Approximate Anchored Densest Subgraph Search on Large Static and Dynamic Graphs

Qi Zhang
Beijing Institute of Technology
Beijing, China
qizhangcs@bit.edu.cn

Yalong Zhang
Beijing Institute of Technology
Beijing, China
yalong-zhang@qq.com

Rong-Hua Li
Beijing Institute of Technology
Beijing, China
lironghuabit@126.com

Guoren Wang
Beijing Institute of Technology
Beijing, China
wanggrbit@gmail.com

ABSTRACT

Densest subgraph search, aiming to identify a subgraph with maximum edge density (i.e., average degree), faces limitations as the edge density metric inadequately reflects biases towards a given vertex set R. To address this issue, the concept of R-subgraph density was introduced, refining the doubled edge density by penalizing vertices in a subgraph but not in R, using the vertex degree as a penalty factor. This advancement leads to the Anchored Densest Subgraph (ADS) search problem, which focuses on finding the subgraph with the highest R-subgraph density for a given set R. Nonetheless, current algorithms for ADS search face significant inefficiencies in handling large-scale graphs or the sizable R set. Furthermore, these algorithms require re-execution to compute the ADS whenever the graph is dynamically updated, complicating the efficient maintenance within dynamic graphs. To tackle these challenges, we investigate the problem of Approximate Anchored Densest Subgraph (AADS) search on large static and dynamic graphs. This problem aims to identify a subgraph $S^* \subseteq V$ within G whose R-subgraph density rounds up from that of the ADS. We present an efficient global algorithm incorporating the re-orientation network flow technique and binary search, operating in time polynomial to the graph's size. Additionally, we propose a novel local algorithm employing shortest-path-based methods for the max-flow computation from s to t around R, markedly boosting performance in scenarios with larger R sets. For dynamic graphs, both basic and improved algorithms are developed to efficiently maintain the AADS when an edge is updated. Extensive experiments and two detailed case studies demonstrate the efficiency, scalability, and effectiveness of our solutions.

PVLDB Reference Format:
Qi Zhang, Yalong Zhang, Rong-Hua Li, and Guoren Wang. Approximate Anchored Densest Subgraph Search on Large Static and Dynamic Graphs. PVLDB, 14(1): XXX-XXX, 2020. doi:XX.XX/XXX.XX

PVLDB Artifact Availability:

The source code, data, and/or other artifacts have been made available at https://anonymous.4open.science/r/Approximate-ADS-F26F/.

Proceedings of the VLDB Endowment, Vol. 14, No. 1 ISSN 2150-8097.
doi:XX.XX/XXX.XX

1 INTRODUCTION

A graph, comprising vertices as entities and edges as relationships, fundamentally models various complex real-world networks [1, 2, 30, 35, 43]. Densest subgraph search, a central problem in graph analysis, aims to identify a subgraph S with the maximum edge density, which is defined as the ratio of edges to vertices, i.e., $\frac{|E(S)|}{|V(S)|}$ [8, 9, 15, 25, 33, 45, 53, 55]. This problem has unveiled their extensive applicability across various domains, such as identifying dense structures in scientific collaborations [40], biological networks [24, 47], and social networks [29, 37, 44]. Furthermore, this investigation has significantly advanced graph database, notably in the areas of query processing [18, 32] and visualization [58, 59].

Densest subgraph search typically identifies only a single, globally densest subgraph, which falls short of fulfilling the requirements of real-world applications. Recent advancements have introduced a new category of research dedicated to enhancing the diversity of the densest subgraph discoveries, especially in finding densest subgraphs related to specific seed sets [9, 15, 25, 45, 53, 55]. Nevertheless, these works face several notable challenges. First, the traditional edge-based density metric fails to accurately capture the bias of a subgraph S towards a given vertex set. Additionally, the requirement that derived subgraphs be disjoint or nested causes all queries in connected and regular graphs to correspond to only one result, thus introducing the degeneracy problem. Moreover, the computational burden imposed by rigid overlapping constraints for large-scale graphs also represents a considerable obstacle.

Addressing the challenges presented, Dai *et al.* [19] integrate the principles of density and locality by introducing a novel metric for a subgraph S, the R-subgraph density, denoted as $\rho_R(S) = \frac{2|E(S)|-\sum_{v\in S\setminus R} d_G(v)}{|V(S)|}$. This metric refines the doubled traditional density, $\frac{2|E(S)|}{|V(S)|}$, by imposing a penalty for vertices in S not included in R, symbolized by $\sum_{v\in S\setminus R} d_G(v)$. It effectively measures S's predisposition towards a specific reference vertex set R, thereby circumventing the degeneracy problem. Utilizing this metric, Dai *et al.* [19] formulate the anchored densest subgraph search problem, which seeks to pinpoint the subgraph $\hat{S}$ with the highest R-subgraph density for a given set R. They propose an algorithm, referred to as ADSGlobal, which exhibits a computational complexity that grows

Table 1: The time complexity of different algorithms.

Static graphs		Dynamic graphs	
Algorithm	Time Complexity	Algorithm	Time Complexity
ADSGlobal [19]	$O(nm \log \frac{n^2}{m})$	ADSGlobal [19]	$O(nm \log \frac{n^2}{m})$
ADSLocal [19]	$\geq O(\log n \text{Vol}^4(R))$	ADSLocal [19]	$\geq O(\log n \text{Vol}^4(R))$
AADSGlobal [Ours]	$O(m^{1.5} \log d_{\max})$	BasicIns/BasicDel [Ours]	$O(m^{1.5})$
AADSLocal [Ours]	$O(\log d_{\max} \text{Vol}^3(R))$	ImproveIns/ImproveDel [Ours]	$O(\text{Vol}^3(R))$

polynomially with the input graph G's size. Moreover, a local algorithm, ADSLocal, is introduced, with its complexity dependent solely on the input set R's size, making it independent of the graph G's overall size.

However, the investigation by Dai *et al.* [19] encounters several notable limitations. First, the ADSGlobal algorithm exhibits significant inefficiencies in processing large-scale graphs, a drawback that persists even with diminutive query sets, indicating a scalability issue. Second, the performance of the ADSLocal algorithm closely resembles that of the ADSGlobal algorithm when dealing with a large reference set R, nullifying anticipated efficiency gains. The fundamental cause of this similarity lies in the ADSLocal algorithm's requirement for iterative subgraph expansion, a method that significantly increases computational complexity. Lastly, updates to the graph necessitate a full recalculation, thereby challenging the efficient maintenance of the ADS for dynamic graphs.

To address these issues, we investigate the approximate anchored densest subgraph search problem on large static and dynamic graphs. This problem aims to find a subgraph $S^* \subseteq V$ of G satisfying $\lceil \rho_R(S^*) \rceil = \lceil \rho_R(\hat{S}) \rceil$, where $\hat{S}$ represents the ADS and S^* is the AADS. Importantly, the difference in R-subgraph density between S^* and $\hat{S}$ is constrained to not exceed 1 (i.e., $\rho_R(\hat{S}) - \rho_R(S^*) < 1$), thus providing a methodological guarantee of approximation accuracy. To search the AADS, we first introduce an efficient global algorithm that incorporates a re-orientation network flow technique and a binary search, addressing concerns of scalability. Subsequently, a novel local algorithm is proposed that leverages shortest-path based methods for localized max-flow computation from s to t around R. This algorithm significantly improves performance in scenarios involving larger values of $|R|$. Additionally, for dynamic graphs, we prove a novel anchored densest subgraph update theorem, based on which we develop both basic and improved algorithms to efficiently maintain the AADS in response to edge insertions and deletions. In summary, our contributions are as follows.

A global algorithm for AADS search. To compute the AADS, we present an algorithm, named AADSGlobal, incorporating the re-orientation network flow technique alongside a binary search method. Specifically, the re-orientation network flow technique evaluates whether $\lceil \rho_R(S^*) \rceil \geq \alpha$ using the max-flow method, while the binary search iteratively adjusts the guessed value α to ascertain the AADS. Remarkably, the AADSGlobal algorithm requires only an integer guess for α, resulting in a time complexity of $O(m^{1.5} \log d_{\max})$ with the classic Dinic's max-flow method [56]. This represents a significant efficiency improvement over the ADSGlobal algorithm, tailored for ADS search, which has a time complexity of $O(nm \log \frac{n^2}{m})$.

A novel local algorithm for AADS search. To improve the efficiency, we propose a novel local algorithm for the AADS search, denoted as AADSLocal. A central innovation of AADSLocal is the introduction of the probing method, LocalReTest, which leverages shortest-path based methods for local max-flow computation from s to t around R. Significantly, for a specified α, whereas the probing method of ADSLocal determines the max-flow through a sequence of progressively expanding subgraphs, LocalReTest eliminates the necessity for this intricate iterative mechanism. Instead, it achieves max-flow computation in tandem with the construction of the locally augmented subgraph. The AADSLocal algorithm exhibits a time complexity of $O(\log d_{\max} \text{Vol}^3(R))$. In contrast, the ADSLocal algorithm, in its last iteration, features $O(\text{Vol}^2(R))$ vertices and $O(\text{Vol}^2(R))$ edges. Employing the state-of-the-art max-flow method, the final iteration complexity reaches $O(\text{Vol}^4(R))$, thus establishing the minimum overall complexity of ADSLocal at $O(\log n \text{Vol}^4(R))$, markedly surpassing that of AADSLocal.

The efficient algorithms for AADS maintenance. To handle dynamic updates of the graph, we develop algorithms for maintaining the AADS. We prove a novel anchored densest subgraph update theorem, which elucidates that the insertion or deletion of an edge can only affect the value of $\lceil \rho_R(\hat{S}) \rceil$ in one of three distinct manners: it may either remain unchanged, increase by 1, or decrease by 1. Based on this theorem, two fundamental algorithms: BasicIns and BasicDel are proposed for handling edge insertions and deletions, respectively. These algorithms enable the efficient maintenance of the AADS through merely two probes of the guessed value α, achieving a computational complexity of $O(m^{1.5})$. Furthermore, we present two improved algorithms, ImproveIns and ImproveDel, which focus on maintaining an *unreversible orientation* $\vec{G}$ and two important subgraphs to preserve the AADS (details see Section 5.2). The worst-case time complexity of our improved algorithms is $O(\text{Vol}^3(R))$ because it necessitates applying the LocalReTest whenever $\lceil \rho_R(\hat{S}) \rceil$ changes. In empirical evaluations, both ImproveIns and ImproveDel demonstrate high efficiency as they require merely limited BFS computations in the majority of cases.

Extensive experiments. We conduct comprehensive experiments on 14 real-life graphs, encompassing both medium and large sizes, to evaluate the efficiency of our algorithms for the search and maintenance of AADS. The results show that: (1) AADSGlobal for AADS search achieves speeds 3 to 13 times faster and reduces memory consumption by 3.5 to 6.5 times compared to ADSGlobal tailored for ADS search; (2) The AADSLocal algorithm demonstrably outperforms ADSLocal, with speedups ranging from 5 to 1500 times, while only marginally increasing memory usage; (3) For AADS maintenance, our basic and improved algorithms are significantly superior to the method that recomputes the ADS (AADS) using ADSLocal (AADSLocal). ImproveIns (ImproveDel) consistently outperforms BasicIns (BasicDel) by at least an order of magnitude across most datasets; (4) The divergence in R-subgraph densities between the ADS $\hat{S}$ and the AADS S^*, namely, $\rho_R(\hat{S}) - \rho_R(S^*)$, invariably remains below 0.2, a figure substantially lower than the theoretical maximum difference of 1. Additionally, we also conduct two case studies on dblp and amazon to illustrate the effectiveness of the AADS. The results show that AADS can better align a given reference set R compared to the exact ADS, while maintaining a close R-subgraph density to that of the ADS. For reproducibility, the source code of this paper is released (see Artifact availability).

2 PRELIMINARIES

Consider an unweighted and undirected graph $G = (V, E)$, where V represents the set of vertices and E denotes the set of edges. Let n and m symbolize the number of vertices and edges in G, respectively. An edge between vertices u and v in G is denoted as (u, v) and equivalently, (v, u). The neighbors of u within G, $N_G(u)$, is defined as $\{v \in V | (u, v) \in E\}$, and the degree of u within G, $d_G(u)$, is given by the cardinality of $N_G(u)$. For a vertex subset S of V, $N_G(S)$ represents the set of neighbors of vertices in S, formally expressed as $N_G(S) = \{v \in V \backslash S | \exists u \in S, (u, v) \in E\}$, and $E(S)$ specifies the edges connecting vertices within S, defined as $E(S) = \{(u, v) \in E | u, v \in S\}$. For two disjoint vertex subsets S and T, the cross edges between S and T are denoted as $E_\times(S, T) = \{(u, v) \in E | u \in S, v \in T\}$, and the additional edges from S with respect to T as $E_\triangle(S, T) = E(S) \cup E_\times(S, T)$.

Given an unweighted and directed graph $\vec{G} = (V, \vec{E})$ with V as the set of vertices and $\vec{E}$ as the set of directed edges. The directed edge from u to v is represented by $\langle u, v \rangle$. A path in this directed graph is a sequence of vertices $v_s = v_0, v_1, ..., v_{l-1}, v_l = v_t$, where $\langle v_{i-1}, v_i \rangle \in \vec{E}$ for $i = 1, ..., l$; for conciseness, this path can also be represented as $v_s \rightsquigarrow v_t$. The set of in-neighbors of u in $\vec{G}$ is denoted as $N_{\vec{G}}(u) = \{v \in V | \langle v, u \rangle \in \vec{E}\}$, and the indegree of u in $\vec{G}$ is $\vec{d}_{\vec{G}}(u) = |N_{\vec{G}}(u)|$. For simplicity, we omit the subscripts G and $\vec{G}$ in the above notations when the context is clear.

Prior to introducing the anchored densest subgraph search problem, we first elucidate the concept of R-subgraph density.

Definition 2.1. (R-subgraph Density [19]) Given a graph $G = (V, E)$ and a reference vertex set $R \subseteq V$, the R-subgraph density of a subgraph $S \subseteq V$ is defined as:

$$\rho_R(S) = \frac{2|E(S)| - \sum_{v \in S \backslash R} d_G(v)}{|V(S)|} \tag{1}$$

Following the R-subgraph density, the concept of anchored densest subgraph is defined as follows.

Definition 2.2. (Anchored Densest Subgraph [19]) Given an undirected graph $G = (V, E)$, an anchor vertex set $A \subseteq V$, a reference vertex set $R \subseteq V$ satisfying $A \subseteq R$ and $E(R) \neq \emptyset$, the anchored densest subgraph, denoted as $\hat{S}$, is a subgraph of G that includes all vertices in A and exhibits the maximum R-subgraph density, i.e., $\hat{S} = \arg\max_{A \subseteq S \subseteq V} \rho_R(S)$.

Example 2.3. Consider a graph $G = (V, E)$ as shown in Figure 1. Let the vertex sets R and A be given as $R = \{v_1, v_3, v_4, v_5, v_6, v_8\}$ and $A = \{v_6\}$, respectively. For the vertex set $S = \{v_1, v_3, v_4, v_5, v_6\}$, the R-subgraph density of S, $\rho_R(S)$, is calculated to be $\frac{2\times 8-0}{5} = 3.2$. It is determined that S possesses the highest R-subgraph density with $A \subseteq S$, identifying it as the anchored densest subgraph (i.e., the area within the red dotted line in Figure 1).

To compute the anchored densest subgraph, Dai *et al.* [19] introduced the ADSGlobal and ADSLocal algorithms. However, these algorithms face significant challenges when applied to large-scale graphs or large R sets. Moreover, any update to the graph necessitates a full recomputation of the ADS, thus undermining its maintenance efficiency. Addressing these limitations, this paper investigates the problem of approximate anchored densest subgraph search, which is formulated as follows.

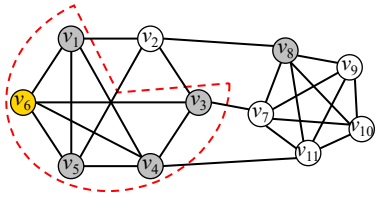


Figure 1: An example graph G

Definition 2.4. (Approximate Anchored Densest Subgraph) Given an undirected graph $G = (V, E)$, an anchor vertex set $A \subseteq V$, a reference vertex set $R \subseteq V$ satisfying $A \subseteq R$ and $E(R) \neq \emptyset$, the approximate anchored densest subgraph, denoted as S^*, is a subgraph of G satisfying: (1) $A \subseteq S^*$; (2) $\lceil \rho_R(S^*) \rceil = \lceil \rho_R(\hat{S}) \rceil$, where $\hat{S}$ is the anchored densest subgraph.

According to Definition 2.4, the following fact is established.

Fact 2.5. *The R-subgraph density of S^* provides an additive approximation to the R-subgraph density of $\hat{S}$, which is less than 1, i.e., $\rho_R(\hat{S}) - \rho_R(S^*) < 1$.*

Problem Statement. Given an undirected graph $G = (V, E)$, an anchor vertex set $A \subseteq V$, a reference vertex set $R \subseteq V$ satisfying $A \subseteq R$ and $E(R) \neq 0$, our goal is to search and maintain the approximate anchored densest subgraph S^* on large static and dynamic graphs.

Remark. Fact 2.5 establishes theoretical guarantees for the additive approximation less than 1. However, our empirical analyses indicate that in real-world graphs, the difference between $\rho_R(\hat{S})$ and $\rho_R(S^*)$ is considerably smaller than 1.

3 AN EFFICIENT GLOBAL ALGORITHM

In this section, we propose an efficient algorithm, i.e., AADSGlobal, aimed at calculating the approximate anchored densest subgraph S^*. The essence of AADSGlobal is to employ a binary search strategy to examine whether $\lceil \rho_R(S^*) \rceil \geq \alpha$. Specifically, if $\lceil \rho_R(S^*) \rceil \geq \alpha$, it is inferred that an increment in α is justified. Conversely, $\lceil \rho_R(S^*) \rceil < \alpha$ indicates a need to reduce α. Below, we begin by checking whether $\lceil \rho_R(S^*) \rceil \geq \alpha$ with the re-orientation network flow technique, followed by a detailed introduction of the AADSGlobal algorithm integrating a binary search approach.

3.1 Checking whether $\lceil \rho_R(S^*) \rceil \geq \alpha$

Here we propose a probing algorithm, called ReTest, equipped with the re-orientation network flow technique to check whether $\lceil \rho_R(S^*) \rceil \geq \alpha$. Prior to delving into ReTest, we introduce three important concepts: *orientation*, *bounty*, and *re-orientation network*.

Definition 3.1. (Orientation) Given a graph $G = (V, E)$, the orientation of G, represented as $\vec{G} = (V, \vec{E})$, is a directed graph where the vertex set remains identical to that of G; while, for each edge (u, v) in E, $\vec{G}$ features two directed edges between u and v.

Definition 3.2. (Bounty) Given a graph G and its orientation $\vec{G} = (V, \vec{E})$, the bounty of a vertex u, denoted as δ_u, is defined as: (1) $\delta_u = +\infty$ if $u \in A$; (2) $\delta_u = \vec{d}(u)$ if $u \in R \backslash A$; (3) $\delta_u = \vec{d}(u) - d(u)$ if $u \in V \backslash R$.

Utilizing the concept of bounty, for a given value of α, the re-orientation network flow technique enables the construction of an

edge-weighted and directed graph $G_\alpha = (V_\alpha, \vec{E}_\alpha, c)$ by augmenting the orientation $\vec{G} = (V, \vec{E})$ as follows.

- Add a source vertex s and a sink vertex t to V_α, and add all vertices in V to V_α, i.e., $V_\alpha = V \cup \{s, t\}$.
- For each directed edge $\langle u, v\rangle \in \vec{E}$, add an arc $\langle u, v\rangle$ to E_α with capacity 1.
- For each vertex $u \in V$, add an arc $\langle s, u\rangle$ to E_α, if $\delta_u < \alpha - 1$ with capacity $(\alpha - 1) - \delta_u$.
- For each vertex $u \in V$, add an arc $\langle u, t\rangle$ to E_α, if $\delta_u > \alpha - 1$ with capacity $\delta_u - (\alpha - 1)$.

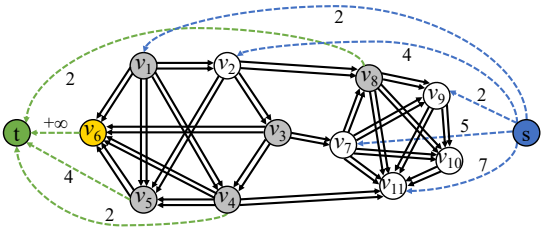


Figure 2: The initial orientation $\vec{G}$ of G

Example 3.3. Consider the graph G depicted in Figure 1. We assume that the orientation $\vec{G}$ of G corresponds to the subgraph of Figure 2, induced by all vertices of G, wherein all directed edges orient towards vertices possessing higher IDs. Given the same sets R and A as in Example 2.3 and $\alpha = 3$, the re-orientation network G_α is illustrated as Figure 2. Within G_α, all edges in $\vec{G}$ are assigned a weight of 1, which we omit for clarity. Take vertex v_2 as an example; it is not included in R, resulting in a bounty $\delta_{v_2} = 2 - 4 = -2$. Since $\delta_{v_2} < \alpha - 1$, there is a directed edge from the s to v_2 with a weight of 4 in G_α. For vertex $v_8 \in R \backslash A$ with $\delta_{v_8} = 4 > \alpha - 1$, G_α includes a directed edge from v_8 to t with a weight of 2. While vertex v_6 belongs to A and thus it is connected to t in G_α through an edge weighted at positive infinity.

Based on the re-orientation network, the ReTest algorithm establishes a relationship between the max-flow computation on G_α and the R-subgraph density of a subgraph to determine whether $\lceil \rho_R(S^*) \rceil \geq \alpha$. The pseudo-code of ReTest is outlined in lines 8-19 of Algorithm 1. Initially, ReTest constructs the orientation $\vec{G}$ by inserting the directed edge $\langle u, v\rangle$ twice for each edge (u, v) in G (lines 10-11). It then calculates the bounty of each vertex in G (lines 12-14) and constructs the re-orientation network G_α with a guess value α (line 15). The algorithm proceeds to compute the maximum flow on G_α and reverses the direction of each edge in $\vec{G}$ if it is saturated in G_α (lines 16-18). Ultimately, the ReTest algorithm yields a subgraph S comprising all vertices with a bounty not less than α, or those capable of reaching a vertex with a bounty that meets or exceeds α (line 19). Regarding the subgraph S, the following theorem establishes the correctness of the ReTest algorithm.

Theorem 3.4. *The subgraph S output by the* ReTest *algorithm exhibits the following properties: (1) $A \subseteq S$ and $\lceil \rho_R(S) \rceil \geq \alpha$ if $\lceil \rho_R(\hat{S}) \rceil \geq \alpha$; (2) $A \subseteq S$ and $\lceil \rho_R(S) \rceil < \alpha$ if $\lceil \rho_R(\hat{S}) \rceil < \alpha$.*

Proof. We begin by proving case (2), where $\lceil \rho_R(\hat{S}) \rceil < \alpha$. The bounty of arbitrary vertex $u \in A$ is assigned to infinite, i.e., $\delta_u = +\infty$. After performing the max-flow computation, δ_u remains significantly greater than α, which unequivocally establishes that $u \in S$

Algorithm 1: AADSGlobal(G, A, R)

Input: An undirected graph $G = (V, E)$, an anchor vertex set $A \subseteq V$, a reference vertex set $R \subseteq V$ satisfying $A \subseteq R$ and $E(R) \neq 0$
Output: The AADS S^*
1 $\alpha_l \leftarrow 1$; $\alpha_u \leftarrow \max_{u \in R} d(u)$;
2 **while** $\alpha_l < \alpha_u$ **do**
3 $\quad \alpha_m \leftarrow \lceil (\alpha_l + \alpha_u + 1)/2 \rceil$;
4 $\quad S^* \leftarrow$ ReTest(G, A, R, α_m);
5 $\quad$ **if** $\lceil \rho_R(S^*) \rceil \geq \alpha$ **then** $\alpha_l \leftarrow \alpha_m$;
6 $\quad$ **else** $\alpha_u \leftarrow \alpha_m - 1$;
7 **return** ReTest(G, A, R, α_l);
8 **Procedure** ReTest(G, A, R, α)
9 $\vec{E} \leftarrow \emptyset$;
10 **foreach** $(u, v) \in E$ **do**
11 $\quad \vec{E} \leftarrow \vec{E} \cup \langle u, v\rangle$; $\vec{E} \leftarrow \vec{E} \cup \langle u, v\rangle$;
12 **foreach** $u \in V \backslash R$ **do** $\delta_u \leftarrow \vec{d}(u) - d(u)$;
13 **foreach** $u \in R \backslash A$ **do** $\delta_u \leftarrow \vec{d}(u)$;
14 **foreach** $u \in A$ **do** $\delta_u \leftarrow +\infty$;
15 Construct the re-orientation network $G_\alpha = (V_\alpha, \vec{E}_\alpha, c)$;
16 Compute a maximum flow $f_{\max}$ in $G_\alpha = (V_\alpha, \vec{E}_\alpha, c)$ by Dinic's algorithm;
17 **foreach** $\langle u, v\rangle \in \vec{E}$ **do**
18 $\quad$ **if** $\langle u, v\rangle$ *is satruated in* G_α **then** Reverse the edge $\langle u, v\rangle \in \vec{E}$;
19 **return** $S \leftarrow \{u \in V | \delta_u \geq \alpha$ *or* u *can reach a vertex* v *with* $\delta_v \geq \alpha\}$;

and, by extension, that $A \subseteq S$. Given that $\hat{S}$ has the highest R-subgraph destiny, it follows that $\rho_R(S) \leq \rho_R(\hat{S}) \leq \lceil \rho_R(\hat{S}) \rceil < \alpha$. Thus, case 2 is established.

Below, we analysis the case of $\lceil \rho_R(\hat{S}) \rceil \geq \alpha$. For an arbitrary vertex set $T \subseteq \hat{S}$, given that $\hat{S}$ possesses the highest R-subgraph density, we establish that $\rho_R(\hat{S} \backslash T) \leq \rho_R(\hat{S})$, i.e.,

$$\frac{2|E(\hat{S}\backslash T)| - \sum_{u \in (\hat{S}\backslash T)\backslash R} d(u)}{|\hat{S}| - |T|} \leq \rho_R(\hat{S}). \tag{2}$$

This can be further expressed as:

$$\frac{2(|E(\hat{S})| - |E(T)| - |E_\times(T, \hat{S}\backslash T)|) - (\sum_{u \in \hat{S}\backslash R} d(u) - \sum_{u \in T\backslash R} d(u))}{|\hat{S}| - |T|} \leq \rho_R(\hat{S}). \tag{3}$$

Rearranging the terms, we have

$$\rho_R(\hat{S}) \times |\hat{S}| - 2(|E(T)| + |E_\times(T, \hat{S}\backslash T)|) + \sum_{u \in T\backslash R} d(u) \leq \rho_R(\hat{S}) \times (|\hat{S}| - |T|). \tag{4}$$

Thus, the inequality

$$2|E(T)| + 2|E_\times(T, \hat{S}\backslash T)| - \sum_{u \in T\backslash R} d(u) \geq \rho_R(\hat{S}) \times |T| \tag{5}$$

is established.

Based on Equation 5, we now prove that $\hat{S} \subseteq S$. Assume, for the sake of contradiction, that $\hat{S} \nsubseteq S$, implying that the set $T = \hat{S} \setminus S$ is non-empty. Given the properties of S, all edges in $E_\times(S, V \setminus S)$ are directed towards $V \setminus S$. Noting that $(\hat{S} \setminus T) \subseteq S$, it follows that all edges in $E_\times(T, \hat{S} \setminus T)$ are directed towards T. This leads to the inequality:

$$\sum_{u \in T} \delta_u = \sum_{u \in T} \vec{d}(u) - \sum_{u \in T\backslash R} d(u) \tag{6}$$

$$\geq 2|E(T)| + 2|E_\times(T, \hat{S} \setminus T)| - \sum_{u \in T\backslash R} d(u) \tag{7}$$

$$\geq \rho_R(\hat{S}) \times |T|. \tag{8}$$

Inequality (7) follows from the fact that there may exist other directed edges pointing from vertex $v \in V \backslash S$ to $u \in T$ and Inequality (8) is deduced from Equation 5. Due to $\sum_{u \in T} \delta_u \geq \rho_R(\hat{S}) \times |T|$, by the pigeonhole principle, there exists at least one $u \in T$ such that $\delta_u \geq \rho_R(\hat{S})$, which implies $\delta_u \geq \lceil \rho_R(\hat{S}) \rceil$. This suggests that u

should be included in S, contradicting the assumption that $T = \hat{S} \setminus S$. Thus, $\hat{S} \subseteq S$ is established.

With the conclusion of $A \subseteq \hat{S} \subseteq S$, we further illustrate that if $\lceil \rho_R(\hat{S}) \rceil \geq \alpha$, then $\lceil \rho_R(S) \rceil \geq \alpha$. Following the maximum flow computation on G_α, it is evident that no $v_s \rightsquigarrow v_t$ path exists where $\delta_{v_s} < \alpha - 1$ and $\delta_{v_t} > \alpha - 1$. The absence of such paths is due to their potential for flow augmentation. Therefore, any a vertex u in S satisfies $\delta_u \geq \alpha - 1$, and we further have

$$\sum_{u \in S \setminus \hat{S}} \delta_u \geq |S \setminus \hat{S}| \times (\alpha - 1). \tag{9}$$

Since $A \subseteq \hat{S}$, the set $S \setminus \hat{S}$ does not contain any vertex from A. Then, the R-subgraph density of S can be expressed as:

$$\rho_R(S) = \frac{2|E(S)| - \sum_{u \in S \setminus R} d(u)}{|S|} \tag{10}$$

$$= \frac{2|E(\hat{S})| - \sum_{u \in \hat{S} \setminus R} d(u)}{|S|} + \frac{2|E(S \setminus \hat{S})| + 2|E_\times(\hat{S}, S \setminus \hat{S})| - \sum_{u \in (S \setminus \hat{S}) \setminus R} d(u)}{|S|} \tag{11}$$

$$\geq \frac{2|E(\hat{S})| - \sum_{u \in \hat{S} \setminus R} d(u)}{|S|} + \frac{\sum_{u \in S \setminus \hat{S}} \vec{d}(u) - \sum_{u \in (S \setminus \hat{S}) \setminus R} d(u)}{|S|} \tag{12}$$

$$= \frac{\rho_R(\hat{S}) \times |\hat{S}|}{|S|} + \frac{\sum_{u \in S \setminus \hat{S}} \delta_u}{|S|} \tag{13}$$

$$> \frac{(\alpha - 1) \times |\hat{S}|}{|S|} + \frac{(\alpha - 1) \times |S \setminus \hat{S}|}{|S|} = \alpha - 1. \tag{14}$$

Inequality (13) is deduced from two main considerations: firstly, the premise that $\lceil \rho_R(\hat{S}) \rceil \geq \alpha$, and secondly, Inequality (9). This analysis supports the conclusion that $\lceil \rho_R(S) \rceil \geq \alpha$.

In summary, Theorem 3.4 is established. □

3.2 The AADSGlobal algorithm

With Section 3.1, we propose the AADSGlobal algorithm which utilizes a binary search on α to identify the approximate anchored densest subgraph. The binary search requires establishing the guess value α's range, where the lower bound is intuitively set to 1, and the upper bound, as inferred from [19], is $\max_{u \in R} d(u)$. The pseudo-code of AADSGlobal is depicted in Algorithm 1. It first initializes the lower bound α_l and the upper bound α_u and then iteratively computes the middle value α_m and invokes ReTest(G, A, R, α_m) to yield a subgraph S^*. If $\lceil \rho_R(S^*) \rceil \geq \alpha$, then α_l is updated to α_m (line 5); otherwise, α_u is adjusted to $\alpha_m - 1$ (line 6). The iterative process terminates once α_l equals α_u, and the algorithm performs ReTest(G, A, R, α_l) again to capture the approximate anchored densest subgraph S^* (line 7).

Example 3.5. Consider the graph G depicted in Figure 1. The AADSGlobal algorithm initializes $\alpha_l = 1$ and $\alpha_u = 5$, and iteratively performs ReTest through binary search. In the first iteration, with the guess value $\alpha_m = 4$, the re-orientation network G_α after executing ReTest$(G, A, R, 4)$ is illustrated in Figure 3, while the orientation $\vec{G}$ is the subgraph of G_α induced by $V = \{v_1, v_2, \ldots, v_{11}\}$. Based on the definition of S, we determine $S_1 = \{v_1, v_3, v_4, v_5, v_6\}$. It is observed that edges in $\vec{G}$ from S_1 to $V \setminus S_1$ point towards $V \setminus S_1$. The R-subgraph density of S_1, $\rho_R(S_1) = 3.2$, satisfies $\lceil \rho_R(S_1) \rceil \geq 4$, leading to an adjustment of α_l to 4. In the subsequent iteration, with $\alpha_m = 5$, invoking ReTest$(G, A, R, 5)$ yields $S_2 = \{v_6\}$, where $\rho_R(S_2) = 0$, thus $\lceil \rho_R(S_2) \rceil = 0 < \alpha = 5$. This result necessitates lowering α by setting $\alpha_u = 4$. With $\alpha_l = \alpha_u$, the iterative process terminates, and the AADSGlobal algorithm performs ReTest$(G, A, R, 4)$ one final time, yielding the subgraph S_1 as the approximate anchored densest subgraph.

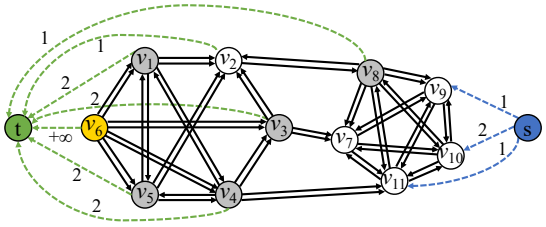


Figure 3: The orientation $\vec{G}$ after invoking ReTest$(G, A, R, 4)$

The following theorem shows the correctness and complexity of Algorithm 1.

Theorem 3.6. *Given an undirected graph $G = (V, E)$, an anchor vertex set $A \subseteq V$, a reference vertex set $R \subseteq V$ satisfying $A \subseteq R$ and $E(R) \neq 0$, Algorithm 1 can output the approximate anchored densest subgraph S^* correctly with the time complexity of $O(m^{1.5} \log d_{max})$.*

Proof. The correctness of Algorithm 1 follows from the correctness of ReTest and the correctness of the binary search. Regarding time complexity, Algorithm 1 executes a total of $(\log d_{\max})$ binary search iterations. In each iteration, ReTest is called to compute the maximum flow, where employing Dinic's algorithm results in a computational cost of $O(m^{1.5})$ time. Thus, the time complexity of Algorithm 1 is $O(m^{1.5} \log d_{\max})$. □

Discussions. The time complexity of our AADSGlobal algorithm for AADS search, $O(m^{1.5} \log d_{\max})$, is significantly lower than the $O(nm \log \frac{n^2}{m})$ held by the ADSGlobal algorithm, tailored for ADS search. This is because AADSGlobal performs a binary search only on integer values, whereas the ADSGlobal algorithm requires a binary search on fractional values. Our later experiments confirm this theoretical result.

4 A NOVEL LOCAL ALGORITHM

Although the AADSGlobal algorithm operates within polynomial time complexity, its scalability is still limited when applied to large-scale graphs. This section presents a novel local algorithm, i.e., AADSLocal, to solve the AADS search problem. A significant innovation within AADSLocal is the introduction of a novel probing method LocalReTest, which enables the localized computation of the max-flow from s to t around R without necessitating access to (or construction of) the entire augmented graph G_α. Below, we first introduce the AADSLocal algorithm and the LocalReTest algorithm and then proceed to the theoretical analysis.

4.1 The framework of AADSLocal algorithm

The AADSGlobal algorithm initializes the orientation $\vec{G}$ by inserting $\langle u, v \rangle$ into $\vec{G}$ twice for each edge $(u, v) \in G$. After this initialization method, it is impossible to determine the bounty of each vertex, thus the algorithm needs to traverse **all** vertices to construct the re-orientation network, and therefore it is not local.

We introduce a novel initialization approach for $\vec{G}$, namely "bi-directional orientation", which entails inserting both $\langle u, v \rangle$ and $\langle v, u \rangle$ for every edge (u, v). This method ensures that the bounty of the vertices in the orientation after initialization is definite and regular since each vertex u receives a precise tally of $d(u)$ incoming edges. Combined with the definition, it follows that vertices within the set $R \setminus A$ have a bounty equal to $d(u)$, whereas all other vertices

Algorithm 2: AADSLocal(G, A, R)

Input: An undirected graph $G = (V, E)$, an anchor vertex set $A \subseteq V$ and a reference vertex set $R \subseteq V$ satisfying $A \subseteq R$ and $E(R) \neq 0$
Output: The AADS S^*

1 $\text{Vol}(R) \leftarrow 0$;
2 $\vec{G} \leftarrow \emptyset$; $V^A \leftarrow \emptyset$;
3 **foreach** $u \in R$ **do**
4 $\quad \text{Vol}(R) \leftarrow \text{Vol}(R) + d(u)$;
5 $\quad V^A \leftarrow V^A \cup \{u\}$;
6 $\quad \text{InitOri}(u)$;
7 $\quad$ **if** $u \in A$ **then** $\delta_u \leftarrow +\infty$;
8 $\quad$ **else** $\delta_u \leftarrow d(u)$;
9 **foreach** $u \in V \backslash R$ **do** $\delta_u \leftarrow 0$;
10 $\alpha_l \leftarrow 1$; $\alpha_u \leftarrow \max_{u \in R} d(u)$;
11 **while** $\alpha_l < \alpha_u$ **do**
12 $\quad \alpha_m \leftarrow \lceil (\alpha_l + \alpha_u + 1)/2 \rceil$;
13 $\quad S^* \leftarrow \text{LocalReTest}(\vec{G}, A, R, \alpha_m, V^A)$;
14 $\quad$ **if** $\lceil \rho_R(S^*) \rceil \geq \alpha$ **then** $\alpha_l \leftarrow \alpha_m$;
15 $\quad$ **else** $\alpha_u \leftarrow \alpha_m - 1$;
16 **return** $\text{LocalReTest}(\vec{G}, A, R, \alpha_l, V^A)$;
17 **Procedure** $\text{InitOri}(u)$
18 **if** $u \notin V^A$ **then**
19 $\quad V^A \leftarrow V^A \cup \{u\}$;
20 $\quad$ **foreach** $v \in N_G(u)$ *and* $v \notin V^A$ **do**
21 $\quad\quad \vec{G} \leftarrow \vec{G} \cup \langle u, v \rangle$; $\vec{G} \leftarrow \vec{G} \cup \langle v, u \rangle$;

in $V \backslash R$ have a bounty of exactly 0. As a result, only the vertices in R are connected to the sink t. Intuitively, the augmenting paths found by the maximum flow algorithm often pass only through vertices near R, while vertices farther from the set R are not considered. This forms the rationale for designing our local algorithm.

With the concept of a bi-directional orientation, we propose the AADSLocal algorithm, detailed in Algorithm 2. This algorithm adopts a binary search framework to compute the AADS, with the following distinctions from AADSGlobal. Firstly, AADSLocal determines if $\lceil \rho_R(S^*) \rceil \geq \alpha$ for a specified guess value α by using LocalReTest (Algorithm 3, detailed in Section 4.2), which locally computes the max-flow from s to t around R in G_α with shortest-path based method (line 13, line 16). Secondly, the orientation $\vec{G}$ is progressively constructed in executing a local search. A set V^A is employed to determine whether u's adjacent edges have been integrated into $\vec{G}$, thereby ensuring that each vertex u is processed only once (line 2). For each vertex $u \notin V^A$, the InitOri procedure is invoked following the bi-directional orientation construction method. It inserts $\langle u, v \rangle$ and $\langle v, u \rangle$ for each edge (u, v) associated with u into $\vec{G}$ (lines 17-21). Before performing LocalReTest, AADSLocal initializes by invoking InitOri for each vertex in R (line 6) and calculating $\text{Vol}(R)$, defined as the sum of the degrees of all vertices in R (line 4). $\text{Vol}(R)$ is crucial for computing the local max-flow of LocalReTest (see Lemma 4.1). Additionally, vertices in A are assigned an infinite bounty, vertices in $R \setminus A$ receive a bounty equal to their degree, and others are allocated a bounty of 0 (lines 7-9).

4.2 Locally checking whether $\lceil \rho_R(S^*) \rceil \geq \alpha$

Before introducing the LocalReTest algorithm for identifying whether $\lceil \rho_R(S^*) \rceil \geq \alpha$, it is essential to delineate the concept of "reverse re-orientation network", which is instrumental in offering a more intuitive elucidation of the algorithm's locality.

Upon specifying a guess value of α, the reverse re-orientation network, represented as $G_\alpha^{-1} = (V_\alpha, E_\alpha^{-1}, c)$, is constructed by augmenting the bi-directional orientation $\vec{G} = (V, \vec{E})$ as follows.

- Add a source vertex s and a sink vertex t to V_α, and add all vertices in V to V_α, i.e., $V_\alpha = V \cup \{s, t\}$.
- For each directed edge $\langle u, v \rangle \in \vec{E}$, add an arc $\langle v, u \rangle$ to E_α^{-1} with capacity 1.
- For each vertex $u \in V$, add an arc $\langle s, u \rangle$ to E_α^{-1}, if $\delta_u > \alpha - 1$ with capacity $\delta_u - (\alpha - 1)$.
- For each vertex $u \in V$, add an arc $\langle u, t \rangle$ to E_α^{-1}, if $\delta_u < \alpha - 1$ with capacity $(\alpha - 1) - \delta_u$.

For a specified guess value α, after performing the max-flow computation on G_α^{-1}, there is no path from s to t, and we can also yield a vertex set S including the vertices either with a bounty of at least α or that can reach another vertex with a bounty of at least α. According to Theorem 3.4, the correctness of the probing method using G_α^{-1} is also established.

Based on the reverse re-orientation network, our goal is to implement a local max-flow calculation starting from R. The probing methodology described in ADSLocal initially constructs a subgraph, followed by a max-flow computation and iterative subgraph expansion until a solution is obtained. The LocalReTest algorithm integrates max-flow computation with subgraph construction to simplify this iterative process. It employs a shortest path-based approach which identifies the shortest augmenting path to compute the max-flow. Upon locating this path, all edges present are reversed in $\vec{G}$. In the computation of max-flow, a critical aspect involves determining the termination of the shortest augmenting path during the BFS procedure. Certainly, the absence of a path from source s to sink t in G_α^{-1} indicates that the search halts at a vertex with bounty less than $\alpha - 1$. To enhance the efficiency of the BFS procedure, we propose Lemma 4.1 that provides a novel approach to determine the path's endpoint.

Lemma 4.1. *Given $\alpha \geq 2$ and a subgraph S with $\rho_R(S) > \alpha - 1$, the degree $d(u)$ for each vertex u in S satisfies $d(u) < Vol(R)$.*

Proof. From Lemma 3.1 in [19], the following relationship is established:

$$\rho_R(S) > \alpha - 1 \Leftrightarrow \text{Vol}(R \setminus S) + E_\times(S, V \setminus S) + (\alpha - 1)|S| < \text{Vol}(R). \tag{15}$$

Given that $\text{Vol}(R \setminus S) \geq 0$ and $\alpha \geq 2$, it holds for any vertex $u \in S$ that:

$$\text{Vol}(R) > E_\times(S, V \setminus S) + |S| \geq E_\times(u, V \setminus S) + |S| \tag{16}$$
$$= d(u) - E_\times(u, S) + |S| \tag{17}$$
$$\geq d(u) + 1. \tag{18}$$

Therefore, for any vertex u in S, $d(u)$ must be less than $\text{Vol}(R)$. □

Equipped with Lemma 4.1, we propose the LocalReTest algorithm detailed in Algorithm 3. This algorithm leverages the bi-directional orientation from the previous iteration to expedite the construction of the reverse re-orientation network. Specifically, it first checks if $\alpha = 1$. If this condition holds, i.e., $\rho_R(\hat{S}) \leq 1$, the algorithm directly outputs the set R as the AADS (line 1). For the case of $\alpha \geq 2$, corresponding to $\rho_R(\hat{S}) > 1$, the LocalReTest algorithm first adds the vertices in V^A with a bounty no less than $\alpha - 1$ into V^N (line 2). After initializing V^N, the algorithm incorporates the adjacent edges of each vertex in V^N into the orientation for α using the InitOri procedure (line 3).

Algorithm 3: LocalReTest($\vec{G}, A, R, \alpha, V^A$)

Input: An undirected graph $G = (V, E)$, an anchor vertex set $A \subseteq V$, a reference vertex set $R \subseteq V$, a non-negative integer α, and the set V^A
Output: A subgraph S satisfying $A \subseteq S$

1 **if** $\alpha = 1$ **then** {$S \leftarrow R$; **return** S};
2 $V^N \leftarrow \{u \in V^A | \delta_u \geq \alpha - 1\}$;
3 **foreach** $u \in V^N$ **do** InitOri(u);
4 Construct the partial reverse re-orientation network $G_{\vec{\alpha}}^{-1}$ according to $\vec{E}$;
5 **while** *true* **do**
6 Perform BFS from the source node s in $G_{\vec{\alpha}}^{-1}$ to try to find a shortest path $\langle s, v_s, \ldots, v_t \rangle$ in $G_{\vec{\alpha}}^{-1}$ where v_t satisfies $\delta_{v_t} < \alpha - 1$ or $d(v_t) \geq \text{Vol}(R)$;
7 For each vertex u visited by the BFS, invoke InitOri(u) to update $\vec{E}$;
8 Maintain $G_{\vec{\alpha}}^{-1}$ according to the updated $\vec{E}$ in the process of BFS;
9 **if** *there exists such a shortest path* $\langle s, v_s, \ldots, v_t \rangle$ *in* $G_{\vec{\alpha}}^{-1}$ **then**
10 Reverse the path $\langle v_s, \ldots, v_t \rangle$ in $\vec{E}$;
11 **if** $v_t \notin V^A$ *and* $d(v_t) < \text{Vol}(R)$ **then** $V^A \leftarrow V^A \cup v_t$;
12 **if** $\delta_{v_t} = \alpha - 1$ *and* $d(v_t) < \text{Vol}(R)$ **then** $V^N \leftarrow V^N \cup v_t$;
13 **else break;**
14 $S \leftarrow \{u \in V | \delta_u \geq \alpha$ or u can reach a vertex v with $\delta_v \geq \alpha\}$;
15 **return** S;

With V^N and $\vec{G}$, the partial reverse re-orientation network, $G_{\vec{\alpha}}^{-1}$, is initialized, obviating the need for construction from scratch (line 4). Next, the LocalReTest algorithm progresses to the construction of the local G_{α}^{-1} based on $G_{\vec{\alpha}}^{-1}$ and the max-flow computation phases, utilizing a computation-while-expanding strategy (lines 5-13). It continuously executes the BFS procedure starting from the source node s, aiming to identify the shortest augmenting path $\langle s, v_s, \ldots, v_t \rangle$, where v_t meets $\delta_{v_t} < \alpha - 1$ or $d(v_t) \geq \text{Vol}(R)$, in accordance with Lemma 4.1. Note that during the BFS process, as each vertex is visited, the InitOri procedure is invoked to extend the orientation, and then $G_{\vec{\alpha}}^{-1}$ is updated (lines 6-8). Upon discovering a shortest augmenting path $\langle s, v_s, \ldots, v_t \rangle$, all edges on this path, except for the one connected to s, are saturated. Consequently, these saturated edges are reversed in the bi-directional orientation, and updates are made to V^A and V^N (lines 10-12). The absence of augmenting paths in $G_{\vec{\alpha}}^{-1}$ signifies the completion of the max-flow computation. At this juncture, Algorithm 3 outputs the vertex set S, which comprises vertices either possessing a bounty of at least α or capable of reaching another vertex with a bounty of at least α.

4.3 Analysis of correctness and locality

Utilizing Theorem 3.4 and Lemma 4.1, we ascertain the correctness of LocalReTest, thereby ensuring that the AADSLocal algorithm correctly outputs the AADS. Next, we establish the locality of the AADSLocal algorithm by demonstrating that LocalReTest is local. LocalReTest constructs a local reverse re-orientation network $G_{\vec{\alpha}}^{-1}$, which we extend to be complete for ease of analysis. Specifically, we introduce $\langle u, v \rangle$ and $\langle v, u \rangle$ into $G_{\vec{\alpha}}^{-1}$ if u and v are not connected in $G_{\vec{\alpha}}^{-1}$ but $(u, v) \in E$. Within the complete G_{α}^{-1}, let l be the distance between s and t and $\text{dist}(s, u)$ be the distance between s and u. Denote V_i by the vertex set defined as $V_i = \{u \in V \mid \text{dist}(s, u) \leq i\}$. The following lemma is established.

Lemma 4.2. *In* G_{α}^{-1}, $|V_{l-1}| \leq \frac{Vol(R)}{\alpha} + |A| \leq 2Vol(R)$ *and* $|V_l| \leq 2Vol^2(R)$.

Proof. In G_{α}^{-1}, for each vertex u in V, we have $\delta_u \geq 0$, and further, we establish that:

$$\sum_{u \in V_{l-1} \setminus A} \delta_u \leq \sum_{u \in V \setminus A} \delta_u = \sum_{u \in V \setminus A} \vec{d}(u) - \sum_{u \in V \setminus A} d(u) \tag{19}$$

$$\leq \sum_{u \in V} \vec{d}(u) - \sum_{u \in V \setminus R} d(u) \tag{20}$$

$$= 2|E| - \sum_{u \in V \setminus R} d(u) \tag{21}$$

$$= \sum_{u \in V} d(u) - \sum_{u \in V \setminus R} d(u) = \text{Vol}(R). \tag{22}$$

Moreover, each vertex in $V_{l-1} \setminus A$ satisfies $\delta_u \geq \alpha$; otherwise, the distance between s and t would not be greater than $l - 1$. Thus, we find that:

$$\sum_{u \in V_{l-1} \setminus A} \delta_u \geq \sum_{u \in V_{l-1} \setminus A} \alpha \geq \alpha \times |V_{l-1} \setminus A| = \alpha \times (|V_{l-1}| - |A|). \tag{23}$$

By combining the above inequalities, we can conclude that $|V_{l-1}| \leq \frac{\text{Vol}(R)}{\alpha} + |A| \leq 2\text{Vol}(R)$. Further, according to Lemma 4.1 and Lemma 4.2, $|V_l| < 2\text{Vol}^2(R)$ is established. □

With the established Lemma 4.2, we analyze the time complexity of LocalReTest using the Dinic algorithm as a paradigmatic example of a shortest path-based max-flow computation technique. Notably, the implementation of the Dinic algorithm [56] herein includes an optimization: during the BFS process, when exploring the adjacent edges of a vertex u in V_{l-1}, the search will terminate upon reaching the edge (u, t). This optimization halts the BFS prematurely, preventing further traversal of other vertices.

Theorem 4.3. *The* LocalReTest *algorithm, which employs the Dinic algorithm for maximum flow computation, exhibits a time complexity of* $O(Vol^3(R))$.

Proof. We begin by examining the time complexity of BFS and DFS when computing the maximum flow using the Dinic algorithm in LocalReTest. In the BFS procedure, the queue contains at most the vertices in $s \cup V_{l-1}$, each with a degree less than or equal to $\text{Vol}(R)$. Consequently, the number of traversed edges is bounded by $|s \cup V_{l-1}| \times \text{Vol}(R)$, falling within $O(\text{Vol}^2(R))$. Therefore, the BFS procedure consumes $O(\text{Vol}^2(R))$ time. Similarly, the DFS phase involves the traversal of the same number of vertices and edges as the BFS phase, resulting in a time complexity for the DFS process also of $O(\text{Vol}^2(R))$.

Next, we determine the number of iterations required by the Dinic algorithm, where each iteration consists of one BFS and one DFS. Given that $|V_{l-1}| \leq 2\text{Vol}(R)$, hence $l \subseteq O(\text{Vol}(R))$. Coupled with the fact that l increases by at least one after each iteration, the Dinic algorithm necessitates a total of $O(\text{Vol}(R))$ rounds of iterations. Consequently, the time complexity of the LocalReTest algorithm is $O(\text{Vol}^3(R))$. □

Theorem 4.3 show that the time complexity of Algorithm 3 is bounded by a polynomial in terms of $O(\text{Vol}(R))$, which is independent on the size of graph G. With this foundation, the AADSLocal algorithm is strongly local, with a time complexity of $O(\log d_{\max} \times \text{Vol}^3(R))$.

Discussions. We analyze the time complexity of the ADSLocal algorithm for searching ADS [19]. First, ADSLocal needs to perform $(\log n)$ local max-flow computations. Second, the local max-flow computation method in ADSLocal iteratively extends the subgraph and performs computations. The number of vertices and edges

of the subgraph in the last iteration are both $O(\mathrm{Vol}^2(R))$, causing the complexity of the last iteration to reach $O(\mathrm{Vol}^4(R))$ using the state-of-the-art max-flow method. Therefore, the time complexity of ADSLocal is $\geq O(\log n \times \mathrm{Vol}^4(R))$. In contrast, the time complexity of our AADSLocal algorithm for searching AADS is $O(\log d_{\max} \times \mathrm{Vol}^3(R))$, which is significantly lower than that of the ADSLocal algorithm. Our subsequent experiments confirm this theoretical analysis.

5 THE MAINTAINANCE OF AADS

This section proposes efficient algorithms for maintaining the approximate anchored densest subgraph amidst updates to the graph. Initially, an anchored densest subgraph update theorem is presented, based on which we propose the BasicIns and BasicDel algorithms in response to edge insertions and deletions, respectively. Moreover, the improved algorithms, i.e., ImproveIns and ImproveDel, are developed, with a focus on maintaining an unreversible orientation $\vec{G}$ to preserve the AADS.

5.1 ADS update theorem and basic algorithms

We first present the anchored densest subgraph update theorem, which forms the foundation of the proposed BasicIns and BasicDel algorithms for maintaining AADS.

Theorem 5.1. *After a single edge insertion (resp., deletion) in G, $\lceil \rho_R(\hat{S}) \rceil$ either remains unchanged, increases by 1, or decreases by 1.*

Proof. We focus on the scenario of edge insertion; the analogous case of edge deletion is omitted for brevity. Upon inserting an edge (u, v), the variations in $\lceil \rho_R(\hat{S}) \rceil = \lceil \frac{2|E(\hat{S})| - \sum_{v \in \hat{S} \setminus R} d_G(v)}{|\hat{S}|} \rceil$ are as follows.

(1) $u, v \notin \hat{S}$: $\rho_R(\hat{S})$ remains constant, and $\lceil \rho_R(\hat{S}) \rceil$ is unaffected.
(2) $u, v \in \hat{S} \setminus R$: $\rho_R(\hat{S})$ remains constant, and $\lceil \rho_R(\hat{S}) \rceil$ is unaffected.
(3) $u, v \in \hat{S} \cap R$: $\rho_R(\hat{S})$ becomes $\rho_R(\hat{S}) + \frac{2}{|\hat{S}|}$, leading to $\lceil \rho_R(\hat{S}) \rceil$ being either unchanged or incremented by 1.
(4) $u \in \hat{S} \cap R$ and $v \in \hat{S} \setminus R$: $\rho_R(\hat{S})$ becomes $\rho_R(\hat{S}) + \frac{1}{|\hat{S}|}$, leading to $\lceil \rho_R(\hat{S}) \rceil$ being either unchanged or incremented by 1.
(5) $u \in \hat{S} \cap R$ and $v \notin \hat{S}$: $\rho_R(\hat{S})$ remains constant, and $\lceil \rho_R(\hat{S}) \rceil$ is unaffected.
(6) $u \in \hat{S} \setminus R$ and $v \notin \hat{S}$: $\rho_R(\hat{S})$ becomes $\rho_R(\hat{S}) - \frac{1}{|\hat{S}|}$, resulting in $\lceil \rho_R(\hat{S}) \rceil$ being either unchanged or reduced by 1.

Consequently, following the insertion of an edge into G, $\lceil \rho_R(\hat{S}) \rceil$ either remains unchanged, increases by 1, or decreases by 1. □

According to Theorem 5.1, $\lceil \rho_R(\hat{S}) \rceil$ remains constant or alters by only one unit following the insertion or deletion of an edge, the same applies to $\lceil \rho_R(S^*) \rceil$. Thus, merely two checks using ReTest are required to identify the updated AADS.

Based on the above rationale, we develop the BasicIns and BasicDel algorithms for edge insertion and deletion. The pseudo-code of BasicIns is delineated in Algorithm 4. Initially, the algorithm inserts the edge (u, v) into G, and calculates the current round-up of the maximum R-subgraph density $\alpha \leftarrow \lceil \rho_R(S^*) \rceil$ (line 2). Subsequently, BasicIns invokes ReTest with the parameter α, updating the subgraph S^* (line 3). The algorithm then checks whether

Algorithm 4: BasicIns$(G, A, R, S^*, (u, v))$

Input: An undirected graph $G = (V, E)$, an anchor vertex set A, a reference vertex set R, the AADS S^* and an edge (u, v) to be inserted
Output: The updated AADS S^*

```
1  E ← E ∪ {(u, v)};
2  α ← ⌈ρ_R(S*)⌉;
3  S* ← ReTest(G, A, R, α);
4  if ⌈ρ_R(S*)⌉ < α then                    // Check if ⌈ρ_R(Ŝ)⌉ decreases by 1
5      S* ← ReTest(G, A, R, α − 1);
6  else
7      S̈ ← ReTest(G, A, R, α + 1);
8      if ⌈ρ_R(S̈)⌉ ≥ α + 1 then             // Check if ⌈ρ_R(Ŝ)⌉ increases by 1
9          S* ← ReTest(G, A, R, α + 1);
10 return S*;
```

$\lceil \rho_R(S^*) \rceil < \alpha$ to determine if $\lceil \rho_R(\hat{S}) \rceil$ decreases by 1 (line 4). If so, BasicIns performs ReTest with guess value $\alpha - 1$ to update S^* as the approximate anchored densest subgraph after inserting edge (u, v); otherwise, the algorithm needs to compute the subgraph $\ddot{S}$ to check if $\lceil \rho_R(\hat{S}) \rceil$ increases by 1 (lines 7-8), and then update S^* if it does increase. For the BasicDel algorithm, which maintains the AADS for edge deletion, its pseudo-code necessitates merely a modification of the first line in Algorithm 4 to $E \leftarrow E - \{u, v\}$. We omit the pseudo-code of BasicDel.

The correctness of the BasicIns and BasicDel algorithms is affirmed by Theorem 3.4 and Theorem 5.1. The following theorem shows their time complexity.

Theorem 5.2. *The time complexity of* BasicIns *and* BasicDel *are $O(m^{1.5})$ because they only need to perform* ReTest *twice.*

5.2 The improved algorithms

The primary limitation of the basic algorithms lies in their dependency on reconstructing the orientation network for an updated graph and performing a max-flow computation. In this subsection, we propose two enhanced algorithms, ImproveIns and ImproveDel, designed to maintain the AADS by preserving an "unreversible orientation" contingent on a parameter α. For a given integer α, we define "reversible path" and "unreversible orientation" as follows.

Definition 5.3. (Reversible Path) Given a graph $G = (V, E)$, its orientation $\vec{G} = (V, \vec{E})$ and an integer α, a path $v_s \rightsquigarrow v_t$ is reversible if: (1) $\delta_{v_s} < \alpha - 1$ and $\delta_{v_t} > \alpha - 1$; or (2) $\delta_{v_s} < \alpha$ and $\delta_{v_t} > \alpha$.

Definition 5.4. (Unreversible Orientation) Given a graph $G = (V, E)$, its orientation $\vec{G} = (V, \vec{E})$ and an integer α, $\vec{G}$ is an unreversible orientation if there is no reversible path in $\vec{G}$.

The rationale behind defining the unreversible path and the unreversible orientation originates from Theorem 3.4. Note that $\lceil \rho_R(S_\alpha) \rceil \geq \alpha$ and $\lceil \rho_R(S_{\alpha+1}) \rceil < \alpha + 1$ hold if and only if $\alpha = \lceil \rho_R(\hat{S}) \rceil$, where S_* is the subgraph returned by invoking the ReTest algorithm with $*$ as the guess value. It can be proven that S_* comprises the vertices that either have a bounty of at least $*$ or can reach another vertex with a bounty of no less than $*$ in an unreversible orientation $\vec{G}$. Notably, S_α consists of vertices within the AADS. By maintaining an unreversible orientation $\vec{G}$ and two subgraphs S_α and $S_{\alpha+1}$, we can determine whether $\lceil \rho_R(\hat{S}) \rceil$ changes due to the insertion or deletion of edges and thus maintain AADS. This constitutes the central idea of our ImproveIns and ImproveDel algorithms.

Algorithm 5: ImproveIns($\vec{G}, A, R, S_\alpha, S_{\alpha+1}, (u,v)$)

Input: A graph $G=(V,E)$, an orientation $\vec{G}=(V,\vec{E})$, an anchor vertex set A, a reference vertex set R, the subgraph S_α, $S_{\alpha+1}$, and the edge (u,v) to be inserted
Output: The updated S_α, $S_{\alpha+1}$, and orientation $\vec{G}$

1 $E \leftarrow E \cup \{(u,v)\}$;
2 **if** $u \notin R$ **then** DecBounty(u);
3 **if** $v \notin R$ **then** DecBounty(v);
4 **if** ($u \in S_{\alpha+1}$ *and* $v \notin S_{\alpha+1}$) *or* ($u \in S_\alpha$ *and* $v \notin S_\alpha$) **then** $p \leftarrow u$; $q \leftarrow v$;
5 **else** $p \leftarrow v$; $q \leftarrow u$;
6 $\vec{G} \leftarrow \vec{G} \cup \{\langle p,q\rangle\}$;
7 IncBounty(q);
8 Repeat lines 4-7;
9 **if** $\lceil \rho_R(S_{\alpha+1}) \rceil \geq \alpha+1$ **then** $S_\alpha \leftarrow S_{\alpha+1}$; $S_{\alpha+1} \leftarrow$LocalReTest($\alpha+2$); α++;
10 **else if** $\lceil \rho_R(S_\alpha) \rceil < \alpha$ **then** $S_{\alpha+1} \leftarrow S_\alpha$; $S_\alpha \leftarrow$LocalReTest($\alpha-1$); α−−;
11 **return** ($S_\alpha, S_{\alpha+1}, \vec{G}$)
12 **Procedure** IncBounty(x)
13 δ_x++;
14 **if** $x \in V \setminus S_\alpha$ *and* $\delta_x = \alpha$ **then**
15 **if** *a path* $y \rightsquigarrow x$ *can be found,* $\delta_y \leq \alpha-2$ **then** Reverse the path; δ_x−−; δ_y++;
16 **else** $S_\alpha \leftarrow S_\alpha \cup \{x\} \cup \{y \mid y$ can reach $x\}$;
17 **else if** $x \in S_\alpha \setminus S_{\alpha+1}$ *and* $\delta_x = \alpha+1$ **then**
18 **if** *a path* $y \rightsquigarrow x$ *can be found,* $\delta_y \leq \alpha-1$ **then** Reverse the path; δ_x−−; δ_y++;
19 **else** $S_{\alpha+1} \leftarrow S_{\alpha+1} \cup \{x\} \cup \{y \mid y$ can reach $x\}$;
20 **Procedure** DecBounty(x)
21 δ_x−−;
22 **if** $x \in S_{\alpha+1}$ **then**
23 **if** $\delta_x = \alpha-1$ **then** Reverse a path $x \rightsquigarrow y$ where $\delta_y \geq \alpha+1$; δ_y−−; δ_x++;
24 $S_{\alpha+1} \leftarrow \{y \mid \delta_y \geq \alpha+1$ or y can reach any vertex with $\delta \geq \alpha+1\}$;
25 **else if** $x \in S_\alpha \setminus S_{\alpha+1}$ **then**
26 **if** $\delta_x = \alpha-2$ **then** Reverse a path $x \rightsquigarrow y$, where $\delta_y \geq \alpha$, δ_y−−, δ_x++;
27 $S_\alpha \leftarrow \{y \mid \delta_y \geq \alpha$ or y can reach any vertex with $\delta \geq \alpha\}$;

According to the main idea of the improved algorithms, the unreversible orientation $\vec{G}$ and two subgraphs S_α and $S_{\alpha+1}$ must be taken as input parameters of the ImproveIns and ImproveDel algorithms. We use the following method to compute the inputs $\vec{G}$, S_α and $S_{\alpha+1}$. First, we perform either AADSGlobal or AADSLocal with $\alpha = \lceil \rho_R(S^*) \rceil$, ensuring that there is no path $v_s \rightsquigarrow v_t$ with $\delta_{v_s} < \alpha-1$ and $\delta_{v_t} > \alpha-1$ in $\vec{G}$. Subsequently, we construct the re-orientation network based on $\vec{G}$ and $\alpha+1$, and then invoke ReTest again to ensure no path $v_s \rightsquigarrow v_t$ exists with $\delta_{v_s} < \alpha$ and $\delta_{v_t} > \alpha$ in $\vec{G}$. With the unreversible orientation $\vec{G}$, two subgraphs S_α and $S_{\alpha+1}$ can be derived easily. Below, we introduce our improved algorithms in detail.

The ImproveIns **algorithm for edge insertion.** When an edge (u,v) is inserted, it is imperative to insert two directed edges into $\vec{G}$ and maintain $\vec{G}$ as an unreversible orientation. To streamline the analysis, we present a three-step procedure for edge insertion, detailed as follows: (1) insert (u,v) into the graph G; (2) insert a directed edge into the orientation $\vec{G}$; (3) repeat step 2. This procedure ensures that the bounty of any given vertex changes by only one unit at each step, simplifying the identification of reversible paths and thereby aiding in maintaining $\vec{G}$.

The ImproveIns algorithm is outlined in Algorithm 5, which accepts three specific parameters: the unreversible orientation $\vec{G}$, S_α and $S_{\alpha+1}$. The main idea of ImproveIns is to maintain the AADS by ensuring that $\vec{G}$ remains an unreversible orientation throughout each stage of the three-step process. In Step 1 (lines 1-3), ImproveIns inserts (u,v) into G, which increments the degrees of u and v by one each, potentially altering their bounties δ_u and δ_v. Taking u as an example, if $u \in R$, then by Definition 3.2, δ_u remains unchanged, thus preserving the irreversibility of $\vec{G}$. Alternatively, if $u \notin R$, the DecBounty procedure is invoked to reduce δ_u by 1 and make $\vec{G}$ be unreversible again (line 2). In DecBounty, when x is in $S_{\alpha+1}$, it checks if δ_x decreases to $\alpha-1$. If it does, since x can reach a vertex y in $\vec{G}$ whose bounty is at least $\alpha+1$, we reverse the reversible path $x \rightsquigarrow y$ and update δ_y and δ_x (line 23). Importantly, DecBounty updates the set $S_{\alpha+1}$, regardless of whether δ_x equals $\alpha-1$ (line 24). For the case of $x \in S_\alpha \setminus S_{\alpha-1}$, if $\delta_x = \alpha-2$, a reversible path $x \rightsquigarrow y$ is identified as x can reach a vertex y with bounty no less than α. This path is then reversed, and updates are applied to both δ_y and δ_x (line 26). Note that the DecBounty procedure maintains the set S_α once $x \in S_\alpha \setminus S_{\alpha-1}$ (line 27). For all other cases, it is clear that $\vec{G}$ remains unreversible, as there are no reversible paths present in the input orientation $\vec{G}$.

In step 2 of ImproveIns (lines 4-7), a directed edge $\langle p,q\rangle$ is inserted into $\vec{G}$, with the orientation determined by the membership of u and v in sets $S_{\alpha+1}$ or S_α. This configuration ensures that after the edge's insertion, $\vec{G}$ hosts at most one reversible path. The insertion allows vertex p to retain its bounty δ_p, simultaneously increasing the in-degree and the bounty of vertex q, regardless of its presence in set R. Following this, the ImproveIns algorithm engages the IncBounty(q) procedure to preserve the irreversibility of $\vec{G}$ by identifying a potential reversal path (lines 12-19). In IncBounty, after increasing x's bounty by 1, if x belongs to $V \setminus S_\alpha$ and δ_x reaches α, the procedure checks for a reversible path $y \rightsquigarrow x$ where $\delta_y \leq \alpha-2$. If such a path is identified, it is reversed, and both δ_x and δ_y are updated to restore the irreversibility of $\vec{G}$. Conversely, S_α is maintained as x now has a bounty equal to α. For the case where x is in $S_\alpha \setminus S_{\alpha+1}$ and δ_x equals $\alpha+1$, if the procedure identifies a reversible path $y \rightsquigarrow x$ with $\delta_y \leq \alpha-1$, then the path is reversed and δ_x and δ_y are adjusted accordingly to make $\vec{G}$ unreversible again. Otherwise, $S_{\alpha+1}$ is maintained since the bounty of x now equals $\alpha+1$. For other cases where these specific conditions do not apply, $\vec{G}$ remains irreversible. Step 3 repeats the operations of step 2 to continuously uphold the irreversibility of $\vec{G}$ (line 8).

After the three-step procedure, ImproveIns ensures that $\vec{G}$ retains an unreversible orientation upon edge insertion, and then it maintains the sets S_α and $S_{\alpha+1}$ as follows. If $\lceil \rho_R(S_{\alpha+1}) \rceil \geq \alpha+1$, indicating that $\lceil \rho_R(\hat{S}) \rceil = \lceil \rho_R(S_{\alpha+1}) \rceil = \alpha+1$, and then S_α is updated to $S_{\alpha+1}$, and ImproveIns performs LocalReTest($\alpha+2$) to re-calculate $S_{\alpha+1}$ and increase α by 1 (line 9). While if $\lceil \rho_R(S_\alpha) \rceil < \alpha$, meaning $\lceil \rho_R(\hat{S}) \rceil = \lceil \rho_R(S_{\alpha+1}) \rceil = \alpha-1$, $S_{\alpha+1}$ is adjusted to S_α, and LocalReTest is employed to manage $S_{\alpha-1}$ and decrease α by 1 (line 10). Ultimately, the ImproveIns algorithm outputs S_α, $S_{\alpha+1}$ and $\vec{G}$ where S_α represents the AADS.

Below, we establish the correctness of the ImproveIns algorithm by demonstrating the correctness of DecBounty and IncBounty.

Theorem 5.5. *The* DecBounty *procedure can ensure that* $\vec{G}$ *remains unreversible and that the subgraphs* S_α *and* $S_{\alpha+1}$ *are correctly maintained.*

Proof. Let δ_x^- and δ_x denote the bounty of x before and after the undirected edge insertion, respectively. We first establish the correctness of $x \in S_{\alpha+1}$ under two scenarios: (1) $\delta_x^- = \alpha$; and (2) $\delta_x^- = \alpha+1$. In case (1), the bounty of x drops to $\alpha-1$, i.e., $\delta_x = \alpha-1$. According to the definition of $S_{\alpha+1}$, x can reach a vertex y in $\vec{G}$

that maintains a bounty of at least $\alpha + 1$. At this point, reversing this reversible path $x \rightsquigarrow y$ increases the in-degree of x by 1 and decreases the in-degree of y by 1, setting $\delta_x = \alpha$ and $\delta_y \geq \alpha$. Thus, $\vec{G}$ is an unreversible orientation, as the input $\vec{G}$ is unreversible. In case (2), δ_x equals α, and by Definition 5.3, there remains no reversible path in $\vec{G}$; otherwise, this would contradict the premise that the input $\vec{G}$ is unreversible. As for the subgraphs S_α and $S_{\alpha+1}$, δ_y potentially drops from $\alpha + 1$ to α in case (1) and δ_x definitely decreases from $\alpha+1$ to α in case (2). Thus, updating $S_{\alpha+1}$ is necessary in both cases. However, since no new vertex has a bounty no less than α, S_α does not need to be updated.

We now consider the case of $x \in S_\alpha \backslash S_{\alpha+1}$ which splits into two subcases: (1) $\delta_x^- = \alpha$; and (2) $\delta_x^- = \alpha - 1$. In subcase (1), δ_x decreases to $\alpha - 1$. Since x cannot reach a vertex with a bounty no less than $\alpha + 1$, there remains no reversible path in $\vec{G}$, and thus $\vec{G}$ remains an unreversible orientation; In subcase (2), δ_x drops to $\alpha - 2$. Due to $x \in S_\alpha \backslash S_{\alpha+1}$, x can reach a vertex y whose bounty is not less than α, signifying the presence of a reversible path $x \rightsquigarrow y$. Reversing this path and adjusting δ_y by decreasing it by 1 and increasing δ_x by 1 ensures that $\vec{G}$ remains unreversible; otherwise, the input $\vec{G}$ would not have been irreversible, causing a contradiction. Concerning S_α and $S_{\alpha+1}$, δ_x definitely decreases from $\alpha + 1$ to α in case (1), and δ_y potentially reduces from α to $\alpha - 1$ in case (2), necessitating the update to S_α. However, since no vertex newly has a bounty of at least $\alpha + 1$, $S_{\alpha+1}$ remains unchanged.

For other cases not discussed, it is evident that $\vec{G}$ is an unreversible orientation and that S_α and $S_{\alpha+1}$ remain unchanged.

In conclusion, Theorem 5.5 is established. □

Theorem 5.6. *The* IncBounty *procedure can ensure that $\vec{G}$ remains unreversible and that the subgraphs S_α and $S_{\alpha+1}$ are correctly maintained.*

Proof. Let δ_x^- and δ_x denote the bounty of x before and after the directed edge insertion, respectively. In IncBounty, the bounty of x increases by 1, i.e., $\delta_x = \delta_x^- + 1$. We first establish the correctness of $x \in V \backslash S_\alpha$. If δ_x does not reach α, by Definition 5.3, $\vec{G}$ remains free of reversible paths, aligning with the initial condition of irreversibility. On the other hand, if δ_x equals α, we consider the possibility of a reversible path terminating at x. Should no such path exist, $\vec{G}$ conclusively remains irreversible. Under this condition, x qualifies as a new vertex within S_α due to its adjusted bounty, thus necessitating an update to S_α by including x and all vertices from which x is reachable. Since α remains constant, $S_{\alpha+1}$ does not require any modifications. Conversely, if a reversible path $y \rightsquigarrow x$ with $\delta_y \leq \alpha - 2$ is discovered, it necessitates a contradiction-based proof to validate the absence of reversible paths after its reversal. Assuming a reversible path $y \rightsquigarrow x$ is reversed, yet another reversible path $s \rightsquigarrow t$ persists. Necessarily, $s \rightsquigarrow t$ and $y \rightsquigarrow x$ overlap. Let z be the last overlapping vertex between these paths. Prior to the reversal, we have $\delta_s < \alpha - 1, \delta_t > \alpha - 1, \delta_y \leq \alpha - 2, \delta_x = \alpha - 1$. The existence of a reversible path $y \rightsquigarrow z \rightsquigarrow t$ directly conflicts with the assertion of the input $\vec{G}$ being irreversible. Hence, after the reversal of $y \rightsquigarrow x$, $\vec{G}$ is unreversible. The reversal also makes the bounties of x and y decrease by 1 and increase by 1, respectively. Thus the subgraphs S_α and $S_{\alpha+1}$ remain unchanged as α does not change.

For the case where x is in $S_\alpha \backslash S_{\alpha+1}$, we distinguish two subcases: (1) $\delta_x = \alpha$; and (2) $\delta_x = \alpha + 1$. In subcase (1), δ_x becomes α, by Definition 5.3, no reversible path remains in $\vec{G}$; otherwise, this would contradict the premise that the input $\vec{G}$ is unreversible. In subcase (2), we evaluate the existence of a reversible path $y \rightsquigarrow x$ with $\delta_y \leq \alpha - 1$. If no such path exists, $\vec{G}$ remains irreversible. x becomes a new vertex with a bounty equal to $\alpha + 1$, prompting an update to $S_{\alpha+1}$ to include x and all vertices that can reach it, while S_α remains unchanged. Conversely, discovering a reversible path ending at x prompts a contradiction-based proof to affirm the absence of reversible paths post-reversal. Hence, following the reversal of $y \rightsquigarrow x$, $\vec{G}$ is unreversible. The reversal adjusts the bounties, decreasing δ_x by 1 and increasing δ_y by 1. Thus, the subgraphs S_α and $S_{\alpha+1}$ remain unchanged as α does not change.

For other cases not discussed, it is clear that $\vec{G}$ is an unreversible orientation and that S_α and $S_{\alpha+1}$ remain unchanged.

In summary, Theorem 5.6 is established. □

Algorithm 6: ImproveDel($\vec{G}, A, R, S_\alpha, S_{\alpha+1}, (u, v)$)

Input: A graph $G = (V, E)$, an orientation $\vec{G} = (V, \vec{E})$, an anchor vertex set A, a reference vertex set R, the subgraph S_α, $S_{\alpha+1}$, and the edge (u, v) to be deleted
Output: The updated S_α, $S_{\alpha+1}$, and orientation $\vec{G}$

```
1  E ← E\{(u, v)};
2  if u ∉ R then IncBounty (u);
3  if v ∉ R then IncBounty (v);
4  ⟨x, y⟩ ← an directed edge in G⃗ between u and v;
5  DecBounty (y,⟨x, y⟩);
6  Repeat lines 4-5;
7  if ⌈ρ_R(S_{α+1})⌉ ≥ α + 1 then S_α ← S_{α+1}; S_{α+1} ←LocalReTest(α + 2); α++;
8  else if ⌈ρ_R(S_α)⌉ < α then S_{α+1} ← S_α; S_α ←LocalReTest(α − 1); α− −;
9  return (S_α, S_{α+1}, G⃗)
10 Procedure IncBounty(x) The same as the IncBounty procedure in ImproveIns;
11 Procedure DecBounty(x, ⟨x, y⟩)
12 δ_x− −;
13 if x ∈ S_{α+1} then
14     if δ_x = α − 1 then Reverse a path x ⇝ y, where δ_y ≥ α + 1; δ_y− −; δ_x++;
15     G⃗ ← G⃗ \ ⟨x, y⟩;
16     S_{α+1} ← {y|δ_y ≥ α + 1 or y can reach any vertex with δ ≥ α + 1};
17 else if x ∈ S_α \ S_{α+1} then
18     if δ_x = α − 2 then Reverse a path x ⇝ y, where δ_y ≥ α; δ_y− −; δ_x++;
19     G⃗ ← G⃗ \ ⟨x, y⟩;
20     S_α ← {y|δ_y ≥ α or y can reach any vertex with δ ≥ α};
21 else G⃗ ← G⃗ \ ⟨x, y⟩;
```

With Theorem 5.5 and Theorem 5.6, we affirm the correctness of the ImproveIns algorithm. Concerning time complexity, ImproveIns requires $O(\mathrm{Vol}^3(R))$ time in the worst-case scenario due to its reliance on LocalReTest. However, ImproveIns exhibits superior efficiency in practical settings for several reasons. Firstly, the BFS process terminates upon encountering vertices with a bounty less than or equal to a specific threshold, limiting the search domain to vertices with bounties exceeding this threshold. Typically, these vertices are located within the graph's densest regions, which are relatively small in practical scenarios, thus offering a restricted search field for BFS. Secondly, LocalReTest is activated only when there is a change in $\lceil \rho_R(\hat{S}) \rceil$. Given that such changes are rare in real-world applications, the execution of LocalReTest becomes infrequent.

The ImproveDel **algorithm for edge deletion.** The deletion of an edge (u, v) necessitates the removal of two directed edges from $\vec{G}$. Analogous to ImproveIns, the ImproveDel algorithm involves a three-step procedure for edge deletion, summarized as follows: (1)

Table 2: Graph datasets statistics

Name	Type	n	m
amazon	Product network	334,863	925,872
citeseer	Citation network	384,414	1,736,145
hyves	Social network	1,402,674	2,777,419
youtube	Social network	1,134,890	2,987,624
trec	Hyperlink network	1,601,788	6,679,248
flixster	Social network	2,523,387	7,918,801
dblp	Citation network	1,653,767	8,159,739
skitter	Computer network	1,696,416	11,095,299
indian	Hyperlink network	1,382,868	13,591,473
pokec	Social network	1,632,804	22,301,964
usaroad	Road network	23,947,347	28,854,312
livejournal	Social network	3,997,962	34,681,189
orkut	Social network	3,072,441	117,185,083
weibo	Social network	58,655,849	261,321,033

delete (u, v) from the graph G; (2) delete a directed edge between u and v from the orientation $\vec{G}$; (3) repeat step 2.

The ImproveDel algorithm, as outlined in Algorithm 6, employs a structured three-step procedure to maintain the AADS by ensuring that $\vec{G}$ remains in an unreversible orientation throughout (lines 1-6). In step 1 (lines 1-3), the deletion of (u, v) from G decreases the degrees of u and v by one, potentially affecting their bounties, δ_u and δ_v. Take vertex u as an example, if $u \in R$, $\vec{G}$ continues to exhibit an unreversible orientation. Conversely, if u falls outside R, δ_u experiences an increment of 1, triggering the IncBounty procedure to preserve $\vec{G}$'s irreversibility (line 2). During step 2 (lines 4-5), suppose that the deleted edge is $\langle u, v\rangle$. Subsequent to this removal, u's bounty remains unchanged, whereas v's indegree reduction decreases δ_v, regardless of its presence in R, necessitating the DecBounty procedure to maintain $\vec{G}$. Step 3 repeats step 2, so that an unreversible orientation $\vec{G}$ can also be obtained (line 6). Finally, the ImproveDel algorithm can derive the AADS based on the unreversible orientation $\vec{G}$ like ImproveIns (lines 7-8).

The correctness and time complexity of ImproveDel is similar to that of ImproveIns.

6 EXPERIMENTS

6.1 Experiment settings

Different algorithms. We implement the proposed algorithms, specifically AADSGlobal (Algorithm 1) and AADSLocal (Algorithm 2), for the problem of approximate anchored densest subgraph search. For comparative analysis, algorithms for searching the anchored densest subgraph presented in [19], namely, ADSGlobal and ADSLocal, are also implemented. For dynamic graphs, we implement the algorithms for maintaining the approximate anchored densest subgraph, which includes BasicIns (Algorithm 4) and ImproveIns (Algorithm 5) for edge insertion, alongside BasicDel and ImproveDel (Algorithm 6) for edge deletion. In our experiments, we also compare these maintenance algorithms with the method that uses ADSLocal to compute from scratch.

Datasets. We select 14 distinct datasets from two sources: the Network Repository (https://networkrepository.com/) and the Koblenz Network Collection (http://konect.cc/), detailed comprehensively in Table 2. This collection spans an extensive variety of network categories, including product, citation, social, hyperlink, road, and computer networks. For the purposes of our study, all graphs are treated as undirected and unweighted.

Experiment environment. All algorithms are coded using C++ and compiled with the GCC compiler using O3 optimization. Experiments are conducted on a PC running a Linux operating system, equipped with a 2.2 GHz AMD 3995WX 64-Core CPU and 256 GB of memory. The cutoff time was 1,000 seconds for each query.

Seed vertex set generation. We adopt a query seed set generation approach used in [19]. The process begins with randomly selecting a vertex, denoted by v, from the overall set of vertices. Subsequently, A is constituted by randomly selecting a predefined number of vertices (default is 8) from v's 1-hop and 2-hop neighbors, explicitly including v itself in A. R is then generated for each vertex u in A through several (default: 3) random walks of a specified length (default: 2 steps), aimed at identifying additional members for R. Detailed insights into the query set generation method can be found in [19].

6.2 Performance studies

Exp-1: The runtime and memory of different algorithms. We generate 100 queries for each dataset and apply four algorithms: ADSGlobal, ADSLocal, AADSGlobal, and AADSLocal. The average running time and memory consumption of these algorithms across datasets are illustrated in Figure 4 and Figure 5, respectively.

The results show that ADSGlobal incurs the highest runtime and memory usage across the algorithms evaluated. In contrast, the proposed AADSGlobal significantly boosts performance, achieving speeds 3 to 13 times faster and consuming 3.5 to 6.5 times less memory than ADSGlobal. This improvement in efficiency is attributed to the definition of the approximate anchored densest subgraph, necessitating merely integer guesses for α, which aligns with our theoretical analysis in Theorem 3.6. In terms of the local algorithms, AADSLocal demonstrably surpasses ADSLocal, exhibiting speedups ranging from 5 to 1500 times, while only marginally increasing memory consumption. This substantial advancement is credited to two primary factors. Firstly, AADSLocal reduces the number of LocalReTest invocations, facilitated by adopting integer guesses for α like AADSGlobal. Secondly, it introduces a "build-as-you-compute" search strategy within LocalReTest, representing a significant enhancement over the iterative approach utilized by ADSLocal. The results clearly demonstrate the superiority of our proposed algorithms over the ADSGlobal and ADSLocal.

Additionally, Figure 4 shows that the running time of AADSGlobal is at least three orders of magnitude slower than that of AADSLocal in finding the approximate anchored densest subgraph. This lag stems from AADSGlobal's necessity to construct a complete re-orientation network for each iteration of the probing with α, a requirement not shared by AADSLocal. For a more nuanced comparison, Figure 6 depicts the runtime acceleration of AADSLocal relative to AADSGlobal. The x-axis denotes the average proportion of vertices explored in 100 queries, while the y-axis indicates the average runtime acceleration for these queries. In general, the performance gain of AADSLocal is inversely related to the number of vertices it explores. Again, the speedup across various graph sizes invariably exceeds a factor of 1000. Regarding memory usage, both AADSGlobal and AADSLocal necessitate identical memory allocations, owing to their requirement for linear-sized data structures dedicated to bounties, shortest path distances, and so on. Moreover, the memory consumption of both algorithms scales linearly with the size of the

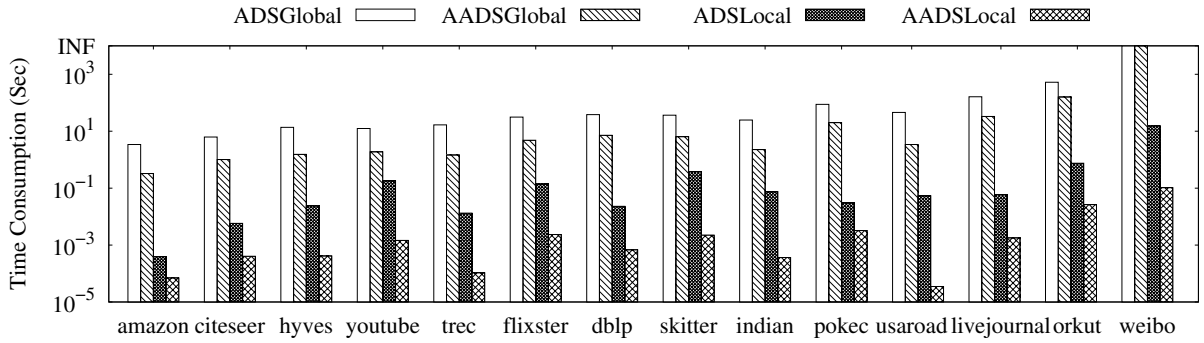


Figure 4: The comparison of the computation time

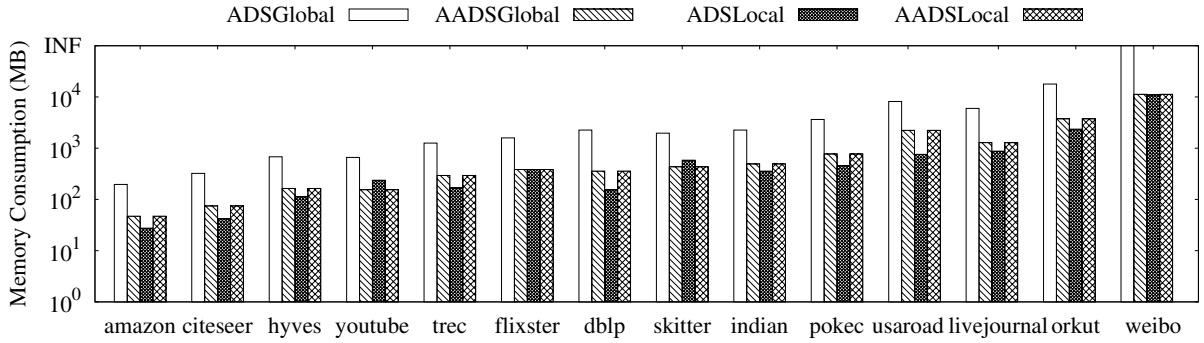


Figure 5: The comparison of the memory cost

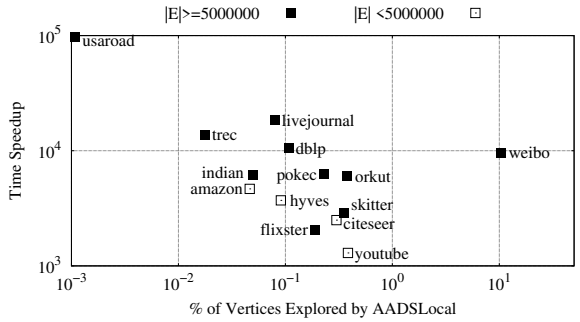


Figure 6: Performance gain of AADSLocal **over** AADSGlobal

dataset. These results confirm the superior efficiency of AADSLocal in comparison to AADSGlobal.

Exp-2: Sensitivity of AADSLocal **on graph density.** We commence with an original graph, denoted by G_0, and generate five subgraphs, $G_1, \ldots, G_5$, to represent different levels of density. We sample each edge in G_0 with a probability of 0.5^i, and then extract the largest connected component from the edge-induced subgraph as G_i. A suite of 1,000 queries, configured with default parameters, is generated for each G_i. Should a subgraph G_i not surpass the threshold of 128 vertices, it is excluded from our analysis. The average runtime and memory occupancy of AADSLocal on the eligible subgraphs are shown in Figure 7 and Figure 8, respectively. It is evident that both time costs and memory consumption of AADSLocal increase with the density of the graph. An average escalation in time cost by a factor of 3.08 and in memory requirements by a factor of 3.89 is recorded when the graph density is doubled. These findings highlight the AADSLocal algorithm's notable scalability, showcasing its efficiency across various graph densities.

Exp-3: Sensitivity of AADSLocal **on reference set size.** Let k be the cardinality of the reference vertex set R, i.e., $k = |R|$. Here we evaluate the running time of AADSLocal by varying k within the set $\{8, 16, 32, 64, 128\}$. We generate the set R with the specified size k using the default method, except that the size of A is resized to $\frac{k}{4}$ when generating A. For each specified k, a total of 100 queries are generated, upon which AADSLocal is executed for each dataset. The running time of AADSLocal with different k is depicted in Figure 9. It is observed that, across the majority of datasets, the runtime of AADSLocal incrementally rises in conjunction with the expansion of R. Remarkably, AADSLocal can efficiently search the approximate anchored densest subgraph in under one second within all parameter settings. When the size of $|R|$ is doubled, the average computational overhead incurred by AADSLocal increases by a factor of only 0.585. In addition, we evaluate the efficiency of AADSLocal and ADSLocal by varying larger k from the set $\{128, 256, ..., 65536, 131072\}$. The results are depicted in Figure 10. Our AADSLocal algorithm efficiently identifies the approximate anchored densest subgraph in under 10 seconds for all values of k. While the ADSLocal algorithm fails to deliver results within the time limitation for larger k. Again, AADSLocal achieves a runtime that is an order of magnitude faster than ADSLocal. These results demonstrate a significant insensitivity of the AADSLocal algorithm to variations in the size of the reference vertex set, thereby evidencing its robust scalability.

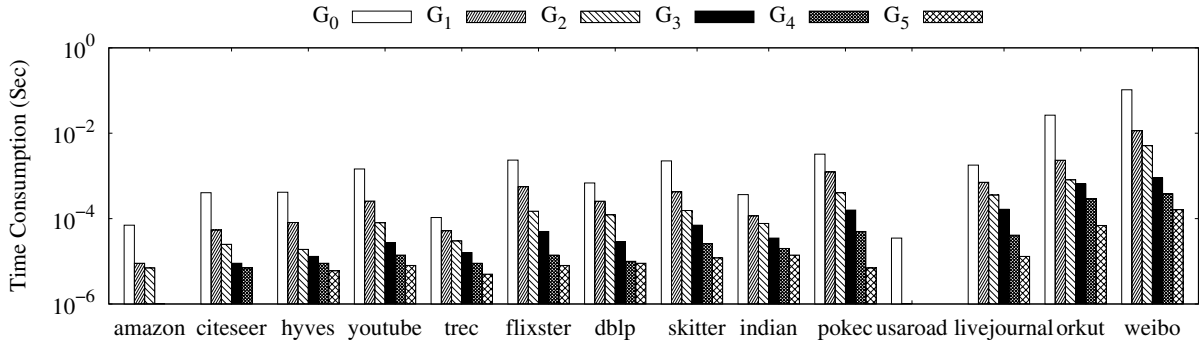


Figure 7: The running time of AADSLocal on graph density

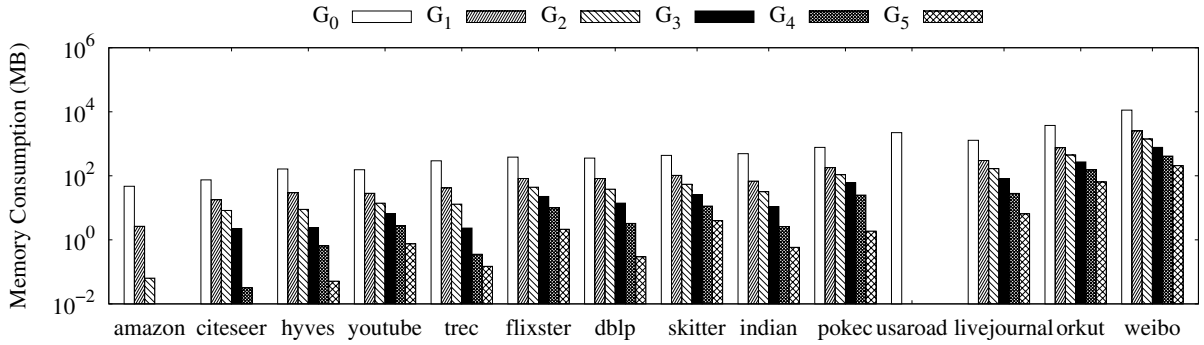


Figure 8: The memory cost of AADSLocal on graph density

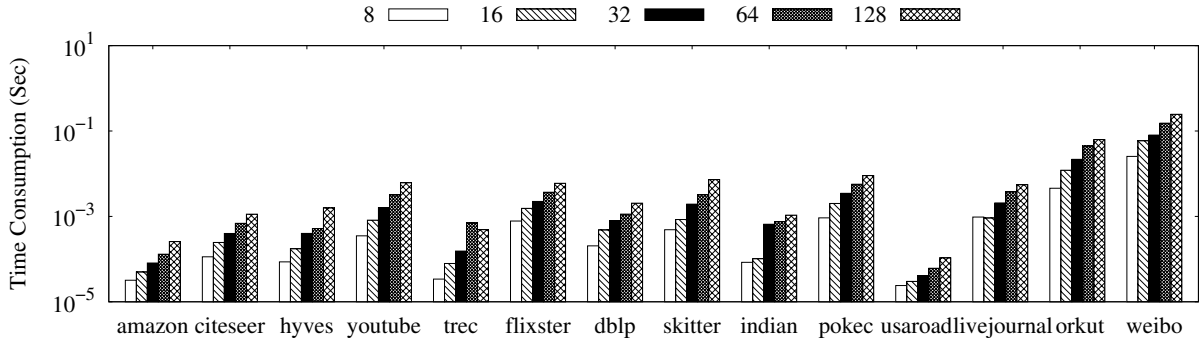


Figure 9: Sensitivity of AADSLocal on reference set size

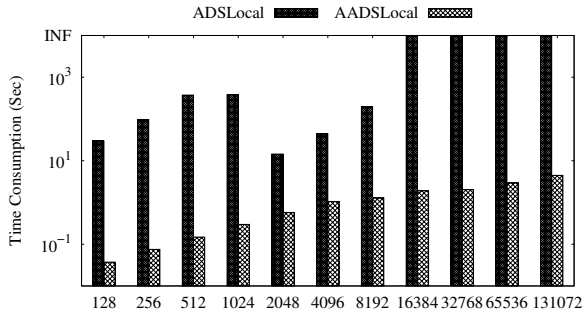


Figure 10: Computation time on larger reference vertex sets

Exp-4: The running time of maintenance algorithms. We generate 10,000 edges at random in each dataset to evaluate our algorithms: BasicIns and ImproveIns for edge insertion, along with BasicDel and ImproveDel for edge deletion. Given a specific dataset, for a query q_i within Exp-1, where $1 \leq i \leq 100$, the processing time for 10,000 updated edges is denoted as $T(q_i)$. Therefore, the average time for handling 100 queries, each subjected to 10,000 updates, denoted as $\overline{T} = \frac{1}{100} \sum_{i=1}^{100} T(q_i)$, is presented in Figure 11.

Building upon the insights from Exp-1, it becomes apparent that the recomputation time for 10,000 updated edges, as necessitated by algorithms aimed at (approximate) anchored densest subgraph search, significantly exceeds the processing time required by our maintenance algorithms. For instance, on the amazon dataset, the recomputation times mandated by the optimal algorithms, i.e.,

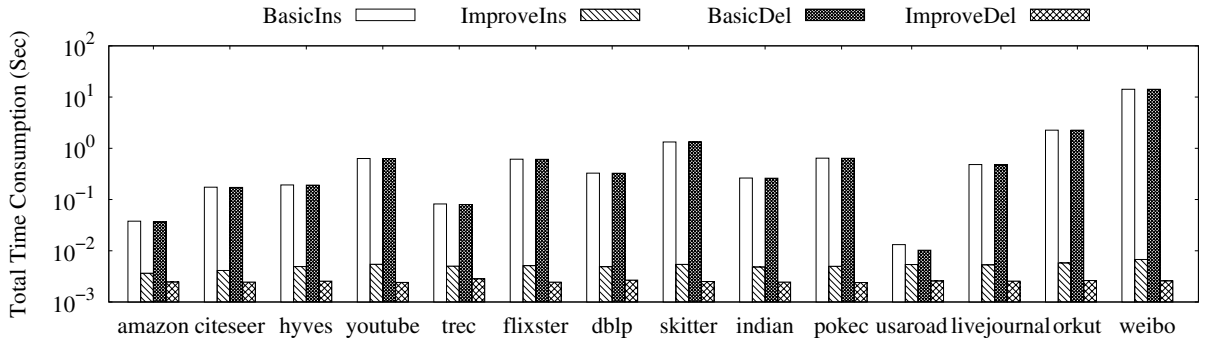


Figure 11: The running time of maintenance algorithms

Table 3: Comparison between ADS and AADS with different metrics

Dataset	Subgraph	R-subgraph density: $\rho_R(S) = \frac{2\|E(S)\| - \sum_{v \in S \setminus R} d_G(v)}{\|V(S)\|}$	Density: $\rho(S) = \frac{\|E(S)\|}{\|V(S)\|}$	Local Conductance: $\pi_R(S) = \frac{\|E(S,\overline{S})\|}{\mathrm{Vol}(R \cap S) - (S \cap \overline{R})}$	Conductance: $\pi(S) = \frac{\|E(S,\overline{S})\|}{\min(\|S\|, n - \|S\|)}$	F_1 score: $F_1(S,R) = \frac{2\|S \cap R\|}{(\|S\| + \|R\|)}$
amazon	ADS	5.6585040	2.835711	0.333572	0.332366	0.498863
	AADS	5.6383130 **(-0.020191)**	2.831374 **(-0.004337)**	0.340935 **(+0.007363)**	0.338488 **(+0.006122)**	0.533579 **(+0.034716)**
citeseer	ADS	7.9976380	4.002918	0.730507	0.730249	0.362857
	AADS	7.9811970 **(-0.016441)**	3.999623 **(-0.003295)**	0.735749 **(+0.005242)**	0.735093 **(+0.004844)**	0.384222 **(+0.021365)**
hyves	ADS	3.2104650	1.626541	0.921962	0.826158	0.321035
	AADS	3.1454330 **(-0.065032)**	1.599281 **(-0.027260)**	0.853966 **(-0.067996)**	0.842798 **(+0.016640)**	0.406393 **(+0.085358)**
youtube	ADS	7.2922350	3.650802	0.918281	0.913857	0.331365
	AADS	7.2484260 **(-0.043809)**	3.625227 **(-0.025575)**	0.929623 **(+0.011342)**	0.929438 **(+0.015581)**	0.37819 **(+0.046825)**
trec	ADS	5.6752320	3.586277	0.469242	0.429304	0.497379
	AADS	5.6024060 **(-0.072826)**	3.490155 **(-0.096122)**	0.508752 **(+0.039510)**	0.448882 **(+0.019578)**	0.576901 **(+0.079522)**
flixster	ADS	6.1586250	3.099062	0.955307	0.947673	0.226052
	AADS	6.1081780 **(-0.050447)**	3.073086 **(-0.025976)**	0.959910 **(+0.004603)**	0.954658 **(+0.006985)**	0.280093 **(+0.054041)**
dblp	ADS	10.543997	5.309548	0.767453	0.765460	0.289283
	AADS	10.527255 **(-0.016742)**	5.299411 **(-0.010137)**	0.770350 **(+0.002897)**	0.768543 **(+0.003083)**	0.303563 **(+0.01428)**
skitter	ADS	10.770263	5.411840	0.920304	0.844299	0.388504
	AADS	10.769975 **(-0.000288)**	5.413363 **(+0.001523)**	0.926973 **(+0.006669)**	0.854317 **(+0.010018)**	0.407545 **(+0.019041)**
indian	ADS	14.143748	7.646244	0.537802	0.513856	0.484087
	AADS	14.110848 **(-0.0329)**	7.653666 **(+0.007422)**	0.543738 **(+0.005936)**	0.520130 **(+0.006274)**	0.523696 **(+0.039609)**
pokec	ADS	6.5524730	3.276936	0.924950	0.924915	0.212633
	AADS	6.5162130 **(-0.03626)**	3.258538 **(-0.018398)**	0.925808 **(+0.000858)**	0.925789 **(+0.000874)**	0.241173 **(+0.02854)**
usaroad	ADS	2.2860640	1.143608	0.161678	0.161505	0.81171
	AADS	2.2797380 **(-0.006326)**	1.141942 **(-0.001666)**	0.160510 **(-0.001168)**	0.160070 **(-0.001435)**	0.829291 **(+0.017581)**
livejournal	ADS	11.759231	5.899040	0.789256	0.788460	0.286649
	AADS	11.743309 **(-0.015922)**	5.897954 **(-0.001086)**	0.791998 **(+0.002742)**	0.790474 **(+0.002014)**	0.299826 **(+0.013177)**
orkut	ADS	9.6347500	4.817375	0.937277	0.937277	0.209141
	AADS	9.6118910 **(-0.022859)**	4.805945 **(-0.01143)**	0.937234 **(-0.000043)**	0.937234 **(-0.000043)**	0.227251 **(+0.01811)**
weibo	ADS	2.6075910	1.456520	0.991176	0.970494	0.196846
	AADS	2.4221570 **(-0.185434)**	1.296905 **(-0.159615)**	0.994090 **(+0.002914)**	0.993316 **(+0.022822)**	0.447595 **(+0.250749)**

ADSLocal and AADSLocal, for insertion or deletion of 10,000 edges, totals to $0.00039 \times 10000 = 3.9$ seconds and $0.00007 \times 10,000 = 0.7$ seconds, respectively. Conversely, our BasicIns and ImproveIns algorithms require 0.037792 and 0.003647 seconds, respectively, to maintain the AADS for edge insertion. For edge deletion, BasicDel and ImproveDel necessitate 0.036612 and 0.002453 seconds, respectively. Moreover, ImproveIns (ImproveDel) consistently outperforms BasicIns (BasicDel) by at least an order of magnitude across all datasets, with the notable exception of usaroad, where it remains approximately 2.45 (4) times faster. The comparable processing times for BasicIns and BasicDel result from both algorithms requiring a mere two max-flow computations. Additionally, the running times of ImproveIns and ImproveDel are comparable in order of magnitude, with the former being slightly longer. These results confirm the efficiency of our algorithms for AADS maintenance.

Exp-5: Comparison of AADS and ADS. We conduct an evaluation of ADS and AADS using five distinct metrics to demonstrate the effectiveness of AADS as an approximation of ADS. This assessment generates 1,000 queries in accordance with the procedures detailed in Section 6.1, albeit with modifications to the number of random walks and steps, set to 15 and 4, respectively. Using these queries, the AADSLocal and ADSLocal algorithms are invoked to search AADS and ADS. For each metric, the average values of AADS and ADS derived from 1,000 queries are presented in Table 3. Investigations into the R-subgraph density across diverse datasets demonstrate that the divergence between $\rho_R(\hat{S})$ and $\rho_R(S^*)$ invariably remains below 0.2, with $\rho_R(\hat{S})$ consistently larger than $\rho_R(S^*)$. This magnitude of discrepancy is notably smaller than the theoretical maximum difference of 1 suggested by Fact 2.5, affirming the effectiveness of the AADS in approximating the ADS. Concerning metrics of density, local conductance, and conductance, the mean values of AADS over 1,000 queries are sometimes slightly greater than those of ADS and sometimes slightly less. However, overall, the mean values of AADS and ADS are comparable. Regarding the F_1 score, the average value of AADS slightly exceeds that of ADS across all datasets. The reason for this observation is as follows. When $\lceil \rho_R(\hat{S}) \rceil \geq 2$, AADS contains ADS exactly, and their R-subgraph densities are very close to each other. This indicates that vertices belonging to AADS but not ADS have a higher probability of being included in R, leading to a slightly higher F_1-score for AADS than for ADS. For the case of $\lceil \rho_R(\hat{S}) \rceil \leq 1$, our AADSLocal algorithm outputs R directly as AADS, which also results in a slightly

Table 4: Update time on temporal graphs (Second)

Dataset	n	m	ADSLocal	BasicIns	ImproveIns	BasicDel	ImproveDel
WikiElec	7,116	100,693	1, 200	3.080363	**0.014525**	3.084948	**0.040677**
Epinions	131,580	711,210	14, 150	8.044448	**0.022647**	8.118113	**0.093406**
Hepph	28,094,	3,148,447	132, 950	25.449169	**0.028956**	25.639931	**0.112757**

higher F_1-score for AADS than for ADS. These results show that AADS is a good approximation to ADS in terms of subgraph density and locality.

Exp-6: Results on temporal graphs. Here, we evaluate the proposed maintenance algorithms on three temporal graphs, namely, WikiElec, Epinions, and Hepph, which can be downloaded from the Network Repository. To simulate the evolution process of edge deletion, we design the following experiment for a dataset: first, we sort all edges according to their timestamps in ascending order and use the complete graph as the initial graph; then, 50,000 edges are deleted in descending order of their timestamps. For edge insertion, we use the opposite process: the graph consisting of $m - 50,000$ edges is used as the initial graph, and then 50,000 edges are inserted in ascending order of their timestamps.

As a comparison, since there is no algorithm to maintain ADS directly (except computation from scratch), we use the new graph formed after each edge update as input to ADSLocal to maintain ADS, which is very costly. Therefore, we roughly estimate the total time required to be $T^* \times 50{,}000$, where T^* is the average running time of 100 queries per dataset when applying ADSLocal. The total running time for all algorithms to update 50,000 edges in WikiElec, Epinions, and Hepph is shown in Table 4. Clearly, our proposed maintenance algorithms significantly outperform the re-computation method equipped with ADSLocal. The improved algorithms (i.e., ImproveIns and ImproveDel) are at least one order of magnitude faster than the basic algorithms (i.e., BasicIns and BasicDel), again indicating their high efficiency. In addition, we sample the R-subgraph density values of AADS S^* and ADS $\hat{S}$ during the edge updating process, as shown in Figure 12. It can be seen that the R-subgraph densities of AADS and ADS are consistently very close to each other over time for all three datasets, with $\rho_R(S^*)$ being slightly lower than $\rho_R(\hat{S})$. These results again show that AADS is a good approximation of ADS, consistent with the observations obtained from Exp-5.

6.3 Case studies

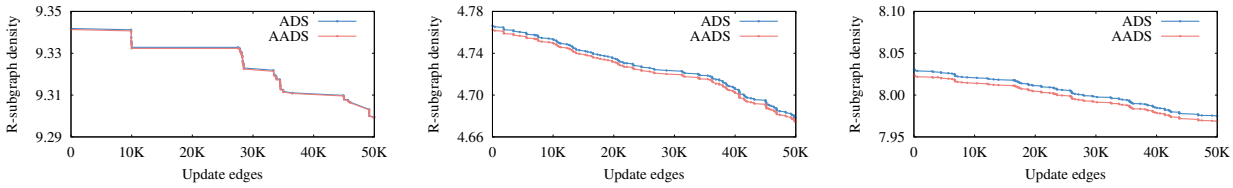
Case study on dblp**.** We conduct a case study on a subgraph, dblpCCF, of the dblp dataset, encompassing authors who have published in conferences and journals recommended by the China Computer Federation, along with their collaborative relationships. The dblpCCF subgraph is composed of 1,207,754 vertices and 5,878,173 edges. To construct the anchored set A and the reference set R, we apply a random walk technique with the parameters set to 15 walks of 4 steps each, following the procedure outlined in Section 6.1. Figure 13 shows the ADS and AADS by performing ADSLocal and AADSLocal, with Lixin Zhou identified as the seed vertex (represented by the red vertex). The yellow vertices, along with the seed vertex, constitute the anchored set A, and all depicted vertices are part of the reference set R. Note that the default size of set A is 8. However, due to the presence of an isolated vertex within A, which is not shown in Figure 13, both ADS and AADS implicitly include this isolated vertex, in alignment with their definitions. From Figure 13, it is evident that ADS, consisting of 43 vertices, is included in AADS, which includes 61 vertices. All vertices within AADS are elements of the reference set R, and AADS covers a wider subset of vertices in R than ADS. The metrics for the R-subgraph density of AADS and ADS are 8.787 and 8.884, respectively, with a marginal difference not surpassing 0.1, which shows that AADS is a good approximation of ADS. Furthermore, Figure 13 distinctly highlights the significant correlation of the additional green vertex on the right side of AADS with the seed vertex. Similarly, the green vertex positioned in the upper left corner shows a close association with ADS. The results show that AADS is not only a good approximation of ADS, but also better aligns with a given reference set R compared to ADS.

Case study on amazon**.** We also conduct a case study on the amazon dataset with the anchored set A and the reference set R generated by the method in the case study on dblp. Figure 14 showcases the ADS and AADS results obtained through the invocation of the ADSLocal and AADSLocal algorithms, with "280680: Genetics, Syndromes, and Communication Disorders" as the seed vertex (the red vertex). The anchored set A consists of the red seed vertex and the yellow vertices, and all vertices in Figure 14 form part of the reference set R. It can be observed that the additional green vertices located in the top right corner of the AADS exhibit a significant structural correlation with the anchor set A, and maintain a tight connection to the reference set R. Upon analyzing the metadata of the amazon dataset, it is evident that all vertices included in AADS are classified as books, encompassing a diverse range of categories including Health, Mind & Body, Medicine, Reference, and Professional & Technical. This illustrates the strong association of AADS's additional vertices with the sets A and R, confirming their practical relevance. In addition, ADS comprises 39 vertices, while AADS includes 48 vertices, with their respective R-subgraph densities being 5.590744 and 5.583333. Clearly, the negligible difference of merely 0.006411. These results again demonstrate the capability of AADS to approximate ADS closely.

7 RELATED WORK

Densest subgraph search. Our work is intricately linked to the Densest Subgraph Search (DSS) problem, which seeks to identify a subgraph exhibiting the highest edge density defined as the ratio of the number of edges to the number of vertices [16, 21, 28]. This problem has been approached through parametric maximum flow methods, achieving a complexity of $O(mn \cdot \log n)$ [28]. However, due to the prohibitive complexity of exact solutions for large-scale graphs, there has been significant advancement in approximation algorithms [8, 16, 21, 51]. Furthermore, the DSS problem has been extensively generalized to various graph types, including weighted [20], directed [16, 33, 36–38], bipartite [4, 29, 41], uncertain [42, 60], and multilayer graphs [26, 27, 31].

Variants of the DSS problem can be broadly classified into two main categories. The first involves introducing new density measures and developing algorithms for the efficient identification of dense subgraphs [21, 41, 48, 49, 52, 54]. The second category focuses on incorporating additional constraints into edge-density-based DSS problem, including size constraints [5, 11, 14, 23, 33, 46, 57], seed set [19, 22, 50], connectivity constraints [13, 39] and so on.



(a) WikiElec (b) Epinions (c) Hepph

Figure 12: The R-subgraph densities of ADS and AADS with three temporal graphs

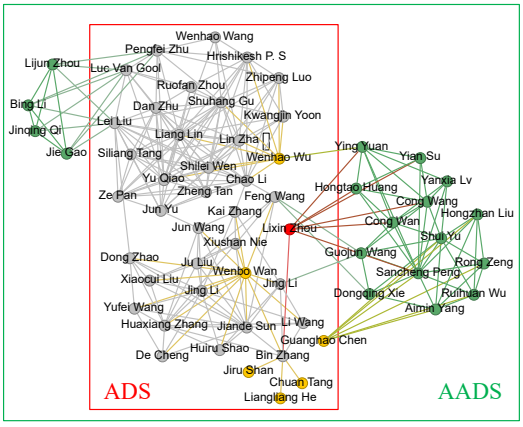


Figure 13: Case study on dblpCCF

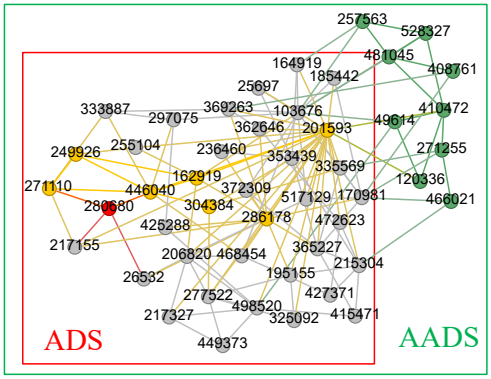


Figure 14: Case study on amazon

Among these variants, the seed set-based variation holds particular relevance to our study. Dai *et al.* established the concept of R-subgraph density for a vertex set S, applying penalties to vertices outside a given reference set R, leading to the formulation of the anchored densest subgraph search problem [19]. They proposed a local algorithm to efficiently identify the vertex set S with the maximum R-subgraph density, whose time complexity is independent of the input graph's size. Sozio and Gionis addressed the problem of seeking a set S that includes all query vertices $Q \subseteq V$, with S possessing the highest minimum degree while meeting criteria like the maximum permissible distance between S and Q [50]. They demonstrated the effectiveness of adopting the Charikar greedy peeling algorithm for optimal solutions. Fazzone *et al.* investigated the dense subgraph with attractors and repulsers problem, aiming to identify a dense subgraph S that is proximal to a set of attractors, A, while maintaining distance from a set of repulsers, R [22]. This paper studies the approximate anchored densest subgraph search and maintenance based on the R-subgraph density defined in [19]. Due to different problem definitions, the algorithms described in [50] and [22] are not applicable to our problem. The algorithms in [19], on the other hand, are significantly less efficient when dealing with large-scale graphs or sizable R sets and cannot maintain the ADS for dynamic graphs. In this paper, we present, for the first time, efficient algorithms for searching the approximate anchored densest subgraph on both static and dynamic graphs.

The re-orientation network flow techniques. Our work is also related to the re-orientation network flow techniques, initially developed to address the minimization of the maximum in-degree problem [3, 6, 17]. The re-orientation network flow can also be used to compute pseudoarboricity since the maximum in-degree of the optimal orientation is equal to the pseudoarboricity [10]. Bezáková proposed a renowned method to compute the exact optimal orientation by using the re-orientation network flow technique and binary search, achieving a time complexity of $O(|E|^{3/2}\log p)$ [10]. Blumenstock later enhanced this method, reducing the complexity to $O(|E|^{3/2}\sqrt{\log\log p})$ [12]. In terms of approximation algorithms, Bezáková developed a 2-approximation algorithm with a runtime of $O(m+n)$ [10], while Kowalik proposed a $(1+\epsilon)$-approximation algorithm based on the early-stopped Dinic algorithm, which operates with time complexity of $O(m \log n \max\{1, \log p\}/\epsilon)$ [34]. Additionally, Asahiro investigated the orientation of edges in a weighted graph to minimize the maximum weighted out-degree [7]. To the best of our knowledge, our research constitutes the first application of the re-orientation network flow technique to the search and maintenance of the approximate anchored densest subgraph.

8 CONCLUSION

In this paper, we investigate the problem of approximate anchored densest subgraph search in static and dynamic graphs, where the R-subgraph densities of the approximate solution and exact anchored densest subgraph are equal after rounding upwards. We propose the AADSGlobal algorithm equipped with the re-orientation network flow technique and a binary search method. To improve the efficiency, an innovative local algorithm is proposed that utilizes shortest-path based methods to compute the max-flow from s to t around R locally. Furthermore, for dynamic graphs, we develop both basic and improved algorithms aimed at efficiently maintaining the AADS for edge insertions and deletions. Comprehensive experiments and two detailed case studies demonstrate the efficiency, scalability, and effectiveness of our solutions.

REFERENCES

[1] Fabeah Adu-Oppong, Casey K Gardiner, Apu Kapadia, and Patrick P Tsang. 2008. Social circles: Tackling privacy in social networks. In *SOUPS*.

[2] Sanjay Agrawal, Surajit Chaudhuri, and Gautam Das. 2002. DBXplorer: A System for Keyword-Based Search over Relational Databases. In *ICDE*. 5–16.

[3] Oswin Aichholzer, Franz Aurenhammer, and Günter Rote. 1995. *Optimal graph orientation with storage applications*. Universität Graz/Technische Universität Graz. SFB F003-Optimierung und Kontrolle.

[4] Reid Andersen. 2008. A local algorithm for finding dense subgraphs. In *SODA*. 1003–1009.

[5] Reid Andersen and Kumar Chellapilla. 2009. Finding Dense Subgraphs with Size Bounds. In *WAW (Lecture Notes in Computer Science, Vol. 5427)*. 25–37.

[6] Srinivasa Rao Arikati, Anil Maheshwari, and Christos D. Zaroliagis. 1997. Efficient Computation of Implicit Representations of Sparse Graphs. *Discret. Appl. Math.* 78, 1-3 (1997), 1–16.

[7] Yuichi Asahiro, Eiji Miyano, Hirotaka Ono, and Kouhei Zenmyo. 2007. Graph Orientation Algorithms to minimize the Maximum Outdegree. *Int. J. Found. Comput. Sci.* 18, 2 (2007), 197–215.

[8] Bahman Bahmani, Ravi Kumar, and Sergei Vassilvitskii. 2012. Densest Subgraph in Streaming and MapReduce. *Proc. VLDB Endow.* 5, 5 (2012), 454–465.

[9] Oana Denisa Balalau, Francesco Bonchi, T.-H. Hubert Chan, Francesco Gullo, and Mauro Sozio. 2015. Finding Subgraphs with Maximum Total Density and Limited Overlap. In *WSDM*. 379–388.

[10] Ivona Bezáková. 2000. Compact representations of graphs and adjacency testing. (2000).

[11] Aditya Bhaskara, Moses Charikar, Eden Chlamtac, Uriel Feige, and Aravindan Vijayaraghavan. 2010. Detecting high log-densities: an O (n 1/4) approximation for densest k-subgraph. In *STOC*. 201–210.

[12] Markus Blumenstock. 2016. Fast Algorithms for Pseudoarboricity. In *ALENEX*. SIAM, 113–126.

[13] Francesco Bonchi, David García-Soriano, Atsushi Miyauchi, and Charalampos E. Tsourakakis. 2021. Finding densest k-connected subgraphs. *Discret. Appl. Math.* 305 (2021), 34–47.

[14] Nicolas Bourgeois, Aristotelis Giannakos, Giorgio Lucarelli, Ioannis Milis, and Vangelis Th. Paschos. 2013. Exact and Approximation Algorithms for Densest k-Subgraph. In *WALCOM (Lecture Notes in Computer Science, Vol. 7748)*. 114–125.

[15] Lijun Chang and Miao Qiao. 2020. Deconstruct Densest Subgraphs. In *WWW*. 2747–2753.

[16] Moses Charikar. 2000. Greedy approximation algorithms for finding dense components in a graph. In *APPROX 2000 (Lecture Notes in Computer Science, Vol. 1913)*. Springer, 84–95.

[17] Marek Chrobak and David Eppstein. 1991. Planar Orientations with Low Outdegree and Compaction of Adjacency Matrices. *Theor. Comput. Sci.* 86, 2 (1991), 243–266.

[18] Edith Cohen, Eran Halperin, Haim Kaplan, and Uri Zwick. 2003. Reachability and Distance Queries via 2-Hop Labels. *SIAM J. Comput.* 32, 5 (2003), 1338–1355.

[19] Yizhou Dai, Miao Qiao, and Lijun Chang. 2022. Anchored Densest Subgraph. In *SIGMOD*. 1200–1213.

[20] Maximilien Danisch, T.-H. Hubert Chan, and Mauro Sozio. 2017. Large Scale Density-friendly Graph Decomposition via Convex Programming. In *WWW*. 233–242.

[21] Yixiang Fang, Kaiqiang Yu, Reynold Cheng, Laks V. S. Lakshmanan, and Xuemin Lin. 2019. Efficient Algorithms for Densest Subgraph Discovery. *Proc. VLDB Endow.* 12, 11 (2019), 1719–1732.

[22] Adriano Fazzone, Tommaso Lanciano, Riccardo Denni, Charalampos E. Tsourakakis, and Francesco Bonchi. 2022. Discovering Polarization Niches via Dense Subgraphs with Attractors and Repulsers. *Proc. VLDB Endow.* 15, 13 (2022), 3883–3896.

[23] Uriel Feige, Guy Kortsarz, and David Peleg. 2001. The Dense k-Subgraph Problem. *Algorithmica* 29, 3 (2001), 410–421.

[24] Eugene Fratkin, Brian T. Naughton, Douglas L. Brutlag, and Serafim Batzoglou. 2006. MotifCut: regulatory motifs finding with maximum density subgraphs. In *ISMB*. 156–157.

[25] Esther Galbrun, Aristides Gionis, and Nikolaj Tatti. 2016. Top-k overlapping densest subgraphs. *Data Min. Knowl. Discov.* 30, 5 (2016), 1134–1165.

[26] Edoardo Galimberti, Francesco Bonchi, and Francesco Gullo. 2017. Core Decomposition and Densest Subgraph in Multilayer Networks. In *CIKM*. 1807–1816.

[27] Edoardo Galimberti, Francesco Bonchi, Francesco Gullo, and Tommaso Lanciano. 2020. Core Decomposition in Multilayer Networks: Theory, Algorithms, and Applications. *ACM Trans. Knowl. Discov. Data* 14, 1 (2020), 11:1–11:40.

[28] Andrew V Goldberg. 1984. Finding a maximum density subgraph. (1984).

[29] Bryan Hooi, Hyun Ah Song, Alex Beutel, Neil Shah, Kijung Shin, and Christos Faloutsos. 2016. FRAUDAR: Bounding Graph Fraud in the Face of Camouflage. In *KDD*. 895–904.

[30] Xin Huang, Hong Cheng, Lu Qin, Wentao Tian, and Jeffrey Xu Yu. 2014. Querying k-truss community in large and dynamic graphs. In *SIGMOD*. 1311–1322.

[31] Vinay Jethava and Niko Beerenwinkel. 2015. Finding Dense Subgraphs in Relational Graphs. In *ECML PKDD (Lecture Notes in Computer Science, Vol. 9285)*. Springer, 641–654.

[32] Ruoming Jin, Yang Xiang, Ning Ruan, and David Fuhry. 2009. 3-HOP: a high-compression indexing scheme for reachability query. In *SIGMOD*. 813–826.

[33] Samir Khuller and Barna Saha. 2009. On Finding Dense Subgraphs. In *ICALP (Lecture Notes in Computer Science, Vol. 5555)*. 597–608.

[34] Lukasz Kowalik. 2006. Approximation Scheme for Lowest Outdegree Orientation and Graph Density Measures. In *ISAAC (Lecture Notes in Computer Science, Vol. 4288)*. Springer, 557–566.

[35] Chengwei Lei and Jianhua Ruan. 2013. A novel link prediction algorithm for reconstructing protein-protein interaction networks by topological similarity. *Bioinform.* 29, 3 (2013), 355–364.

[36] Wensheng Luo, Zhuo Tang, Yixiang Fang, Chenhao Ma, and Xu Zhou. 2023. Scalable Algorithms for Densest Subgraph Discovery. In *ICDE*. 287–300.

[37] Chenhao Ma, Yixiang Fang, Reynold Cheng, Laks V. S. Lakshmanan, Wenjie Zhang, and Xuemin Lin. 2020. Efficient Algorithms for Densest Subgraph Discovery on Large Directed Graphs. In *SIGMOD*. 1051–1066.

[38] Chenhao Ma, Yixiang Fang, Reynold Cheng, Laks V. S. Lakshmanan, Wenjie Zhang, and Xuemin Lin. 2021. On Directed Densest Subgraph Discovery. *ACM Trans. Database Syst.* 46, 4 (2021), 13:1–13:45.

[39] Wolfgang Mader. 1972. Existenz n-fach zusammenhängender Teilgraphen in Graphen genügend grosser Kantendichte. In *Abhandlungen aus dem mathematischen Seminar der Universität Hamburg*, Vol. 37. 86–97.

[40] Alberto O. Mendelzon. 2000. Review - Authoritative Sources in a Hyperlinked Environment. *ACM SIGMOD Digit. Rev.* 2 (2000).

[41] Michael Mitzenmacher, Jakub Pachocki, Richard Peng, Charalampos E. Tsourakakis, and Shen Chen Xu. 2015. Scalable Large Near-Clique Detection in Large-Scale Networks via Sampling. In *KDD*. 815–824.

[42] Atsushi Miyauchi and Akiko Takeda. 2018. Robust Densest Subgraph Discovery. In *ICDM*. 1188–1193.

[43] Gergely Palla, Imre Derényi, Illés Farkas, and Tamás Vicsek. 2005. Uncovering the overlapping community structure of complex networks in nature and society. *Nature* 435, 7043 (2005), 814–818.

[44] B. Aditya Prakash, Ashwin Sridharan, Mukund Seshadri, Sridhar Machiraju, and Christos Faloutsos. 2010. EigenSpokes: Surprising Patterns and Scalable Community Chipping in Large Graphs. In *PAKDD (Lecture Notes in Computer Science, Vol. 6119)*. Springer, 435–448.

[45] Lu Qin, Rong-Hua Li, Lijun Chang, and Chengqi Zhang. 2015. Locally Densest Subgraph Discovery. In *KDD*. 965–974.

[46] Frederic Roupin and Alain Billionnet. 2008. A deterministic approximation algorithm for the densest k-subgraph problem. *International Journal of Operational Research* 3, 3 (2008), 301–314.

[47] Barna Saha, Allison Hoch, Samir Khuller, Louiqa Raschid, and Xiao-Ning Zhang. 2010. Dense Subgraphs with Restrictions and Applications to Gene Annotation Graphs. In *RECOMB (Lecture Notes in Computer Science, Vol. 6044)*. 456–472.

[48] Raman Samusevich, Maximilien Danisch, and Mauro Sozio. 2016. Local triangle-densest subgraphs. In *ASONAM*. 33–40.

[49] Saurabh Sawlani and Junxing Wang. 2020. Near-optimal fully dynamic densest subgraph. In *SIGACT*. 181–193.

[50] Mauro Sozio and Aristides Gionis. 2010. The community-search problem and how to plan a successful cocktail party. In *KDD*. ACM, 939–948.

[51] Hsin-Hao Su and Hoa T. Vu. 2020. Distributed Dense Subgraph Detection and Low Outdegree Orientation. In *DISC (LIPIcs, Vol. 179)*. 15:1–15:18.

[52] Bintao Sun, Maximilien Danisch, T.-H. Hubert Chan, and Mauro Sozio. 2020. KClist++: A Simple Algorithm for Finding k-Clique Densest Subgraphs in Large Graphs. *Proc. VLDB Endow.* 13, 10 (2020), 1628–1640.

[53] Nikolaj Tatti and Aristides Gionis. 2013. Discovering Nested Communities. In *ECML PKDD (Lecture Notes in Computer Science, Vol. 8189)*. 32–47.

[54] Charalampos E. Tsourakakis. 2015. The K-clique Densest Subgraph Problem. In *WWW*. 1122–1132.

[55] Elena Valari, Maria Kontaki, and Apostolos N. Papadopoulos. 2012. Discovery of Top-k Dense Subgraphs in Dynamic Graph Collections. In *SSDBM (Lecture Notes in Computer Science, Vol. 7338)*. 213–230.

[56] A YEFIM. 1970. DINITZ: Algorithm for solution of a problem of maximum flow in a network with power estimation. In *Soviet Math. Doklady*, Vol. 11. 1277–1280.

[57] Peng Zhang and Zhendong Liu. 2021. Approximating Max k-Uncut via LP-rounding plus greed, with applications to Densest k-Subgraph. *Theor. Comput. Sci.* 849 (2021), 173–183.

[58] Yang Zhang and Srinivasan Parthasarathy. 2012. Extracting Analyzing and Visualizing Triangle K-Core Motifs within Networks. In *ICDE*. 1049–1060.

[59] Feng Zhao and Anthony K. H. Tung. 2012. Large Scale Cohesive Subgraphs Discovery for Social Network Visual Analysis. *Proc. VLDB Endow.* 6, 2 (2012), 85–96.

[60] Zhaonian Zou. 2013. Polynomial-time algorithm for finding densest subgraphs in uncertain graphs. In *Proceedings of MLG Workshop*.