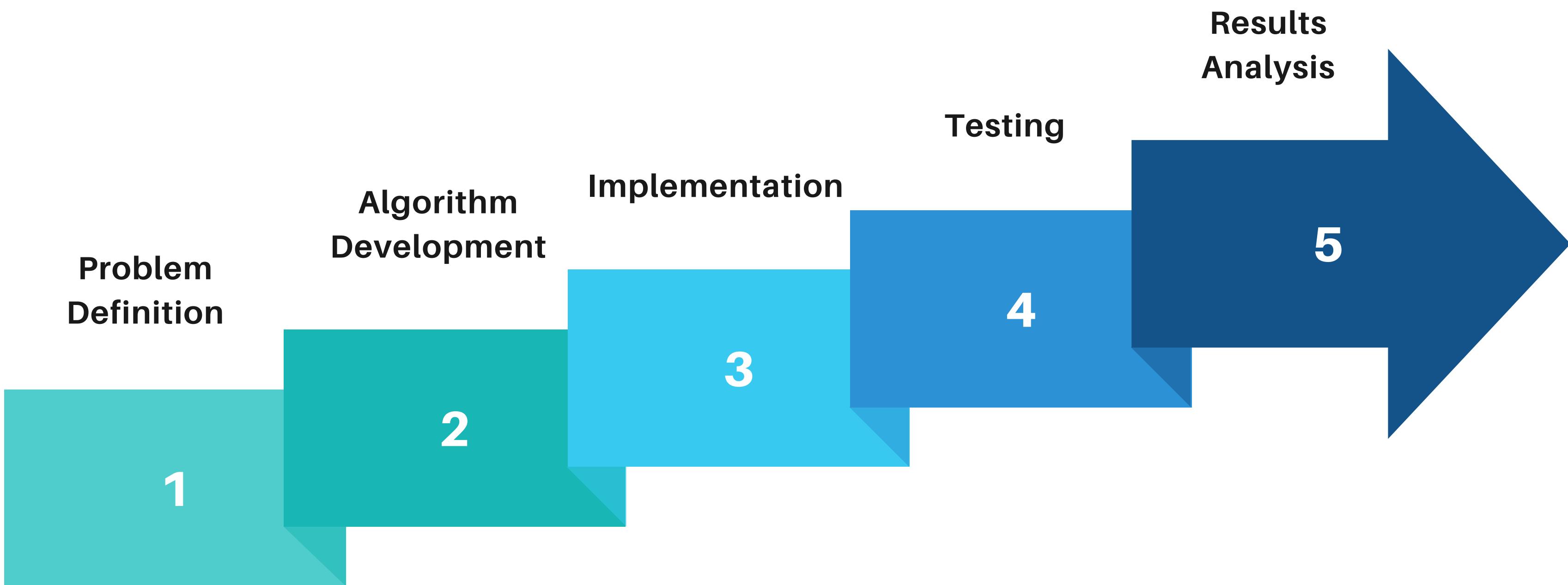
## **Final Output:**

The algorithm produces distinct, well-defined communities within the network.

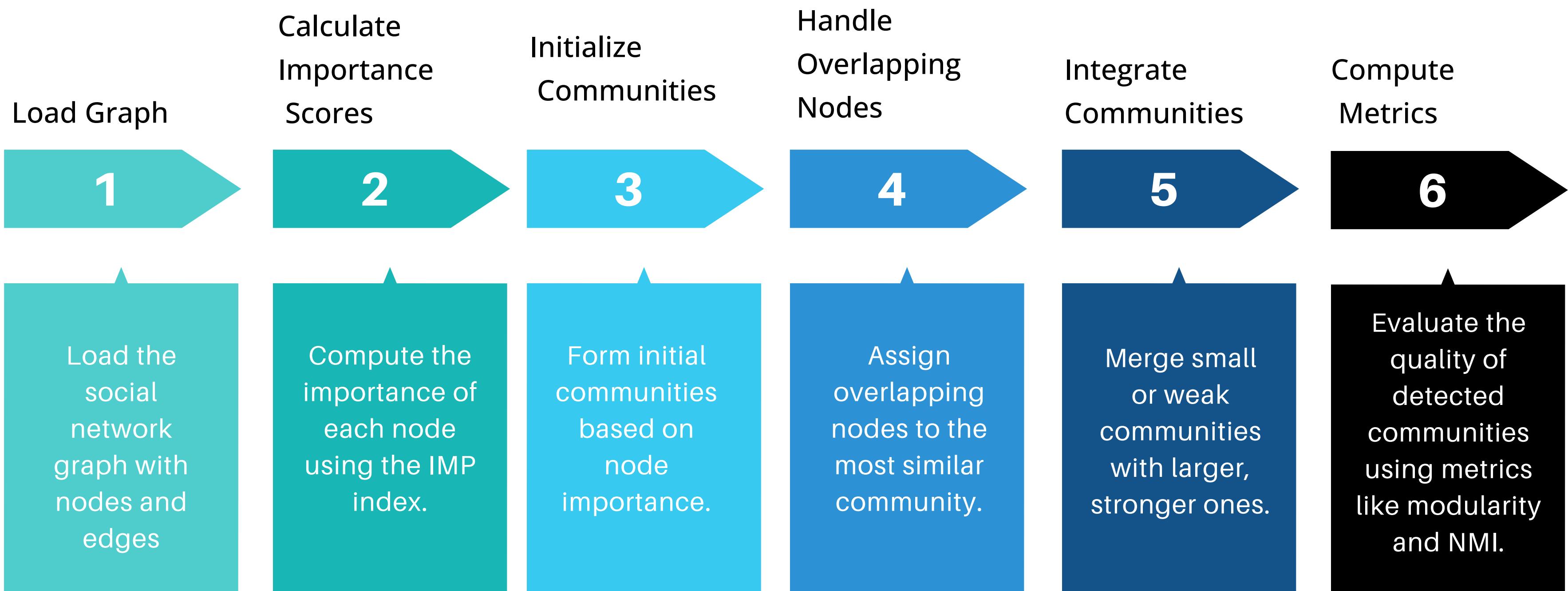


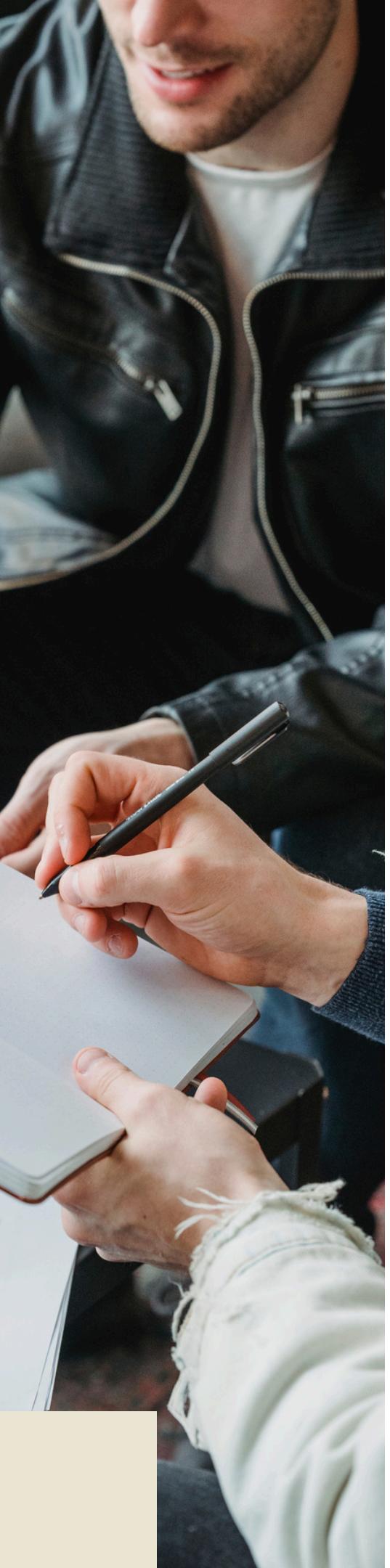
# LOCAL COMMUNITY DETECTION - SOCIAL NETWORK

## PROJECT FLOW



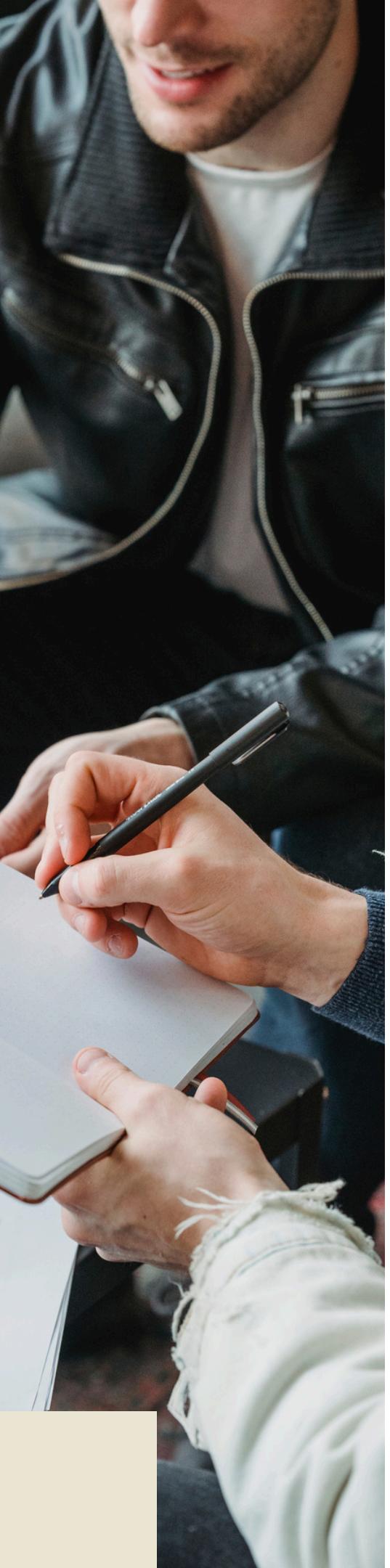
# LCD-SN ALGORITHM FLOWCHART





# REFERENCES:

1. Mursel Tasgin, Amac Herdagdelen, Haluk Bingol - 2008 "Community Detection in Complex Networks Using Genetic Algorithms".
2. Jierui Xie; Boleslaw K. Szymanski - 2013 "LabelRank: A Stabilized Label Propagation Algorithm for Community Detection in Networks".
3. Xia Lei; Zhang Lejun; Zhang Jianpei; Yang Jing; Guo Lin - 2013 "An Improved Community Detection Algorithm Based on Local Information in Social Networks".
4. Zeineb Dhouioui; Jalel Akaichi - 2014 "Overlapping community detection in social networks".
5. Shidong Li; Sheng Ge - 2015 "Overlapping Local Community Detection in Directed Weighted Networks".



# REFERENCES:

- 6.Cuijuan Wang; Wenzhong Tang; Yanyang Wang; Jing Fang; Shan Yao - 2017 "Local community detection algorithm based on links and content".
- 7.Meilian Lu ; Zhenglin Zhang; Zhihe Qu; Yu Kang - 2018 "LPANNI: Overlapping Community Detection Using Label Propagation in Large-Scale Complex Networks".
- 8.Yanjie Xu;Tao Ren ; Shixiang Sun - 2022 "Community Detection Based on Node Influence and Similarity of Nodes".
- 9.Hamid Roghani, Asgarali Bouyer - 2022 "A Fast Local Balanced Label Diffusion Algorithm for Community Detection in Social Networks (LBLD)".
- 10.Sahar Bakhtar; Hovhannes A. Harutyunyan - 2022 "Dynamic Local Community Detection Algorithms".

# THANK YOU

● FOR YOUR KIND ATTENTION