GSO-DPCD: A Fuzzy Density-Peak Approach for Overlapping Community Detection

Swetha Balasubramanian
Department of Computer Science and Engineering
South Asian University
New Delhi, India
swethab@students.sau.ac.in

Pranab K. Muhuri
Department of Computer Science and Engineering
South Asian University
New Delhi, India
pranabmuhuri@cs.sau.ac.in

***Abstract*—Overlapping community detection in large-scale networks faces a constant trade-off between structural accuracy and computational complexity. Although traditional Density Peak Clustering (DPC) presents an intuitive mechanism for identifying community centers (density peaks), its applicability is limited due to $O(N^2)$ complexity. In this paper, we propose a Gravity-driven Structural Overlapping Density Peak-based community detection (GSO-DPCD) that models community detection as a fluid-dynamic process within a Structural Gravity Field. Using a structural compactness measure derived from triadic closure, we detect density peaks that are topologically central in their communities. Further, we present a Heuristic Path Approximation that overcomes the quadratic computational bottleneck, achieving near-linear $O(N \log N)$ complexity. Extensive experiments on LFR benchmarks and large-scale real-world networks show that GSO-DPCD consistently outperforms existing methods in terms of modularity and accuracy, offering a robust and efficient framework for uncovering hidden structures in highly complex networks with deep overlaps.**

***Index Terms*—Overlapping Community Detection, Density Peak clustering, Structural Gravity Field, Gravitational Membership Propagation, Fuzzy Clustering**

I. INTRODUCTION

The rapid growth of large-scale interconnected systems, from protein-protein interactions to social platforms, has redefined the requirements of community detection. In these complex domains, crisp community boundaries fail to reflect the true underlying structure. When real-world data are modelled as networks, it forms overlapping communities in which a node can belong to multiple communities simultaneously [1]. To model this "fuzziness" of membership, we need to shift from "hard" partitioning to "soft" partitioning [2]. The application of fuzzy logic enables a mathematically rigorous representation of node membership across overlapping communities [3]. Although many fuzzy-based approaches have been extended to network structures like Fuzzy C-Means [4] (FCM) and Possibilistic C-Means [5], their practical applicability remains limited as they require a predefined number of communities and are sensitive to this initialization. Numerous methods were proposed to address these gaps [6], [7], but they were heavily dependent on local heuristics and sensitive to structural bias. Recent research has focused on Evolutionary Fuzzy Clustering and Swarm-based Optimization, which aim to bridge this gap but are computationally complex, making them unsuitable for large-scale networks.

The Density Peak Clustering (DPC) is an intuitive framework that identifies density peaks as high-density nodes which are at a large distance from other high-density nodes [8]. The DPC algorithm transformed center selection by combining local density (ρ) and minimum distance (δ). Many spatial and fuzzy-based extensions have been proposed to enhance DPC. Spatial variants like DPC-KNN [9] use kNN to stabilize density estimates while fuzzy variants like FKNN-DPC [10] integrate fuzzy set theory to accommodate uncertainty in data. Although DPC-based approaches are effective, they are limited by their inability to capture the true overlapping nature of real-world networks [11]. They are also bottlenecked by the $O(n^2)$ computation required for the all-pairs distance matrix. Further, they also suffer from the domino effect: if a node is assigned a wrong community label, its neighbours with lower density will also be assigned incorrectly. These limitations hinder its applicability to large networks.

The existing literature reveals a fundamental tradeoff in community detection between Scalability and structural resolution. Methods like Louvain [12] and Leiden [13] are highly scalable yet disjoint. Methods like NMF/Fuzzy-DPC, understand the network but are computationally expensive. The proposed framework, Gravity-driven Structural Overlapping Density Peak-based community Detection, GSO-DPCD, bridges this gap, overcoming both the Euclidean limitation of DPC-based approaches and the high iterative overhead of fuzzy frameworks. GSO-DPCD models nodes as objects in a topological potential field guided by a Structural Gravity Paradigm. It leverages Structural similarity to quantify Topological closeness. Density is defined using structural compactness, computed from triadic closure and local cohesion. This ensures that the peaks are structurally central nodes and not just high-degree nodes. Furthermore, utilizing gravitational membership propagation, it replaces the iterative optimization mechanism with a near-linear mechanism that enables deep fuzzy accuracy at $O(N \log N)$ scale.

Our key contribution is the Gravitational Membership Propagation strategy. Drawing inspiration from the principles of fluid dynamics, we model community influence as a directed flow originating from high community peaks down a topological potential gradient towards the community boundaries.

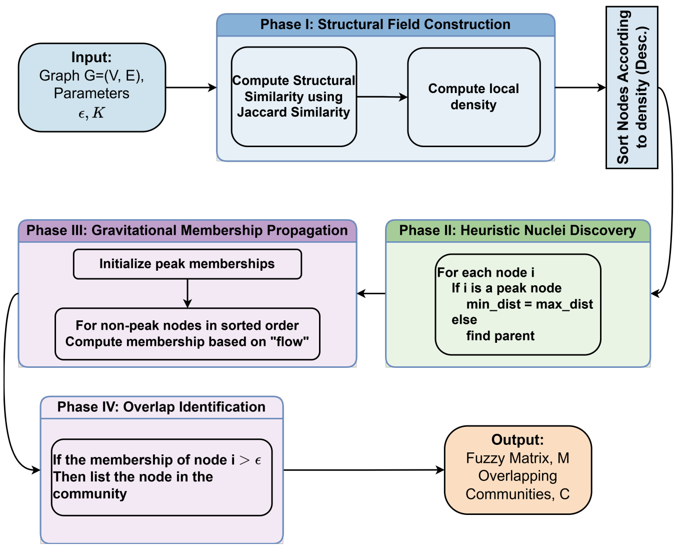

Fig. 1. Flowchart of the proposed algorithm.

In this approach, membership is treated as a recursive fuzzy distribution rather than a discrete label. As the influence propagates, the nodes in the "valleys" between the peaks naturally integrate signals from multiple sources, acquiring a fuzzy membership vector that captures the degree of structural overlap. By incorporating a Heuristic Path Approximation for center identification, we overcome the quadratic complexity bottleneck of DPC, thus, achieving a near-linear complexity of $O(N \log N)$. This enables GSO-DPCD to handle large-scale networks, surpassing the scalability limits of conventional DPC and fuzzy algorithms. Our results demonstrate that the depth of fuzzy interpretability and large-scale computational efficiency can be integrated in an effective framework for community detection.

The remainder of the paper is organized as follows: Section II details the methodology of the proposed framework. Section III presents the experimental setup, followed by an analysis of the results. Finally, section IV concludes the paper.

II. Proposed Approach

The design of GSO-DPCD re-conceptualizes community detection as a dynamic flow within a structural gravity field [14]. The methodology proceeds through three fundamental stages: 1) The formation of density through structural "mass", 2) The identification of density peaks, and 3) Gravitational propagation of fuzzy membership vectors across the networks. Fig. 1 illustrates the flow of the proposed approach. We will discuss the framework in detail in this section.

A. Constructing the Structural Gravity Field

In topological graphs, traditional Euclidean distance is inadequate to reflect the relational strength between the nodes. To reflect the true topology of the networks, we define the "mass" of a node through structural compactness, employing a weighted similarity metric $\mathcal{S}(i, j)$ derived from the Jaccard coefficient of shared neighbourhoods. For any pair of pair of connected nodes i and j, the similarity is:

$$\mathcal{S}(i, j) = \frac{|N(i) \cap N(j)|}{|N(i) \cup N(j)|} \quad (1)$$

Where $N(i)$ represents the set of neighbors of node i. This metric captures Triadic Closure. From this "mass," we derive the Local Density (ρ_i) as the sum of structural ties:

$$\rho_i = \sum_{j \in N(i)} \mathcal{S}(i, j) \quad (2)$$

Where we define ρ_i as the "height" of the terrain at node i. High-density nodes occupy the peaks of the cohesive structural cores, while low-density nodes are found at the sparse "valleys" between communities.

Algorithm 1: GSO-DPCD: Gravity-driven Structural Overlapping Density Peak Community Detection

Input: Graph $G = (V, E)$, Overlap Threshold ϵ, Number of Peaks K
Output: Fuzzy Membership Matrix $M \in \mathbb{R}^{n \times K}$, Community Set $\mathcal{C}$
// Phase I: Structural Field Construction
foreach *edge* $(u, v) \in E$ **do**
  $W_{uv} \leftarrow \frac{|N(u) \cap N(v)|}{|N(u) \cup N(v)|}$ // Structural Jaccard Similarity
foreach *node* $u \in V$ **do**
  $\rho_u \leftarrow \sum_{v \in N(u)} W_{uv}$ // Calculate Local Density
// Phase II: Heuristic Nuclei Discovery
$V_{sorted} \leftarrow$ Sort nodes in V by ρ in descending order;
foreach *node* $u \in V_{sorted}$ **do**
  $\mathcal{N}^+(u) \leftarrow \{v \in N(u) \mid \rho_v > \rho_u\}$
  //Topological Parent Search
  if $\mathcal{N}^+(u)$ *is not empty* **then**
    $\delta_u \leftarrow \min_{v \in \mathcal{N}^+(u)}(1 - W_{uv})$;
    $parent(u) \leftarrow \arg\min_{v \in \mathcal{N}^+(u)}(1 - W_{uv})$;
  else
    $\delta_u \leftarrow \max(\delta)$ //Potential Structural Peak
// Phase III: Gravitational Membership Propagation
Identify top K peaks using $\gamma_u = \rho_u \times \delta_u$;
Initialize M as an $n \times K$ zero matrix;
for $k \leftarrow 1$ **to** K **do**
  $M_{peak_k, k} \leftarrow 1.0$;
foreach *node* $u \in V_{sorted}$ *not in peaks* **do**
  $\mathcal{N}^+(u) \leftarrow \{v \in N(u) \mid \rho_v > \rho_u\}$;
  if $\mathcal{N}^+(u)$ *is not empty* **then**
    $M_u \leftarrow \frac{\sum_{v \in \mathcal{N}^+(u)}(W_{uv} \cdot M_v)}{\sum_{v \in N^+(u)} W_{uv}}$ // Recursive Fuzzy Flow
// Phase IV: Overlap Extraction
foreach *node* $u \in V$ **do**
  Assign u to community C_k **if** $M_{u,k} > \epsilon$;
return $M, \mathcal{C}$;

B. "Peak" selection

In this phase, we will identify the community peaks. In traditional DPC, determining δ requires a global search for the nearest neighbor of higher density, resulting in quadratic complexity. GSO-DPCD employs a Heuristic Path Approximation. For every node i, we limit the search for a "parent" (a node with $\rho_j > \rho_i$) to its immediate topological neighbors. This leads us to two cases: 1) If a neighbor j exists such that $\rho_j > \rho_i$, we define the distance δ_i as the inverse of their structural similarity: $\delta_i = 1 - \mathcal{S}(i, j)$. 2) If no such neighbor exists, it indicates that the node i is a local maximum, a Structural "Peak". By limiting our search to a node's topological neighbors, we reduce the computational complexity from quadratic $O(N^2)$ to linear $O(m)$. Then, we construct a decision graph based on the combined metric

$$\gamma_i = \rho_i \times \delta_i \tag{3}$$

The "spikes" indicate the K structural peaks of the network's communities, detected using a nonparametric elbow-detection strategy.

C. Gravitational Membership Propagation

This phase is the core of GSO-DPCD, where we shift from crisp "peaks" to Fuzzy "Flow". Once we identified the K peaks, they are used as the source of community influence. For every node i we define a membership vector $M_i = [m_{i,1}, m_{i,2}, \ldots, m_{i,K}]$. The identified peaks are initialized with absolute membership (e.g., $m_{k,k} = 1.0$), after which the remaining nodes are iteratively updated in decreasing order of density. This "flow" of membership resembles water flowing down a mountain. The influence of the peaks propagates through the network using a recursive fuzzy relation:

$$M_i = \frac{\sum_{j \in \mathcal{N}^+(i)} (\mathcal{S}(i, j) \cdot M_j)}{\sum_{j \in \mathcal{N}^+(i)} \mathcal{S}(i, j)} \tag{4}$$

Where $\mathcal{N}^+(i)$ is the set of neighbors of i with a higher density than i. As the membership vectors traverse through the network topology, they naturally diffuse and merge. If a node is located in a "valley" between two "peaks" (boundary between two communities), that node will receive equal "gravitational pull" from both peaks, resulting in a fuzzy membership vector (e.g., $M_i = [0.48, 0.52]$), which is a natural representation of its overlapping nature.

D. Time and Space Complexity Analysis

GSO-DPCD achieves a transition from quadratic to near-linear complexity. Let n represent the number of nodes, m the number of edges, and K the number of detected community peaks. In the Structural Similarity Calculation, we compute the Jaccard similarity for every edge $(u, v) \in E$, between neighborhood sets. In sparse graphs, the intersection of two sets takes $O(d_{avg})$ using a hash-based lookup. The total cost is $T_{sim} = O(m \cdot d_{avg}) \approx O(m)$. Next, we compute the Local Density (ρ) , which involves a linear scan of the adjacency list to sum the weights of incident edges. This costs $T_{density} = O(m)$. Then we sort the n nodes in descending order of ρ, which costs $T_{sort} = O(n \log n)$. Since our Heuristic Path Approximation only inspects the local neighborhood $N(i)$ for each node. The total number of lookups is equivalent to the sum of degrees, hence $T_{delta} = \sum_{i=1}^{n} d(i) = O(m)$. Finally, for the Gravitational Membership Propagation, which costs $T_{prop} = O(m \cdot K)$. The total time complexity of these phases is $O(m \cdot K + n \log n)$. Since K is typically small ($K \ll n$) and real-world graphs are sparse ($m \propto n$), GSO-DPCD operates in near-linear $O(n \log n)$ time.
Memory overhead of GSO-DPCD is $O(m + nK)$, resulting from an adjacency list which requires $O(m)$ space, $O(n)$ for storing ρ, δ, and γ values for each node and $O(n * K)$ for storing the fuzzy state of the network. This indicates that GSO-DPCD can reside entirely in RAM even for networks exceeding millions of edges.

E. Theoretical Foundations

We formally establish the mathematical properties validating the GSO-DPCD framework's structural density definition and fuzzy membership propagation logic.

1) Convergence and Acyclicity of Influence Flow: The membership propagation strategy forms a Directed Acyclic Graph (DAG), ensuring convergence in a single network traversal. **Proof:** Let $\mathcal{G} = (V, E)$ be the input graph. The propagation is an ordering where nodes are sorted such that $\rho(v_1) \geq \rho(v_2) \geq \cdots \geq \rho(v_n)$. The membership vector M_i of node i is derived from the set $\mathcal{N}^+(i) = \{j \in N(i) \mid \rho_j > \rho_i\}$. By definition, for any node $j \in \mathcal{N}^+(i)$, its position in the ordering $\mathcal{O}$ must precede i (i.e., $index(j) < index(i)$). The resulting structures form a Directed Acyclic Graph (DAG) of influence. The absence of cycles guarantees that propagation is strictly feed-forward and non-iterative, ensuring that each node i obtains a deterministic membership vector M_i in a single evaluation [15].

2) Preservation of the Fuzzy Partition: The Gravitational Membership Propagation inherently maintains a valid fuzzy partition where $\sum_{k=1}^{K} m_{ik} = 1$ for all nodes. **Proof:** Initialized community peaks act as base cases with pure unit vectors. Assuming inductively that all denser neighbors $j \in \mathcal{N}^+(i)$ possess valid fuzzy vectors ($\sum_k m_{jk} = 1$), the membership of node i is computed as a convex combination:

$$M_i = \frac{\sum_{j \in \mathcal{N}^+(i)} w_{ij} M_j}{\sum_{j \in \mathcal{N}^+(i)} w_{ij}} \tag{5}$$

Taking the sum across all K communities reduces the numerator to $\sum_j w_{ij}(1)$, which cancels with the denominator to equal 1. Thus, the "flow" of membership preserves the fuzzy distribution throughout the network [16].

F. Robustness Against Topological Hubs

Structural Jaccard Density (ρ) is robust against high-degree "bridge" nodes lacking local cohesion. **Proof:** Consider a "hub" node h that bridges two disjoint cliques but shares only one edge with each. While h has a high degree ($d = 2$), its Jaccard similarity $\mathcal{S}(h, j)$ with any neighbor j will be

TABLE I
SPECIFIC PARAMETERS FOR LFR BENCHMARK GENERATION

Parameter Description	Notation	Value
Number of nodes	N	10^3 to 10^6
Mixing parameter	μ	$[0.1, 0.7]$
Fraction of overlapping nodes	on	$5\% - 50\%$
Overlapping memberships	om	$[2, 10]$
Average degree	k	20
Maximum degree	$maxk$	50
Degree exponent	τ_1	2
Community exponent	τ_2	1
Min community size	$minc$	20
Max community size	$maxc$	100

TABLE II
STATISTICS OF REAL-WORLD DATASETS

Dataset	Acronym	Nodes	Edges
Email-Enron	EE	36,692	183,831
Github	GM	37,700	289,003
Facebook	FB	134,833	1,380,293
Twitch-Gamers	TG	168,114	6,797,557

near zero because their shared neighborhood $|N(h) \cap N(j)|$ is empty (excluding themselves). Conversely, a "core" node c within a clique of size n has $n-1$ neighbors, and for each neighbor, the shared neighborhood is large. Because $\rho_i = \sum S(i,j)$, the core node c will naturally yield a significantly higher density than the bridge hub h, effectively preventing bridge nodes from being incorrectly identified as community peaks [17].

III. EXPERIMENTAL RESULTS AND ANALYSIS

To evaluate the effectiveness and scalability of GSO-DPCD, we conducted a comprehensive experimental study across a diverse set of synthetic and large-scale real datasets. Our analysis emphasizes the algorithm's ability to detect high-quality communities under increasing structural noise, overlapping density, and network size.

A. Experimental Setup

1) Comparison algorithms: We compared GSO-DPCD against a diverse set of state-of-the-art community detection methods. As we already discussed, COPRA [18] and SLPA [19] are standard label propagation-based overlapping community detection algorithms. FuzAg [6] is based on fuzzy agglomeration, EdMot [20] is an edge-enhancement-based modularity optimization algorithm. OCLN [21] is local expansion-based while RaidB [22] is a clique-based algorithm, and CoDeSEG [23] is based on structural entropy game.

2) Datasets and metrics: We use the LFR benchmark generator to generate a comprehensive set of networks by varying the network size, overlaps, complexity and noise. We also use four large-scale real datasets with varying topological structures to test the performance of the proposed approach with its competitors. The details of the datasets and parameters are given in Tables I and II. To evaluate the accuracy of communities detected, we use metrics like Overlapping Normalized Mutual Information (ONMI) [24] and the Omega

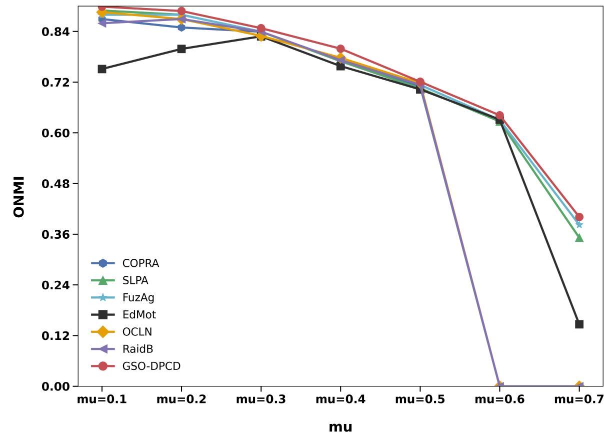

Fig. 2. Plot of the performance of algorithms with increasing noise.

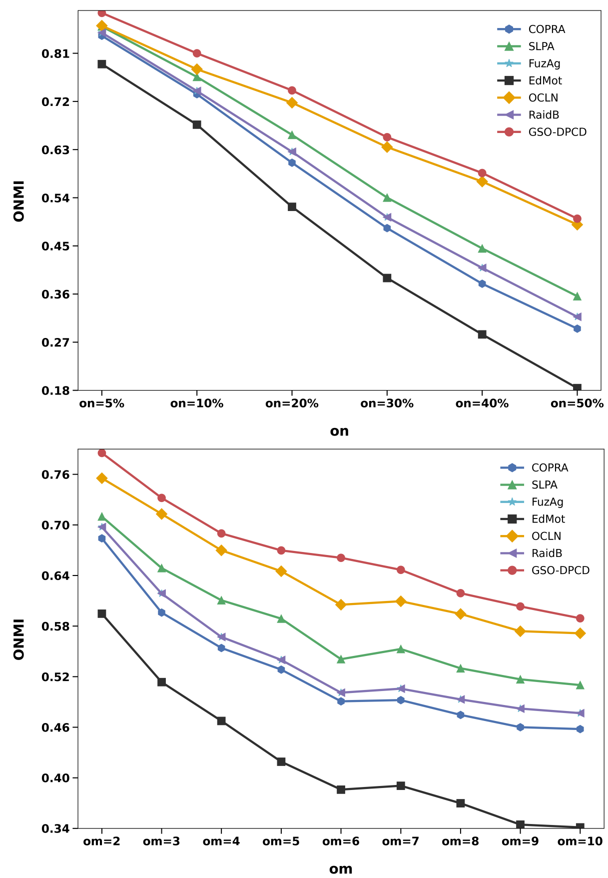

Fig. 3. Plot depicting the performance of algorithms with increasing network overlap.

Index [25] against the ground truth. To assess the quality of communities identified, we use extended modularity (eQ) [26] and Soft Modularity [27] for real datasets.

B. Performance on Synthetic LFR Benchmarks

We used the LFR benchmark networks to test the algorithm across 4 critical dimensions: robustness to increasing network complexity (noise), ability to detect overlapping communities

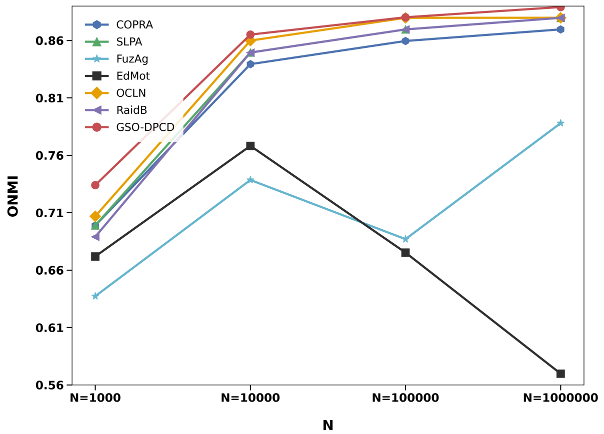

Fig. 4. Plot showing the performance of the algorithms with increasing network size.

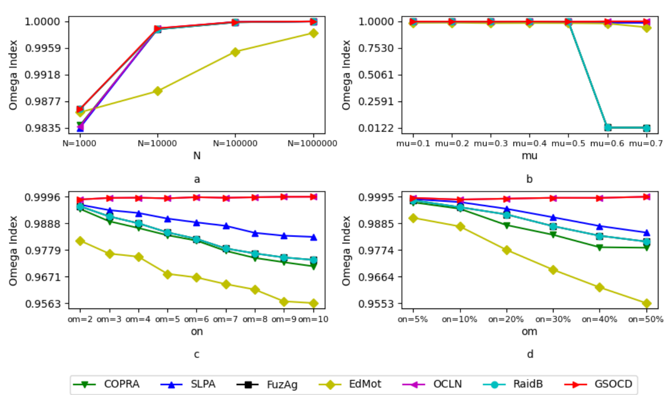

Fig. 5. Comparison of performance of the algorithms with varying network structure evaluated using Omega Index.

in networks with deep overlaps, and accuracy with increasing network size.

1) Robustness Against Mixing Parameter (μ): The mixing parameter μ measures the extent of overlap and noise at the community boundaries. Increasing μ introduces substantial structural noise, which causes the underlying community structure to become increasingly indistinct. Despite this, as shown in fig.2, GSO-DPCD demonstrates strong robustness. At $\mu = 0.7$, while competing methods like COPRA and OCLN fail to identify meaningful communities, GSO-DPCD maintains a 0.4 ONMI, outperforming SLPA and FuzAg. These results demonstrate that the structural gravity field effectively filters inter-community noise.

2) Impact of Overlap Density (om and on): GSO-DPCD was evaluated under varying degrees of community overlap to assess its robustness. As we can observe from fig.3, as the number of communities per node (om) increases to 10, GSO-DPCD achieves an ONMI score of 0.589, surpassing both SLPA and OCLN. At an overlap ratio of 50%, where 50% of the network consists of overlapping nodes, GSO-DPCD remains the top-performing method with an ONMI of 0.5, whereas the performance other methods starts declining.

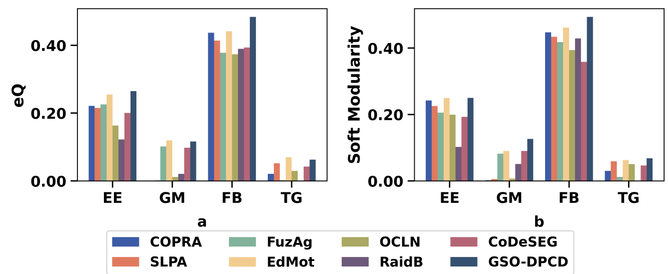

Fig. 6. Performance of the algorithm on large-scale real datasets.

3) Scalability Analysis (N): A central claim of this paper is the $O(N \log N)$ scalability of GSO-DPCD, which is empirically validated across networks ranging from 10^3 to 10^6 nodes. On large-scale networks, with one million nodes, the proposed approach achieves an ONMI of 0.889, outperforming its competitors as shown in fig.4. Notably, the accuracy of the proposed approach increases with network size, demonstrating its readiness for industrial-scale graphs.

4) Performance evaluation using Omega Index: Omega index serves as a measure of agreement between the detected overlapping community assignments to the ground truth. The results of our experiments, as given in the fig. 5, demonstrate that the proposed GSO-DPCD consistently preserves structural integrity even under increasing size, extreme network noise and deep overlaps.

C. Performance on real-world data

For real-world networks where ground truth is unavailable, structural quality is evaluated using modularity-based metrics. As we can observe from fig.6, On the Facebook dataset, GSO-DPCD achieved an eQ of 0.484, outperforming its competitors. On the GitHub network, GSO-DPCD demonstrated strong performance with a Soft Modularity score of 0.12, significantly outperforming its competitors. These results demonstrate that the gravitational flow model of the proposed GSO-DPCD naturally captures the network's structure.

D. Analysis

Under noisy conditions like $\mu \geq 0.7$, motif (EdMot, RaidB) and local-expansion based methods (OCLN) suffer a "cliff-edge" and over-expansion collapse. Our GSO-DPCD employs Structural Jaccard, which ensures cores maintain high "Structural Gravity" despite high inter-community bridges. As Overlap increases ($on \geq 50\%$), LPA-based approaches (COPRA, SLPA) suffer from "label competition" and washout, GSO-DPCD avoids this using deterministic Gravitational Membership Propagation. As the network scale increases, conventional fuzzy approaches (FuzAg) degrade because of global optimization noise. Since our Heuristic Path Approximation keeps the "signal" local and pure, GSO-DPCD performs well in large-scale networks.

IV. Conclusion

In this paper, we introduced GSO-DPCD, a high-performance framework that models overlapping community detection as a Dynamic Gravitational Flow problem. Using the structural mass derived from Jaccard similarity, we bypass the "small-world" distance biases that have historically limited the density-peak-based methods in discrete graph topologies. Extensive experiments provide strong validation for the proposed approach. On LFR benchmark datasets, the algorithm shows remarkable robustness in extremely noisy large networks with deep overlaps. Using Heuristic Path Approximation, the algorithm delivers outstanding performance in near-linear $O(N \log N)$ time. Ultimately, GSO-DPCD proves that fuzzy accuracy and computational efficiency are not mutually exclusive. By treating community detection as a deterministic physical flow, the algorithm provides a stable, scalable and robust tool for uncovering latent structures from a network. A promising extension of GSO-DPCD would be the development of an Incremental Fuzzy Flow mechanism that enables membership vectors to be updated in real time without recomputing the gravitational field. With the rapid growth of large-scale interconnected systems, GSO-DPCD provides a robust foundation for fuzzy-driven structural analysis, enabling a precise lens to view the deep overlaps of our interconnected world.

REFERENCES

[1] G. Palla, I. Derenyi, I. Farkas, and T. Vicsek, "Uncovering the overlapping community structure of complex networks in nature and society," *Nature*, vol. 435, no. 7043, pp. 814–818, Jun. 2005, arXiv:physics/0506133. [Online]. Available: http://arxiv.org/abs/physics/0506133

[2] L. A. Zadeh, "Fuzzy sets," *Information and Control*, vol. 8, no. 3, pp. 338–353, Jun. 1965. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S001999586590241X

[3] M. Blue, B. Bush, and J. Puckett, "Unified approach to fuzzy graph problems," *Fuzzy Sets and Systems*, vol. 125, no. 3, pp. 355–368, Feb. 2002. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0165011401000112

[4] J. C. Bezdek, R. Ehrlich, and W. Full, "FCM: The fuzzy *c*-means clustering algorithm," *Computers & Geosciences*, vol. 10, no. 2, pp. 191–203, Jan. 1984. [Online]. Available: https://www.sciencedirect.com/science/article/pii/0098300484900207

[5] R. Krishnapuram and J. Keller, "A possibilistic approach to clustering," *IEEE Transactions on Fuzzy Systems*, vol. 1, no. 2, pp. 98–110, May 1993. [Online]. Available: https://ieeexplore.ieee.org/document/227387

[6] A. Biswas and B. Biswas, "FuzAg: Fuzzy Agglomerative Community Detection by Exploring the Notion of Self-Membership," *IEEE Transactions on Fuzzy Systems*, vol. 26, no. 5, pp. 2568–2577, Oct. 2018. [Online]. Available: https://ieeexplore.ieee.org/document/8263601

[7] U. K. Roy, P. K. Muhuri, and S. K. Biswas, "NeSiFC: Neighbors' Similarity-Based Fuzzy Community Detection Using Modified Local Random Walk," *IEEE Transactions on Cybernetics*, vol. 52, no. 10, pp. 10 014–10 026, Oct. 2022. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/9464160

[8] A. Rodriguez and A. Laio, "Clustering by fast search and find of density peaks | Science," Jun. 2014. [Online]. Available: https://www.science.org/doi/10.1126/science.1242072

[9] M. Du, S. Ding, and H. Jia, "Study on density peaks clustering based on k-nearest neighbors and principal component analysis," *Knowledge-Based Systems*, vol. 99, pp. 135–145, May 2016. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0950705116000794

[10] J. Xie and B. K. Szymanski, "LabelRank: A stabilized label propagation algorithm for community detection in networks," in *2013 IEEE 2nd Network Science Workshop (NSW)*. West Point, NY, USA: IEEE, Apr. 2013, pp. 138–143. [Online]. Available: http://ieeexplore.ieee.org/document/6609210/

[11] D. J. Watts and S. H. Strogatz, "Collective dynamics of 'small-world' networks," *Nature*, vol. 393, no. 6684, pp. 440–442, Jun. 1998. [Online]. Available: https://www.nature.com/articles/30918

[12] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, "Fast unfolding of communities in large networks," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2008, no. 10, p. P10008, Oct. 2008. [Online]. Available: https://dx.doi.org/10.1088/1742-5468/2008/10/P10008

[13] V. A. Traag, L. Waltman, and N. J. van Eck, "From Louvain to Leiden: guaranteeing well-connected communities," *Scientific Reports*, vol. 9, no. 1, p. 5233, Mar. 2019. [Online]. Available: https://www.nature.com/articles/s41598-019-41695-z

[14] S. Gómez, P. Jensen, and A. Arenas, "Analysis of community structure in networks of correlated data," *Physical Review E*, vol. 80, no. 1, p. 016114, Jul. 2009. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevE.80.016114

[15] A. B. Kahn, "Topological sorting of large networks," *Commun. ACM*, vol. 5, no. 11, pp. 558–562, Nov. 1962. [Online]. Available: https://dl.acm.org/doi/10.1145/368996.369025

[16] J. C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms*. Boston, MA: Springer US, 1981. [Online]. Available: http://link.springer.com/10.1007/978-1-4757-0450-1

[17] L. Hamers, Y. Hemeryck, G. Herweyers, M. Janssen, H. Keters, R. Rousseau, and A. Vanhoutte, "Similarity measures in scientometric research: The Jaccard index versus Salton's cosine formula," *Information Processing & Management*, vol. 25, no. 3, pp. 315–318, Jan. 1989. [Online]. Available: https://www.sciencedirect.com/science/article/pii/0306457389900484

[18] S. Gregory, "Finding overlapping communities in networks by label propagation," *New Journal of Physics*, vol. 12, no. 10, p. 103018, Oct. 2010. [Online]. Available: https://dx.doi.org/10.1088/1367-2630/12/10/103018

[19] J. Xie, B. K. Szymanski, and X. Liu, "SLPA: Uncovering Overlapping Communities in Social Networks via a Speaker-Listener Interaction Dynamic Process," in *2011 IEEE 11th International Conference on Data Mining Workshops*, Dec. 2011, pp. 344–349. [Online]. Available: https://ieeexplore.ieee.org/document/6137400

[20] P.-Z. Li, L. Huang, C.-D. Wang, and J.-H. Lai, "EdMot: An Edge Enhancement Approach for Motif-aware Community Detection," in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, ser. KDD '19. New York, NY, USA: Association for Computing Machinery, Jul. 2019, pp. 479–487. [Online]. Available: https://doi.org/10.1145/3292500.3330882

[21] F. Cheng, C. Wang, X. Zhang, and Y. Yang, "A Local-Neighborhood Information Based Overlapping Community Detection Algorithm for Large-Scale Complex Networks," *IEEE/ACM Transactions on Networking*, vol. 29, no. 2, pp. 543–556, Apr. 2021. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/9288733

[22] P. G. Sun, X. Wu, Y. Quan, and Q. Miao, "Rearranging 'indivisible' Blocks for Community Detection," *IEEE Transactions on Knowledge and Data Engineering*, vol. 35, no. 6, pp. 6252–6263, Jun. 2023. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/9771068

[23] Y. Xian, P. Li, H. Peng, Z. Yu, Y. Xiang, and P. S. Yu, "Community detection in large-scale complex networks via structural entropy game," in *Proceedings of the ACM on Web Conference 2025*, 2025, pp. 3930–3941.

[24] A. Lancichinetti and S. Fortunato, "Normalized mutual information to evaluate overlapping community finding algorithms," https://www.researchgate.net/publication/51944343, 2009, accessed: May 4, 2025.

[25] L. M. Collins and C. W. Dent, "Omega: A General Formulation of the Rand Index of Cluster Recovery Suitable for Non-disjoint Solutions," *Multivariate Behavioral Research*, vol. 23, no. 2, pp. 231–242, Apr. 1988.

[26] A. Lancichinetti, S. Fortunato, and J. Kertész, "Detecting the overlapping and hierarchical community structure in complex networks," *New Journal of Physics*, vol. 11, no. 3, p. 033015, Mar. 2009. [Online]. Available: https://dx.doi.org/10.1088/1367-2630/11/3/033015

[27] T. C. Havens, J. C. Bezdek, C. Leckie, K. Ramamohanarao, and M. Palaniswami, "A Soft Modularity Function For Detecting Fuzzy Communities in Social Networks," *IEEE Transactions on Fuzzy Systems*, vol. 21, no. 6, pp. 1170–1175, Dec. 2013. [Online]. Available: https://ieeexplore.ieee.org/document/6450075/similar