

Maximal D-truss Search in Dynamic Directed Graphs

Anxin Tian
HKUST
atian@ust.hk

Yue Wang
Shenzhen Institute of Computing Sciences
yuewang@sics.ac.cn

Alexander Zhou
HKUST
atzhou@cse.ust.hk

Lei Chen
HKUST
HKUST(GZ)
leichen@ust.hk

ABSTRACT

Community search (CS) aims at personalized subgraph discovery which is the key to understanding the organisation of many real-world networks. CS in undirected networks has attracted significant attention from researchers, including many solutions for various cohesive subgraph structures and for different levels of dynamism with edge insertions and deletions, while they are much less considered for directed graphs. In this paper, we propose incremental solutions of CS based on the D-truss in dynamic directed graphs, where the D-truss is a cohesive subgraph structure defined based on two types of triangles in directed graphs. We first analyze the theoretical boundedness of D-truss given edge insertions and deletions, then we present basic single-update algorithms. To improve the efficiency, we propose an order-based D-Index, associated batch-update algorithms and a fully-dynamic query algorithm. Our extensive experiments on real-world graphs show that our proposed solution achieves a significant speedup compared to the SOTA solution, the scalability over updates is also verified.

PVLDB Reference Format:

Anxin Tian, Alexander Zhou, Yue Wang, and Lei Chen. Maximal D-truss Search in Dynamic Directed Graphs. PVLDB, 16(9): 2199-2211, 2023. doi:10.14778/3598581.3598592

PVLDB Artifact Availability:

The source code, data, and/or other artifacts have been made available at <https://github.com/ExpCodeBase/DTrussMain>.

1 INTRODUCTION

The graph is a concise structure consisting of vertices and edges and is widely used in multiple fields. Community search (CS) is a popular topic of graph processing, which intends to find a specifically connected subgraph for the given query vertices [22]. The community is a concept generated from social networks, while CS works in not only social analysis [1, 5], but also recommendation [3, 12], industry intelligence [30] and bio-medicine [20, 31, 33].

Many works have proposed various solutions based on different metrics [4, 25, 29] or cohesive subgraph structures [7, 9, 41] for solving CS problems. Though CS is well-studied in undirected graphs,

it is less considered for directed graphs. For capturing asymmetric relationships between entities in directed graphs, the D-truss [28] is proposed as a cohesive subgraph structure, which is defined based on two types of triangles in directed graphs. Different from undirected graphs, triangles can be classified into flow triangles and cycle triangles according to directions of edges among the three vertices [37]. Figure 1 (a) shows an example of directed triangles, where v_2, v_3 and v_4 form a cycle triangle, v_2, v_5 and v_6 form a flow triangle. It is worth to mention that the same three vertices can be a part of multiple triangles, e.g. v_2, v_3 and v_4 also form a flow triangle since v_2 and v_4 are bi-directionally connected. We can find a hub vertex in a flow triangle where the other two vertices always point to it, while we cannot distinguish such a vertex out of three vertices which are connected cyclically. In a D-truss (or (k_c, k_f) -truss) [28], each edge is involved in cycle (flow resp.) triangles with at least k_c (k_f resp.) other vertices. The k -truss [9] is proposed for undirected graphs, where each edge is contained in at least k triangles, applying the k -truss in directed graphs by ignoring directions of edges cannot distinguish different triangle participations of each edge. The D-truss shows advantages in CS over directed graphs compared with the k -truss. Formally, given a set of query vertices, D-truss community search (DCS) [28] is defined as finding the connected subgraph that: (1) contains the query vertices; (2) fulfils the requirement of D-truss number; (3) minimizes the diameter. Specifically, the existing solution to DCS [28] obtains the maximal satisfied subgraph first, and then peels off vertices which are deemed as not tightly related to query vertices by minimizing the diameter.

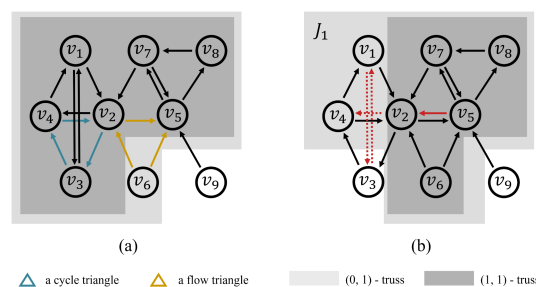


Figure 1: The D-truss example in a toy directed graph

However, many existing CS works based on cohesive subgraph structures [17, 21, 39] which only require the first two conditions and then return the maximal satisfied subgraph directly as the query answer, because many applications need the maximality of

This work is licensed under the Creative Commons BY-NC-ND 4.0 International License. Visit <https://creativecommons.org/licenses/by-nc-nd/4.0/> to view a copy of this license. For any use beyond those covered by this license, obtain permission by emailing info@vldb.org. Copyright is held by the owner/author(s). Publication rights licensed to the VLDB Endowment. Proceedings of the VLDB Endowment, Vol. 16, No. 9 ISSN 2150-8097. doi:10.14778/3598581.3598592

the community such as network analysis [18, 23], recommendation [40] and event organisations [16]. We take the biological food web analysis [23] as an example, the maximal subgraph is necessary for discovering all taxa given the cohesiveness requirement. For proposing a general CS scheme based on the D-truss, we study the problem called maximal D-truss search (MDS), which removes the diameter constraint of DCS.

Considering directed graphs in the real-world, they are always changing with edges or vertices being continuously inserted into or removed from the original graph. Let the directed graph be G , the updates of it be ΔG , and the result of the query given G be $R(G)$, an incremental algorithm \mathcal{A}_Δ aims at obtaining the updated result and reusing $R(G)$ as much as possible, well-designed incremental algorithms are always more efficient and more scalable than from-scratch solutions [14, 15].

Applications. Finding the D-truss has many real-world applications, such as the analysis of social networks and the discovery of web networks [28]. It is inherent that many real-world directed graphs are updated frequently, thus having the D-truss in the latest graph maintained correctly and efficiently is extremely desired in many scenarios. We list some examples below.

- *Social network.* Figure 1 (a) shows an example of the social network, the direction of an edge represents the followed-following relationship between two users, i.e. the edge $\langle v_1, v_2 \rangle$ means the user v_1 follows the user v_2 . Users always follow and unfollow others according to their interests, which will result in the update of the whole graph. In Figure 1 (a), we mark all D-trusses in different colors in the original graph G . In Figure 1 (b), let ΔG be $\{-\langle v_1, v_3 \rangle, -\langle v_3, v_1 \rangle, -\langle v_2, v_4 \rangle, +\langle v_5, v_2 \rangle\}$, where we denote the deletion by $-$ and the insertion by $+$. After applying ΔG , the result of MDS becomes different from the result in Figure 1 (a).

- *Associative thesaurus.* The word associative network works as a useful tool for computational linguistics and natural language processing. On the network of the Edinburgh Associative Thesaurus (EAT) [6]: the word w_i points to another word w_j if w_j appears as the response when w_i is received as the stimulus. Given a query word, the D-truss will include all associative words of it and distinguish them via cycle triangles and flow triangles. In the case study of EAT [28], the $(3, 0)$ -truss of the word “drink” consists of words that have an equal relationship with it, including “glass”, “bottle”, “wine” and “water”. Instead the $(0, 7)$ -truss of it shows the reminding hierarchy relationships around “drink”. There exist words are in the upstream of the hierarchy relationship, e.g. the word “drunk” always reminds “drink” and other words in the subgraph, the word “drink” also reminds many words that in the downstream of the relationship, e.g. “rum” and “liquor”. Once the thesaurus is updated with new edges added and out-of-time edges deleted, such relationships found by the D-truss may be changed, it is strongly motivated that maintaining these D-trusses instead of recomputing them.

Challenges. Though effective and efficient solutions for k -truss maintenance problem of both single-update [21] and batch-update [42] are developed, it is not trivial to extend those solutions to the D-truss. Instead, we must address the following challenges in order to solve the problem.

- *Challenge 1.* Compared with the k -truss model in undirected graphs, the D-truss takes the directions of edges within triangles

into consideration. The direction of an edge makes it mean differently from an undirected edge, this fact implies two key points to notice: (1) the triangles in directed graphs can be classified into two types (i.e. cycle triangles and flow triangles) due to different directions within a triangle; (2) the k -truss ignore duplicate edges, while two edges with different directions are allowed between two vertices u and v . The theoretical hardness of the D-truss maintenance is unknown because of these two points.

- *Challenge 2.* According to the two different types of triangles, the D-truss decomposition [28] is naturally related to two types of supports. For single edge updates, existing work [21] proposes theorems that indicate the rules of k -truss maintenance depending on the only type of support. The fact that the same three vertices can be contained in multiple triangles in the D-truss makes such theorems cannot be extended correctly.

- *Challenge 3.* For more efficient batch-update solutions, the existing index [42] of the k -truss stores edges in the non-ascending order of k values. However, there are a pair of values to consider (i.e. (k_c, k_f)) for the D-truss, the dominant property of (k_c, k_f) [28] makes it impossible to construct such an index for the D-truss.

In this paper, we aim at an efficient incremental solution for addressing the challenges above. We denote that applying ΔG to G by $G \oplus \Delta G$. Many existing solutions of CS have to recompute from scratch for $G \oplus \Delta G$, the cost of recomputing given ΔG always depends on the size of the whole graph, the incremental solution would be more efficient if its cost is polynomial with a specific range of data accessed, which is always smaller than the size of the graph. Furthermore, the solution is more practical if it can handle the updates of different query vertices and (k_c, k_f) requirements except for ΔG , which is called fully-dynamic.

Contributions. For proposing a general incremental scheme for solving CS based on the D-truss, we first analyze the hardness of D-truss maintenance theoretically and then we propose solutions for both single-update and batch-update. More concretely, our contributions can be concluded below:

- We define the maximal D-truss search problem for generality and extend the existing solution to it.
- We conduct theoretical analysis about the boundedness and the unboundedness of the maximal D-truss maintenance.
- We present support update operators and basic single-update algorithms for edge insertion and deletion respectively for the maximal D-truss maintenance problem.
- We propose an order-based index structure, its associated batch-update algorithms and the fully-dynamic query algorithm.

Outline. In Section 2, we review the related work on this topic. Section 3 gives definitions of notations, concepts and problems. We show theoretical discussions in Section 4. Our basic methods are presented in Section 5 and index-based methods are introduced in Section 6. Section 7 gives batch-update algorithms and the query algorithm for the proposed index. We show experimental results in Section 8 and conclude this paper in Section 9.

2 RELATED WORK

In this section, we review existing works of cohesive subgraph models, their maintenance techniques in directed graphs and effectiveness measures of incremental algorithms.

Cohesive subgraph structures over directed graphs. We conclude popular cohesive subgraph structures into three types below.

- *Core-like models:* k -core [7, 8] is a cohesive subgraph structure of undirected graphs where every vertex has at least k degrees. For directed graphs, the D-core (or (k, l) -core) [17, 19] is proposed to find the community with minimum k in-degree and l out-degree requirements being fulfilled. Specifically, the D-core is limited by its inherent property that the fixed requirement of combinations of in-degree and out-degree. The degree-based features of some vertices can be identified explicitly, e.g. vertices with high in-degree and low out-degree imply that entities have influences over others like authorities, and vertices with low in-degree and high out-degree imply that entities tend to be followers [28]. However, compared to the D-truss, for containing two aforementioned types in the D-core, it will be too sparse and unable to capture edge cohesiveness [9, 28].
- *Truss-like models:* In undirected graphs, k -truss [9] is defined based on the number of triangles that each edge is involved in. In this paper, we develop the work based on the D-truss (or (k_c, k_f) -truss) [28], which distinguishes two different types of triangles in semantics for directed graphs. The D-truss can alleviate the limitation of the D-core by varying different k_c and k_f [28]. With the computation of the skyline trussness set for all edges of the given graph, all possible queries of MDS can be efficiently answered from the constructed index [28]. A query algorithm [28] is given which can find the maximal D-truss that contains Q by sparing the procedures for minimizing the diameter.

Cohesive subgraph maintenance over dynamic graphs. In the real world, graphs are always changing over time, instead of re-computing subgraphs from scratch, many works focus on efficient incremental algorithms. For core maintenance, a suite of incremental k -core decomposition algorithms [35] are verified to be efficient, and a tree-like index [27] can be constructed for more efficient maintenance. A fully dynamic approximate solution for the k -core in hypergraphs [36] is also considered. In bipartite graphs, the problem of (α, β) -core decomposition and maintenance is solved by an index-based solution [39]. For truss maintenance, different index structures are proposed based on the preserved information of triangle connectivity [21] and the edge decomposition order [42]. For clique maintenance, maintaining the k -clique in dynamic graphs is tackled by single-update [13] and batch-update methods [10].

3 PRELIMINARIES

Given a directed, simple and unweighted graph denoted by $G = (V(G), E(G))$. Here $V(G)$ denotes the set of vertices and $E(G)$ denotes the set of edges that $E(G) = \{\langle u, v \rangle \mid u, v \in V(G)\}$. Note that double edges which point in opposite directions are allowed. The edge $\langle u, v \rangle$ is the directed edge that points from u to v , where u is an in-neighbor of v and v is an out-neighbor of u .

For a vertex $u \in V(G)$, we denote the set of in-neighbors of u by $N^{\text{in}}(u, G) = \{v \mid \langle v, u \rangle \in E(G)\}$ and the set of out-neighbors of v by $N^{\text{out}}(u, G) = \{v \mid \langle u, v \rangle \in E(G)\}$. $\text{deg}^{\text{in}}(u, G) = |N^{\text{in}}(u, G)|$ is used to denote the in-degree of u and similarly $\text{deg}^{\text{out}}(u, G) = |N^{\text{out}}(u, G)|$ denotes the out-degree of u . The degree of u is defined as the sum of its in-degree and out-degree. We use n to denote the number of vertices in G and m to denote the number of edges in G .

In this section, we give the definitions of the D-truss and existing solutions based on it.

3.1 Problem Definition

Before we give the formal problem statement, some basic concepts are defined as follows. We denote three vertices within a triangle by u, v, w respectively. A cycle triangle is a triangle where each vertex within has exactly 1 in-degree and 1 out-degree, we denote it by $\Delta_{u,v,w}^c$. A flow triangle is a triangle where the out-degrees of vertices within are 0, 1, 2 and in-degrees of them are 2, 1, 0 correspondingly, we denote a flow triangle by $\Delta_{u,v,w}^f$. Figure 1 shows an example for directed triangles where v_2, v_3 and v_4 form a cycle triangle, v_2, v_5 and v_6 form a flow triangle. The D-truss [28] is proposed together with two types of support below.

Table 1: Notations

Notation	Description
$G, \Delta G$	A directed graph and updates to it
G' or $G \oplus \Delta G$	The graph after applying ΔG to G
$V(G), E(G)$	The set of vertices and edges respectively
$\text{sup}_G^c(e), \text{sup}_G^f(e)$	The cycle support and flow support of $\forall e \in E(G)$
$N_G^c(e), N_G^f(e)$	The cycle set and flow set of $\forall e \in E(G)$
$E_G^c(e)$	The set of edges that are spanned between $N_G^c(e)$ and $\{u, v\}$, where $e = \langle u, v \rangle$.
$E_G^f(e)$	The set of edges that are spanned between $N_G^f(e)$ and $\{u, v\}$, where $e = \langle u, v \rangle$.
H_{k_c, k_f}	The D-truss $((k_c, k_f)$ -truss) in G
$Q, \Delta Q$	A set of query vertices in $V(G)$ and updates to this set
Q' or $Q \oplus \Delta Q$	The set of query vertices after applying ΔQ to Q
$\mathcal{H}_{k_c, k_f, Q}(G)$	The maximal (k_c, k_f) -truss community for Q given G

Definition 1. Cycle Support [28]. Given a directed graph G and an edge $e = \langle u, v \rangle \in E(G)$, the **cycle support** of e in G is the number of vertices that form cycle triangles with e in G , denoted by $\text{sup}_G^c(e) = |\{w \in V(G) \mid \Delta_{u,v,w}^c \text{ in } G\}|$. We denote the **cycle set** of e by $N_G^c(e) = \{w \mid w \in N^{\text{in}}(u) \cap N^{\text{out}}(v)\}$, which stores all vertices that form cycle triangles with e in G . Let all edges that are spanned between $N_G^c(e)$ and $\{u, v\}$ be $E_G^c(e)$.

Definition 2. Flow Support [28]. Given a directed graph G and an edge $e = \langle u, v \rangle \in E(G)$, the **flow support** of e in G is defined similarly to the cycle support as $\text{sup}_G^f(e) = |\{w \in V(G) \mid \Delta_{u,v,w}^f \text{ in } G\}|$. We denote the **flow set** of e by $N_G^f(e) = \{w \mid w \in (N^{\text{in}}(u) \cap N^{\text{in}}(v)) \cup (N^{\text{out}}(u) \cap N^{\text{in}}(v)) \cup (N^{\text{out}}(u) \cap N^{\text{out}}(v))\}$, which stores all vertices that form flow triangles with e in G . Let all edges that are spanned between $N_G^f(e)$ and $\{u, v\}$ be $E_G^f(e)$.

Definition 3. Maximal D-truss [28]. Given a directed graph G and two non-negative integers k_c and k_f , a subgraph $H_{k_c, k_f} \subseteq G$ is a D-truss $((k_c, k_f)$ -truss) if

- $\text{sup}_G^c(e) \geq k_c$ and $\text{sup}_G^f(e) \geq k_f, \forall e \in E(H_{k_c, k_f})$;
- H_{k_c, k_f} is maximal (i.e. H_{k_c, k_f} is not contained in any other (k_c, k_f) -trusses of G).

The pair (k_c, k_f) is called a **trussness** of e , where k_c is the **cycle truss number** of e and k_f is the **flow truss number** of e . Note that for a single edge $e \in E(G)$, it can be contained in different D-trusses, so there may exist multiple trussnesses for it. We denote the **trussness set** of e as $\mathcal{T}(e) = \{(k_{c_i}, k_{f_i}) \mid e \in E(H_{c_i, f_i})\}$.

Based on the definitions above, we formally define a series of problems of maximal D-truss search below. First, we define the problem of maximal D-truss search (MDS) in Problem Statement 1. **Problem Statement 1. Maximal D-truss Search (MDS)** [28]. Given a directed graph G , two non-negative integers k_c and k_f , and a set of query vertices $Q \subseteq V(G)$, the maximal D-truss search (MDS) is to find a subgraph $\mathcal{H}_{k_c, k_f, Q}(G)$ such that: 1) $\mathcal{H}_{k_c, k_f, Q}(G)$ is connected and $Q \subseteq V(\mathcal{H}_{k_c, k_f, Q}(G))$; 2) $\mathcal{H}_{k_c, k_f, Q}(G)$ is a subgraph of the (k_c, k_f) -truss (i.e. $\mathcal{H}_{k_c, k_f, Q}(G) \subseteq H_{k_c, k_f}$); 3) $\mathcal{H}_{k_c, k_f, Q}(G)$ is the maximal subgraph that satisfying two conditions above.

We extend the problem of MDS to two advanced problems considering the scenario of dynamic directed graphs below.

Problem Statement 2. Maximal D-truss Maintenance (MDM) Given a directed graph G and a set of to-be-updated edges ΔG , the problem of maximal D-truss maintenance (MDM) is to compute new trussness set of any edges $e \in E(G \oplus \Delta G)$, where \oplus means that changes are applied. We use G' for short of $G \oplus \Delta G$.

Problem Statement 3. Maximal D-truss Search Maintenance (MDSM). Given a directed graph G , the previously queried maximal D-truss $\mathcal{H}_{k_c, k_f, Q}$ in G , a set of to-be-updated edges ΔG , new thresholds k'_c and k'_f and a set of to-be-updated query vertices ΔQ , the maximal D-truss search maintenance (MDSM) is to search the new maximal D-truss $\mathcal{H}_{k'_c, k'_f, Q'}(G')$.

Example 3.1. Consider the given graph in Figure 1 (a), let (k_c, k_f) be $(1, 1)$ and let Q be $\{v_2\}$, in this way the result of MDS is the subgraph H_3 . Then in Figure 1 (b), let ΔG be $\{\ominus \langle v_1, v_3 \rangle, \ominus \langle v_3, v_1 \rangle, \ominus \langle v_2, v_4 \rangle, \oplus \langle v_5, v_2 \rangle\}$, let (k'_c, k'_f) be $(0, 1)$, let ΔQ_1 be $\{\oplus v_5\}$ and let ΔQ_2 be $\{\ominus v_2, \oplus v_3\}$. For $Q \oplus \Delta Q_1$, the result of MDSM is J_1 , since both v_2 and v_5 are contained in the maximal $(0, 1)$ -truss. For $Q \oplus \Delta Q_2$, the result of MDSM is null, it is obvious that v_3 is not contained in the maximal $(0, 1)$ -truss J_1 .

Note that the solution of MDS performs inefficiently when ΔG is given, because it have to recompute the result from scratch. The solution of MDM is expected to handle ΔG much more efficiently, which is the data-update of MDS. What's more, the solution of MDSM shall answer the new query by maintaining the old result, which is the query-update of MDS.

4 THEORETICAL ANALYSIS OF MDM

For proposing the incremental solutions for the MDM problem, we need to analyze the hardness of MDM first in this section.

4.1 The Theoretical Model and Metrics

Following many works on the boundedness of graph problems [2, 14, 32, 42], we require \mathcal{A}_Δ to be *locally persistent*. An algorithm is locally persistent [2, 32] if, 1) it may use a block of storage for each edge, where it saves pointers to (the blocks of storage for) its adjacent edges; 2) no global auxiliary information is allowed, such as the blocks of each edge are not allowed to save pointers to non-adjacent

edges; 3) the block of edge e may contain auxiliary status information $\text{status}(e)$. Given the updates ΔG , a locally persistent algorithm \mathcal{A}_Δ starts from the pointers of edges in $E(\Delta G)$ and traverses the original graph G via tracing pointers. The next-to-be-visited edge only depends on the status of all visited edges.

For the MDM problem, we denote the set of edges whose trussness set changes by CHANGED. An incremental algorithm \mathcal{A}_Δ is *bounded* [38] if its cost is a polynomial function of $\|\text{CHANGED}\|_c$, where $\|\cdot\|_c$ denotes the size of c -hop neighbors for a positive integer c . The problem of MDM is bounded if there exists a bounded \mathcal{A}_Δ , or is unbounded otherwise. However, the boundedness is such a strong property that many graph problems cannot be solved boundedly [15]. Regarding this gap of hardness, the *relative boundedness* [14] is proposed for measuring the efficiency of incremental algorithms for graph problems. Let a batch algorithm be \mathcal{A} , the data accessed by \mathcal{A} during the computing be $G_{\mathcal{A}}$, the difference between $(G \oplus \Delta G)_{\mathcal{A}}$ and $G_{\mathcal{A}}$ be AFF. An incremental algorithm \mathcal{A}_Δ is bounded relative to \mathcal{A} if its cost is a polynomial function of $|\text{AFF}|$. Existing works [14, 15, 42] show that an unbounded incremental algorithm is still practically efficient if it is of relative boundedness.

4.2 The Boundedness Discussion of MDM

Existing works study the maintenance problems of other cohesive subgraph structures, including the truss maintenance [21, 42] and the core maintenance [27, 35] for undirected graphs. It is proved that both the truss maintenance and the core maintenance are asymmetric regarding the boundedness of edge insertions and deletions [42]. We show the theoretical conclusion in Theorem. 4.1 similarly and prove it in the rest of this section.

THEOREM 4.1. *Under the model of locally persistent algorithms, the problem of MDM is bounded for edge deletions, but unbounded for edge insertions.*

The unboundedness of edge insertions. The proof sketch is similar to the proof for the k -truss [42]. We construct a directed graph G and two specific sets of edge updates ΔG_1 and ΔG_2 . We list two key points of the proof sketch, 1) By applying ΔG_1 and ΔG_2 to the original graph G respectively, the size of $|\text{CHANGED}|$ for each update is constant (i.e. $O(f(\|\text{CHANGED}\|_c)) = O(1)$). If there exists a bounded algorithm \mathcal{A}_Δ , it can obtain the correct result with a constant cost for ΔG_1 and ΔG_2 respectively; 2) For any locally persistent algorithms, the cost of applying ΔG_1 and the cost of applying ΔG_2 are $\Omega(l)$, where l is not constant.

Note that the second point directly contradicts the existence of the bounded algorithm. Because if there exists a bounded algorithm, the cost of applying ΔG_1 plus the cost of applying ΔG_2 should also be constant. According to the proof sketch introduced above, we construct the directed graph G in Figure 2. Let the insertion of $e_1 = \langle v_0, u_0 \rangle$ be the update ΔG_1 and the insertion of $e_2 = \langle v_1, u_1 \rangle$ be the update ΔG_2 . Considering all edges in the original graph, we have the trussness set for each edge $e \in E(G)$ marked in Figure 2. We assume that there exists a bounded insertion algorithm \mathcal{A}_Δ for the MDM problem. First, after applying ΔG_1 singly, $\text{CHANGED} = \{e_1, \langle v_0, u_1 \rangle, \langle u_0, v_1 \rangle, \langle v_1, u_1 \rangle\}$, where these 4 edges have their trussness changed to $(0, 2)$. This implies that \mathcal{A}_Δ costs $O(f(\|\text{CHANGED}\|_c)) = O(1)$ time for the update ΔG_1 .

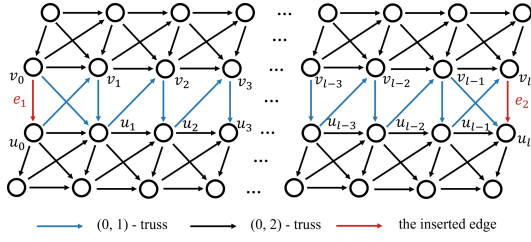


Figure 2: The unboundedness of edge insertions for MDM

Similarly, CHANGED is also of constant size for the update ΔG_2 singly, where $\text{CHANGED} = \{e_2, \langle v_{l-1}, u_l \rangle, \langle u_{l-1}, v_l \rangle, \langle v_{l-1}, u_{l-1} \rangle\}$. The time cost for the update ΔG_2 is $O(f(\|\text{CHANGED}\|_c)) = O(1)$, too. Let $E_{\mathcal{A}}(G, \Delta G)$ be the sequence of edges that \mathcal{A}_{Δ} visits for applying ΔG to G . Both $|E_{\mathcal{A}}(G, \Delta G_1)|$ and $|E_{\mathcal{A}}(G, \Delta G_2)|$ are of constant size. Then, we denote the subgraph that applies ΔG_1 to G by H , and then we compare the sequences visited by \mathcal{A}_{Δ} between applying ΔG_2 to G and H . According to the definition of the boundedness, \mathcal{A}_{Δ} chooses the next edge only depending on all visited edges. Thus the sequences are the same for $G \oplus \Delta G_2$ and $H \oplus \Delta G_2$ until an edge e is visited whose status(e) are different in these two sequences. Such e must exist because CHANGED for these two updates in Figure 2 are different. Furthermore, this e is definitely visited in $E_{\mathcal{A}}(G, \Delta G_1)$ since status(e) are different in H and G . In this way, $E_{\mathcal{A}}(G, \Delta G_1)$ overlaps $E_{\mathcal{A}}(G, \Delta G_2)$ with e . For any locally persistent algorithms, there exist a path from e_1 to e in $E_{\mathcal{A}}(G, \Delta G_1)$ and a path from e_2 to e in $E_{\mathcal{A}}(G, \Delta G_2)$ respectively, then we obtain a path between e_1 and e_2 with the length of $\Omega(l)$ which is obvious in Figure 2. This implies that $|E_{\mathcal{A}}(G, \Delta G_1)| + |E_{\mathcal{A}}(G, \Delta G_2)| = \Omega(l)$, which contradicts the fact that both $|E_{\mathcal{A}}(G, \Delta G_1)|$ and $|E_{\mathcal{A}}(G, \Delta G_2)|$ are of constant size. Above all, \mathcal{A}_{Δ} cannot be bounded for edge insertions of MDM.

The boundedness of edge deletions. For the boundedness of edge deletions of the MDM problem, we provide the proof in Section 7.1, where we present a bounded batch-edge-deletion algorithm.

5 BASIC SINGLE-UPDATE ALGORITHMS

In this section, we consider the cases of single-edge deletion and insertion for the MDM problem, since the single update is the most basic scenario for dynamic directed graphs. Different from the k -truss maintenance problem [42], consider a D-truss, the fact that the same three vertices can be contained in multiple triangles means the existing index [42] cannot have the cycle trussness and the flow trussness correctly updated. Before presenting the proposed algorithms, we introduce some typical operations of the support maintenance. Then we elaborate on algorithms and conduct the complexity analysis for them regarding Theorem 4.1. Without loss of generality, we assume all updates are valid, it must delete an existing edge or insert a new edge for each update. Refer to the full paper [34] for the proofs of all lemmas and theorems.

5.1 Support Single-Update Operators

For the k -truss, the support of each edge $e = (u, v)$ is the size of the intersection of the neighbors of u and v , which is equal to the number of triangles that e is contained in. Existing works [21, 42] observe the fact that each single edge deletion (resp. insertion)

results in the decrement (resp. the increment) of supports by 1 for a specific range of edges. However, this fact does not hold for the D-truss. Because in directed graphs, (1) there are two different types of supports, (2) two edges are allowed between the same vertices (i.e. $\langle u, v \rangle$ and $\langle v, u \rangle$). We propose 4 operators for single-update scenarios below, applying these operators to maintain supports can directly update the trussness set of some edges in CHANGED to save the computation cost. To distinguish different connections between two vertices u and v in directed graphs, we say u and v are *uni-linked* if there exists an edge either $\langle u, v \rangle$ or $\langle v, u \rangle$, they are *bi-linked* if there exist both $\langle u, v \rangle$ and $\langle v, u \rangle$, and they are *not linked* if there exists no edge between them.

LEMMA 5.1. *Given a directed graph G , consider any two vertices $u, v \in V(G)$. For the insertion $e^+ = \langle u, v \rangle$ (resp. the deletion $e^- = \langle u, v \rangle$) if u and v are not uni-linked before $\oplus e^+$ (resp. u and v are not bi-linked before $\oplus e^-$), then for any affected edge, its cycle support, or its flow support, or both increase (resp. decrease) by at most 1.*

The affected edges are composed of two parts, one is the changed edge e^* (i.e. the inserted edge e^+ or the deleted edge e^-), and the other is all edges that are incident to all vertices in $N_G^c(e^*)$ and $N_G^f(e^*)$. For the changed edge e^* , it is inevitable to compute its supports from scratch, the cost of computing is different for the insertion and the deletion. It costs $O(\deg(u) + \deg(v))$ for the insertion because of computing the intersection between the neighbor sets of two endpoints u and v , and it only costs $O(1)$ for the deletion by zeroing them. For the the other part, it costs $O(\sum_z (\deg(z) + \deg(N(z))))$ for recomputing them from scratch, where z denotes the shared neighbors of u and v and is bounded by $O(\min(\deg(u), \deg(v)))$. We observe that it is not necessary to recompute supports of the latter part of affected edges but it can be maintained by two proposed operators $\text{Ins}(e^+)$ and $\text{Del}(e^-)$. Both operators can update supports efficiently for all edges that are incident to all vertices in $N_G^c(e^*)$ and $N_G^f(e^*)$ with the time cost of $O(\sum_z \deg(z))$ only. The correctness is guaranteed by Lemma 5.1.

$\text{Ins}(e^+)$. We compute supports for the newly inserted edge e^+ . According to Lemma 5.1, for any edge e^c , its cycle support increases by 1; for any edge e^f , its flow support increases by 1; note that there may exist the same edges in both $E_G^c(e^+)$ and $E_G^f(e^+)$.

$$\text{Ins}(e^+) \begin{cases} \sup_G^c(e^+) \leftarrow |N_G^c(e^+)|; \sup_G^c(e^c) \leftarrow \sup_G^c(e^c) + 1, \forall e^c \in E_G^c(e^+); \\ \sup_G^f(e^+) \leftarrow |N_G^f(e^+)|; \sup_G^f(e^f) \leftarrow \sup_G^f(e^f) + 1, \forall e^f \in E_G^f(e^+). \end{cases}$$

$\text{Del}(e^-)$. For the deleted edge, we zero its support. Similarly to $\text{Ins}(e^+)$, we only need to update the cycle support of e^c and the flow support of e^f .

$$\text{Del}(e^-) \begin{cases} \sup_G^c(e^-) \leftarrow 0; \sup_G^c(e^c) \leftarrow \sup_G^c(e^c) - 1, \forall e^c \in E_G^c(e^-); \\ \sup_G^f(e^-) \leftarrow 0; \sup_G^f(e^f) \leftarrow \sup_G^f(e^f) - 1, \forall e^f \in E_G^f(e^-). \end{cases}$$

Now we consider the cases that u and v are *bi-linked* before and after applying the changed edge $e^* = \langle u, v \rangle$, we propose operators $\text{Enga}(e^+)$ and $\text{DisEnga}(e^-)$, and the correctness of them is shown in Lemma 5.2.

LEMMA 5.2. *Given a directed graph G , consider any two vertices $u, v \in V(G)$. For the insertion $e^+ = \langle u, v \rangle$ (resp. the deletion $e^- = \langle u, v \rangle$), if u and v are uni-linked before $\oplus e^+$ (resp. u and v are bi-linked*

before $\oplus e^-$), then $\text{Enga}(e^+)$ (resp. $\text{DisEnga}(e^-)$) updates supports correctly for all affected edges.

$\text{Enga}(e^+)$. Consider a single edge insertion $e^+ = \langle u, v \rangle$, it is known that u and v are *uni-linked* before the insertion (i.e. there exists an edge $\tilde{e}^+ = \langle v, u \rangle \in E(G)$). For e^+ , it costs $O(\deg^{\text{in}}(u) + \deg^{\text{out}}(v))$ for updating its cycle support and flow support. The cycle support increases by 1 for any edge e^c and the flow support increases by 1 for any edge e^f with the time cost of $O(\deg(u) + \deg(v))$.

$$\text{Enga}(e^+) \begin{cases} \sup_{G'}^c(e^+) \leftarrow |\text{N}^{\text{in}}(u) \cap \text{N}^{\text{out}}(v)|; \\ \sup_{G'}^c(e^c) \leftarrow \sup_G^c(e^c) + 1, \\ \forall e^c \in \{\langle v, w \rangle, \langle w, u \rangle \mid \\ \forall w \in \text{N}^{\text{in}}(u) \cap \text{N}^{\text{out}}(v) \setminus \text{N}^{\text{in}}(v) \cap \text{N}^{\text{out}}(u)\}; \\ \sup_{G'}^f(e^+) \leftarrow \sup_G^f(\tilde{e}^+) - \sup_{G'}^c(e^+) + \sup_G^f(\tilde{e}^+); \\ \sup_{G'}^f(e^f) \leftarrow \sup_G^f(e^f) + 1, \\ \forall e^f \in \{\langle u, x \rangle, \langle x, v \rangle \mid \\ \forall x \in \text{N}^{\text{in}}(v) \cap \text{N}^{\text{out}}(u) \setminus (\text{N}(u) \cap \text{N}^{\text{out}}(v)) \cup (\text{N}^{\text{in}}(u) \cap \text{N}^{\text{in}}(v))\}. \end{cases}$$

$\text{DisEnga}(e^-)$. Consider a single edge deletion $e^- = \langle u, v \rangle$, it is known that u and v are *bi-linked* before the deletion. For e^- , it costs $O(1)$ for zeroing supports. The cycle support decreases by 1 for any edge e^c and the flow support decreases by 1 for any edge e^f with the time cost of $O(\deg(u) + \deg(v))$.

$$\text{DisEnga}(e^-) \begin{cases} \sup_{G'}^c(e^-) \leftarrow 0; \sup_{G'}^c(e^c) \leftarrow \sup_G^c(e^c) - 1, \\ \forall e^c \in \{\langle v, w \rangle, \langle w, u \rangle \mid \\ \forall w \in \text{N}^{\text{in}}(u) \cap \text{N}^{\text{out}}(v) \setminus \text{N}^{\text{in}}(v) \cap \text{N}^{\text{out}}(u)\}; \\ \sup_{G'}^f(e^-) \leftarrow 0; \sup_{G'}^f(e^f) \leftarrow \sup_G^f(e^f) - 1, \\ \forall e^f \in \{\langle u, x \rangle, \langle x, v \rangle \mid \\ \forall x \in \text{N}^{\text{in}}(v) \cap \text{N}^{\text{out}}(u) \setminus (\text{N}(u) \cap \text{N}^{\text{out}}(v)) \cup (\text{N}^{\text{in}}(u) \cap \text{N}^{\text{in}}(v))\}. \end{cases}$$

5.2 Single Edge Deletion Algorithm

We denote the to-be-deleted edge by $e^- = \langle u, v \rangle$ and denote the updated skyline trussness sets by $\mathcal{ST}'(e)$ for $\forall e \in E(G \setminus e^-)$. First, we discuss the range of affected skyline trussness sets in Lemma 5.3.

LEMMA 5.3. *Given a directed graph G , consider the deletion $e^- = \langle u, v \rangle \in E(G)$. For each edge $e \in E_G^c(e^-) \cup E_G^f(e^-)$ and its skyline trussness set $\mathcal{ST}(e) = \{(k_{c_i}, k_{f_i}) \mid \text{for all } i\}$, (1) if (k_{c_i}, k_{f_i}) dominates any skyline trussness $(k_c(e^-) + 1, k_f(e^-) + 1)$ in $\mathcal{ST}(e^-)$, then (k_{c_i}, k_{f_i}) will NOT be affected; (2) otherwise (k_{c_i}, k_{f_i}) may be affected.*

We present an algorithm that can update an edge deletion correctly for the MDM problem in Algorithm 1. In line 1, we initialize variables. In line 2, we update supports of corresponding edges with operators $\text{DisEnga}(e^-)$ and Del . In lines 4-5, we start processing the queue that is used for detecting affected edges. We check each trussness of the current edge $\langle u_1, v_1 \rangle$ in lines 6-12. In lines 7-9, we maintain trussness sets that can be done instantly after support maintenance. In line 10, we peel off edges whose cycle support is not qualified. We check the local cycle support and the local flow support of $\langle u_1, v_1 \rangle$ in lines 11-12. At last, we mark $\langle u_1, v_1 \rangle$ and continue to push edges into *Queue* in lines 13-14. When there is no edge in *Queue*, we have all skyline trussness sets correctly updated and return them as $\mathcal{ST}'(e)$, $\forall e \in E(G')$.

5.3 Single Edge Insertion Algorithm

We denote the to-be-inserted edge by $e^+ = \langle u, v \rangle$ and denote the updated skyline trussness set by $\mathcal{ST}'(e)$ for $\forall e \in E(G \oplus e^+)$. First, we discuss the range of affected skyline trussness set in Lemma 5.4.

Algorithm 1: Single Edge Deletion Algorithm

Input: $G, e^- = \langle u, v \rangle, \mathcal{ST}(e), \forall e \in E(G)$
Output: $\mathcal{ST}'(e), \forall e \in E(G \setminus e^-)$

- 1 $G' \leftarrow G \setminus e^-; \text{Queue} \leftarrow \emptyset; H \leftarrow \emptyset; \text{DelCheck2}(e^-);$
- 2 Call $\text{DisEnga}(e^-)$ if $\langle v, u \rangle \in E(G)$, or call $\text{Del}(e^-)$ otherwise;
- 3 **while** !*Queue.empty()* **do**
- 4 $H \leftarrow G'; \langle u_1, v_1 \rangle \leftarrow \text{Queue.top}();$
- 5 **if** $\langle u_1, v_1 \rangle$ is NOT visited **then**
- 6 **for each** k_{c_i} in associated trussnesses of $\langle u_1, v_1 \rangle$ **do**
- 7 **if** $k_{c_i} == \sup_G^c(\langle u_1, v_1 \rangle)$ and
- 8 $\sup_{G'}^c(\langle u_1, v_1 \rangle) < \sup_G^c(\langle u_1, v_1 \rangle)$ **then**
- 9 $k_c \leftarrow k_c - 1;$
- 10 Repeat lines 7-8 similarly for the associated $k_{f_i};$
- 11 $H \leftarrow H \setminus e', \text{ s.t. } k_c(e') < k_{c_i};$
- 12 **if** $\text{LCS}(\langle u_1, v_1 \rangle, H) < k_{c_i}$ **then** $k_{c_i} \leftarrow k_{c_i} - 1;$
- 13 **if** $\text{LFS}(\langle u_1, v_1 \rangle, H) < k_{f_i}$ **then** $k_{f_i} \leftarrow k_{f_i} - 1;$
- 14 Mark $\langle u_1, v_1 \rangle$ as visited;
- 15 **if** $\langle u_1, v_1 \rangle$ is affected **then** $\text{DelCheck2}(\langle u_1, v_1 \rangle);$
- 16 **return** $\mathcal{ST}'(e), \forall e \in E(G');$

Procedure. $\text{DelCheck2}(e)$
Put all unvisited edges in $E_G^c(e) \cup E_G^f(e)$ together with their skyline trussness set belong to case (2) in Lemma 5.3 into *Queue*;
Procedure. $\text{LCS}(e, H)$
Return the cycle support of the edge e in the subgraph H ;
Procedure. $\text{LFS}(e, H)$
Return the flow support of the edge e in the subgraph H ;

LEMMA 5.4. *Given a directed graph G , consider the insertion $e^+ = \langle u, v \rangle$. Then for each edge $e \in E_G^c(e^+) \cup E_G^f(e^+)$ and its skyline trussness set $\mathcal{ST}(e) = \{(k_{c_i}, k_{f_i}) \mid \text{for all } i\}$, if (k_{c_i}, k_{f_i}) dominates any skyline trussness $(k_c'(e^+) + 1, k_f'(e^+) + 1)$ in $\mathcal{ST}(e^+)$, then (k_{c_i}, k_{f_i}) will NOT be affected, otherwise (k_{c_i}, k_{f_i}) may be affected.*

We present a solution that can update an edge insertion correctly for the MDM problem in Algorithm 2. In lines 1-3, we initialize variables. We call operators to update supports in line 4. In lines 5-15, we maintain edges in $E_G^c(e^+) \cup E_G^f(e^+)$ after applying operators. In line 14, note that the support checked within this loop has been correctly updated, in this way, $\sup_{G'}(e) > k$ can guarantee an increase in trussness. Because for those edges who forms triangles with e^+ and have their cycle truss or the flow truss of k , the increase in trussness is guaranteed with operators updating supports correctly. In line 16, we compute the skyline trussness set for e^+ . We start processing the queue in lines 17-18 and check the change of cycle truss numbers and flow truss numbers successively in lines 19-35. In lines 20-22, we mark edges whose cycle truss number may increase and append the queue. In lines 23-25, we unmark edges whose cycle truss number remains the same. In lines 26-34, we conduct the checking for flow truss numbers similarly. We update trussness sets according to their marks in line 35 and return the result in line 36.

5.4 Theoretical Analysis

THEOREM 5.5. *For Algorithm 1, the time complexity is $O(\Delta k_c \cdot \sum_{e \in E_{DC2}} \deg(u_1) + \deg(v_1))$, where $e = \langle u_1, v_1 \rangle$ and $|E_{DC2}|$ is the number of edges that are inspected by DelCheck2 in Algorithm 1 and is bounded by $\|\text{CHANGED}\|_1$, and the space complexity is $O(m)$.*

THEOREM 5.6. *For Algorithm 2, the time complexity is $O(\Delta k_c \cdot \sum_{e \in E_{IC2}} \deg(u_1) + \deg(v_1))$, where $e = \langle u_1, v_1 \rangle$ and $|E_{IC2}|$ is the number of edges that are inspected by InsCheck2 in Algorithm 2, and the space complexity is $O(m')$.*

Algorithm 2: Single Edge Insertion Algorithm

Input: $G, e^+ = \langle u, v \rangle, \mathcal{ST}(e), \forall e \in E(G)$
Output: $\mathcal{ST}'(e), \forall e \in E(G \oplus e^+)$

- 1 $G' \leftarrow G \oplus e^+; Queue \leftarrow \emptyset; H \leftarrow \emptyset; \text{InsCheck2}(e^+);$
- 2 **for each** $e \in E(G')$ **do**
- 3 $c_{\text{inc}}(e) \leftarrow \text{false}; f_{\text{inc}}(e) \leftarrow \text{false}; D_{\text{unc}}(e) \leftarrow \text{false};$
- 4 Call $\text{Enga}(e^+)$ if $\langle v, u \rangle \in E(G)$, or call $\text{Ins}(e^+)$ otherwise;
- 5 $k_{\text{cmin}} \leftarrow \min\{k_c \mid (k_c, k_f) \in \mathcal{ST}'(e'), \forall e' \in E_G^c(e^+) \cup E_G^f(e^+);$
- 6 $k_{\text{cmax}} \leftarrow \max\{k_c \mid (k_c, k_f) \in \mathcal{ST}'(e'), \forall e' \in E_G^c(e^+) \cup E_G^f(e^+);$
- 7 **for** $k_c \leftarrow k_{\text{cmin}}$ **to** k_{cmax} **do**
- 8 **if** $\text{sup}_{G'}^c(e^+) == k_c$ **then**
- 9 $k'_c(e^+) \leftarrow k_c;$
- 10 **for each** $e' \in E_G^c(e^+) \cup E_G^f(e^+)$ **do** $D_{\text{unc}}(e') \leftarrow \text{true};$
- 11 **else if** $\text{sup}_{G'}^c(e^+) < k_c$ **then**
- 12 **for each** $e' \in E_G^c(e^+) \cup E_G^f(e^+)$ **do** $D_{\text{unc}}(e') \leftarrow \text{true};$
- 13 **else**
- 14 $k'_c(e') \leftarrow k_c + 1$ s.t. $k_c(e') == k_c;$
- 15 Repeat lines 5-14 similarly for flow truss numbers;
- 16 Compute the skyline trussness set for $e^+;$
- 17 **while** $!Queue.empty()$ **do**
- 18 $H \leftarrow G'; \langle u_1, v_1 \rangle \leftarrow Queue.top();$
- 19 **for each** k_{c_i} in associated trussnesses (k_{c_i}, k_{f_i}) of $\langle u_1, v_1 \rangle$ **do**
- 20 **if** $\text{LCUB}(\langle u_1, v_1 \rangle, k_{c_i}) \geq k_{c_i} + 1$ **then**
- 21 **if** $!c_{\text{inc}}(\langle u_1, v_1 \rangle)$ **then**
- 22 $c_{\text{inc}}(\langle u_1, v_1 \rangle) \leftarrow \text{true}; \text{InsCheck2}(\langle u_1, v_1 \rangle);$
- 23 **else**
- 24 **if** $c_{\text{inc}}(\langle u_1, v_1 \rangle)$ **then**
- 25 $c_{\text{inc}}(\langle u_1, v_1 \rangle) \leftarrow \text{false};$
- 26 **if** $\text{LFUB}(\langle u_1, v_1 \rangle, k_{f_i}) \geq k_{f_i} + 1$ **then**
- 27 **if** $!f_{\text{inc}}(\langle u_1, v_1 \rangle)$ **then**
- 28 $f_{\text{inc}}(\langle u_1, v_1 \rangle) \leftarrow \text{true};$
- 29 $\text{InsCheck2}(\langle u_1, v_1 \rangle);$
- 30 **else**
- 31 **if** $f_{\text{inc}}(\langle u_1, v_1 \rangle)$ **then**
- 32 $f_{\text{inc}}(\langle u_1, v_1 \rangle) \leftarrow \text{false};$
- 33 $D_{\text{unc}}(\langle u_1, v_1 \rangle) \leftarrow \text{true};$
- 34 $\text{InsCheck2}(\langle u_1, v_1 \rangle);$
- 35 **if** $c_{\text{inc}}(e)$ or $f_{\text{inc}}(e)$ **then** Update k_{c_i} or k_{f_i} by 1;
- 36 **return** $\mathcal{ST}'(e), \forall e \in E(G');$

Procedure. $\text{InsCheck2}(e)$
Put all unvisited edges in $E_G^c(e) \cup E_G^f(e)$ belong to case (2) together with their undominated skyline trussnesses into $Queue$;

Procedure. $\text{LCUB}(e, k_c)$
Compute the local cycle support upper bound of e by counting the number of cycle supports where the other two edges in the Δ^c have their $k_{c_i} \geq k_c$ and $D_{\text{unc}}(\cdot)$ is false ;

Procedure. $\text{LFUB}(e, k_f)$
Compute the local flow support upper bound of e by counting the number of flow supports where the other two edges in the Δ^f have their $k_{f_i} \geq k_f$ and $D_{\text{unc}}(\cdot)$ is false ;

6 EFFICIENT ORDER-BASED INDEX

Though the MDM problem can be solved by single update algorithms proposed above, there are 2 weaknesses: (1) both algorithms are only able to handle to-be-updated edges one by one, (2) Algorithm 2 cannot be proved to show neither boundedness nor relative boundedness for edge insertions. In this section, we propose an order-based index called D-Index, which can perform batch updates. We present the index design first and then illustrate its rationale for maintenance. Corresponding batch-update algorithms and query algorithms are further presented in Section 7. Refer to the full paper [34] for the proofs of theorems.

6.1 D-Index Overview

The existing index [42] of the k -truss stores edges in the non-ascending order of k values. However, there are a pair of values to consider (i.e. (k_c, k_f)) for the D-truss, the dominant property of (k_c, k_f) [28] confirms that it is impossible to construct such a single order for the two dimensions of trussnesses in the D-truss. This fact also increases the problem dimension for handling two types of trussness compared to k -truss, which makes the MDM problem harder. For addressing the hardness, we design our D-Index and batch-update algorithms as follows.

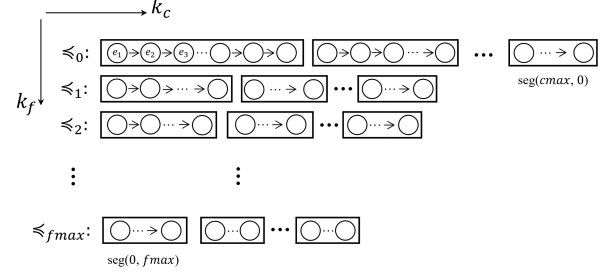


Figure 3: The structure of D-Index

Definition 6. Cycle Decomposition Order (CD Order). Given a directed graph G and a flow truss number k_{f_i} , the cycle decomposition order (CD order) of k_{f_i} is a sequence of edges in the subgraph $H_{0, k_{f_i}}$ that follows the order of the decomposition along cycle truss numbers, we denote it by $\leq_{k_{f_i}} = \{e_1, e_2, \dots, e_{|H|}\}$. For any $\leq_{k_{f_i}}$ of G , it is divided into segments where edges within a segment share the same cycle truss number. We denote a segment by $\text{seg}(k_{c_j}, k_{f_i})$.

Then a CD Order $\leq_{k_{f_i}}$ can be represented by segments $\leq_{k_{f_i}} = \{\text{seg}(0, k_{f_i}), \text{seg}(1, k_{f_i}), \text{seg}(2, k_{f_i}), \dots, \text{seg}(k_{c_j}, k_{f_i})\}$. Figure 3 illustrates the structure of the D-Index. Given a directed graph G , the D-Index of G is the list of all $\leq_{k_{f_i}}$ in the ascending order of k_{f_i} , we denote the D-Index of G and G' by \mathbb{I}_D and \mathbb{I}'_D respectively.

This index can be inherently constructed by the D-truss decomposition algorithm [28]. Note that there may exist many valid CD orders for a single k_{f_i} , since all edges within the same segment can be peeled off in any orders during the process of D-truss decomposition. We keep an arbitrary valid CD order for each k_{f_i} in D-Index and maintain these CD orders one by one for the MDM problem.

6.2 The Rationale of D-Index

Maintenance in a CD order. Given the updated edges for this CD order, the maintenance of it is illustrated below. Since we fix k_{f_i} for each $\leq_{k_{f_i}}$, we only need to decide new cycle truss numbers for all edges in $\leq_{k_{f_i}}$ and the maintenance can be deemed as the movement of edges among segments. Figure 4 illustrates the pattern of such movement, which is extended from the order index [42].

We consider edge insertions first for the ease of illustration, the details for edge deletions are presented in Section 7.1. In Figure 4, we choose an arbitrary CD order $\leq_{k_{f_i}}$ and itself after maintaining $\leq'_{k_{f_i}}$. We use 3 segments to show the principle of maintaining this CD order (i.e. $\text{seg}(k_{c_j}, k_{f_i}), \text{seg}(k_c, k_{f_i})$ and $\text{seg}(k_c + 1, k_{f_i})$), and themselves after applying updates (i.e. $\text{seg}'(k_{c_j}, k_{f_i}), \text{seg}'(k_c, k_{f_i})$).

and $\text{seg}'(k_c + 1, k_{f_i})$). Within $\text{seg}(k_c, k_{f_i})$, we mark edges whose trussness will change in two boxes, where edges in box1 will move to $\text{seg}'(k_c, k_{f_i})$, and edges in box2 will move to $\text{seg}'(k_c + 1, k_{f_i})$. Within $\text{seg}(k_c, k_{f_i})$, we mark edges whose trussness remain same in box3, and mark edges that will move to $\text{seg}'(k_c + 1, k_{f_i})$ in box4. We denote the edges set where each edge exists in the $(k_c, k_{f_i})'$ -truss but not in the (k_c, k_{f_i}) -truss by $\text{Nseg}(k_c, k_{f_i})$. In this way, we have $\text{Nseg}(k_c, k_{f_i}) = \{\text{Box1} \cup \text{Box2}\}$, $\text{seg}(k_c, k_{f_i}) = \{\text{Box3} \cup \text{Box4}\}$, $\text{Nseg}'(k_c, k_{f_i}) = \{\text{Box2} \cup \text{Box4}\}$ and $\text{seg}'(k_c, k_{f_i}) = \{\text{Box1} \cup \text{Box3}\}$. We can regard all these unions as additions because each pair of boxes is disjoint, we naturally obtain that $\text{Nseg}(k_c, k_{f_i}) \cup \text{seg}(k_c, k_{f_i}) = \text{Nseg}(k_c + 1, k_{f_i}) \cup \text{seg}'(k_c, k_{f_i})$. For each CD order and itself after maintaining, we have such $\text{Nseg} - \text{seg}$ relationship hold for all flow truss numbers k_{f_i} .

Let e^* be the edge we currently visit and we initialize it as the first one in $\leq_{k_{f_i}}$, two sets $C(k_c, k_{f_i})$ and $R(k_c, k_{f_i})$ are prepared for storing $\text{Nseg}(k_c + 1, k_{f_i})$ and $\text{seg}'(k_c, k_{f_i})$. First, we initialize $C(k_c, k_{f_i}) = \text{Nseg}(k_c, k_{f_i})$ and $R(k_c, k_{f_i}) = \emptyset$. For any edge in $\text{seg}(k_c, k_{f_i})$ that is visited before e^* , if it exists in $\text{Nseg}(k_c, k_{f_i})$, then it can be checked whether it exists in $\text{seg}'(k_c, k_{f_i})$ or not. We move it into $R(k_c, k_{f_i})$ if the answer is yes, and move it into $C(k_c, k_{f_i})$ otherwise.

We construct and maintain 3 auxiliary variables for the correctness of maintenance in each CD order, including $\text{rem}_{\leq}(\cdot)$, $s(\cdot)$ and $\text{ext}(\cdot)$. For $\forall e \in \leq_{f_i}$, we denote the number of cycle support of e which exist in the edges after e in \leq_{f_i} by $\text{rem}_{\leq}(e)$. For $\forall e \in C(k_c, k_{f_i})$, we maintain $s(e)$ that is equal to the number of cycle support of e in $(k_c, k_{f_i})'$ -truss $\setminus R(k_c, k_{f_i})$. For the current visiting e^* and edges after e^* in \leq_{f_i} , $\text{ext}(e)$ is maintained such that $\text{ext}(e) + \text{rem}_{\leq}(e)$ always equal the number of cycle support in the subgraph which consists of $C(k_c, k_{f_i})$ and the edges after e in \leq_{f_i} . We design dif-

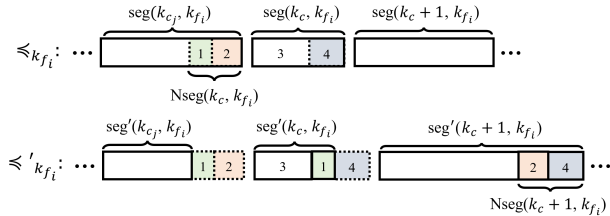


Figure 4: The principle of order-based maintenance idea

ferent procedures for processing all 3 cases of e^* , that is case 1 ($\text{ext}(e^*) = 0$), case 2 ($\text{ext}(e^*) > 0$ and $\text{ext}(e^*) + \text{rem}_{\leq}(e^*) > k_c$) and case 3 ($\text{ext}(e^*) > 0$ and $\text{ext}(e^*) + \text{rem}_{\leq}(e^*) < k_c$). For case 1, we can verify that e^* exists in $\text{seg}'(k_c, k_{f_i})$ straightforwardly, so we remove it from the CD order and move it into $R(k_c, k_{f_i})$. all $s(\cdot)$ and $\text{ext}(\cdot)$ remain same. For case 2, e^* may exist in the $(k_c + 1, k_{f_i})$ -truss of G' , we keep it in the CD order and add it into $C(k_c, k_{f_i})$, all corresponding auxiliary variables need to be maintained according to their definitions. For case 3, e^* cannot exist in the $(k_c + 1, k_{f_i})$ -truss of G' , we remove it from the CD order and move it into $R(k_c, k_{f_i})$, meanwhile, we remove any edge $e \in C(k_c, k_{f_i})$ recursively such that $s(e) \leq k_c$, and update auxiliary variables.

Example 6.1. Consider the graph G in Figure 1(a), assume that we insert a new edge $\langle v_8, v_9 \rangle$ to G , and the procedures that how \leq_0 is

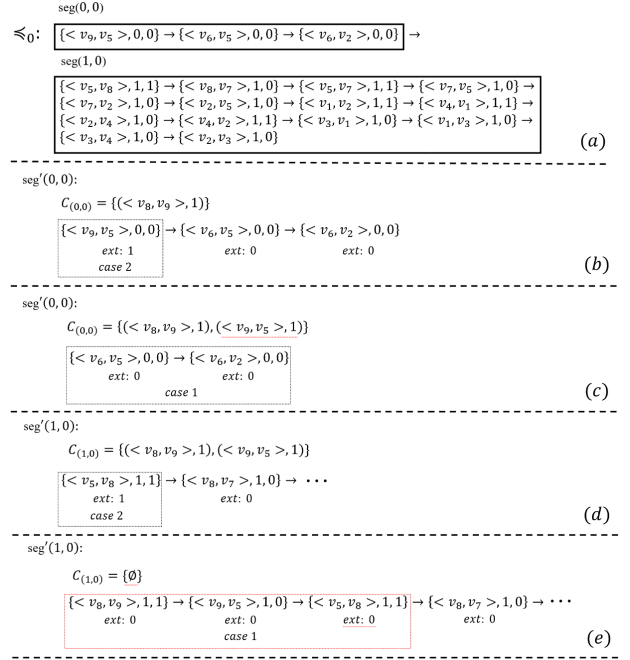


Figure 5: The illustration of Example 6.1

maintained to \leq'_0 are in Figure 5. We underline changes between consecutive subgraphs by red dashed lines. Figure 5 (a) shows a feasible CD order \leq_0 , which consists of $\text{seg}(0, 0)$ and $\text{seg}(1, 0)$. Note that each edge is stored in the format of $(e, k_c(e), \text{rem}_{\leq}(e))$. After the insertion of $\langle v_8, v_9 \rangle$, we start to maintain $\text{seg}(0, 0)$ to $\text{seg}'(0, 0)$. First, in Figure 5 (b), we initialize $C_{(0,0)} = \{(\langle v_8, v_9 \rangle, 1)\}$ in the format of the pair of $(e, s(e))$, and process $C_{(0,0)}$ as in lines 5-15 in Algorithm 4, where $C_{(0,0)}$ remains unchanged. Then we start to process edges in $\text{seg}(0, 0)$, since $\langle v_9, v_5 \rangle$ belongs to case 2, so it is added to $C_{(0,0)}$. In Figure 5 (c), the remained two edges belong to case 1, by now $\text{seg}'(0, 0)$ is obtained and $C_{(0,0)} = \text{Nseg}(1, 0)$.

In Figure 5 (d), we focus on $\text{seg}(1, 0)$ and initialize $C_{(1,0)} = \{(\langle v_8, v_9 \rangle, 1), (\langle v_9, v_5 \rangle, 1)\}$, the checking turns out that they will be removed from $C_{(1,0)}$ and added at the head of $\text{seg}'(1, 0)$. Figure 5 (e) shows $\langle v_8, v_9 \rangle$ and $\langle v_9, v_5 \rangle$ belong to case 1. Due to the change of $C_{(1,0)}$, $\text{ext}(\langle v_5, v_8 \rangle)$ changes from 1 to 0, its case also changes from case 2 to case 1. By now all remained edges belong to case 1 thus are omitted. $C_{(1,0)}$ becomes empty, thus the maintenance to \leq'_0 is finished.

Maintenance among CD orders. Instead of simply maintaining all orders independently, we utilize the dominant property of D-truss to prune unnecessary to-be-checked orders and accelerate the maintenance process as follows. It is obvious that all edges within $\leq_{k_{f_i}}$ exactly form the subgraph $H_{0, k_{f_i}}$. For correctly maintain the index, each $\leq'_{k_{f_i}}$ should form $H'_{0, k_{f_i}}$ after applying edge updates. We start maintaining D-Index from \leq_0 given edge updates E^+ , instead of maintaining CD orders one by one simply, the squeezing search based on the dominant property of D-truss can be conducted below.

For edge deletions, we first obtain $\leq'_{k_{f_i}}$ by maintaining the CD order $\leq_{k_{f_i}}$, then we need to maintain $\leq_{k_{f_i}+1}$, based on the dominant

property, we can pick a tighter upper bound for the cycle truss number of any edge e in $\leq_{k_{f_i}+1}$: $\{\min(k_c^1, k_c^2) \mid e \in \text{seg}(k_c^1, k_{f_i} + 1), e \in \text{seg}'(k_c^2, k_{f_i})\}$. For edge insertions, the cycle truss of e has a lower bound k_c^1 exist in $\leq_{k_{f_i}+1}$, it also has an upper bound k_c^2 exist in $\leq'_{k_{f_i}}$ for $\forall e \in \leq_{k_{f_i}+1}$. Then the maintenance of $\leq_{k_{f_i}+1}$ can be conducted more efficiently from both the head and the tail of the CD order. Note that we need to update E^+ by $E(H_{0,k_{f_i}+1}) \setminus E(H'_{0,k_{f_i}+1})$ for the next to-be-maintained CD order, and if E^+ becomes empty, which implies that no updated edges exist in the next CD order and all remained CD orders are unchanged, in this way, we can early-stop the maintenance correctly and efficiently.

7 INDEX-BASED ALGORITHMS

In this section, we propose the solutions for the problem of MDM and MDSM based on the D-Index, including the algorithms for batch edges deletion and insertion respectively for the MDM problem, the query algorithm for the MDSM problem, together with theoretical analysis for these algorithms.

7.1 Batch Edges Update Algorithms

We denote the batch of to-be-deleted edges by E^- . We present our batch edges deletion algorithm in Algorithm 3. We initialize variables in line 1, note that E_{up} is initialized as E^- and is updated for the next CD order in lines 2-19. In lines 3-7, we start maintaining trussness sets for edges in E_{up} by initializing the queue. In lines 8-16, we maintain the order by auxiliary variables $rem_{\leq}(\cdot)$ and $ts(\cdot)$, where $ts(e)$ is defined as the number of cycle triangles where the other two edges e' and e'' both have their cycle truss numbers no smaller than k_c of e . Note that all existing trussness sets are obtained from \mathbb{I}_D . In lines 17-19, we update the flow truss numbers of edges that exist in the next CD order by picking a tighter upper bound to squeeze the search space. We update E_{up} and necessary variables in line 20 and return the result in line 21.

We denote the batch of to-be-inserted edges by E^+ . Our batch edges insertion algorithm is presented in Algorithm 4. We initialize variables in line 1. As illustrated in Section 6.2, E_{up} is initialized and maintained in lines 2-33. We initialize $C_{(k_c, k_f)}$ is initialized in line 5 and is processed in lines 6-13. In lines 14-15, we update $ext(\cdot)$ for edges after e^* inclusively in \leq_{k_f} . In lines 16-26, process the edge according to 3 cases of different $ext(\cdot)$ and $rem_{\leq}(\cdot)$ mentioned in Section 6.2. We initialize $N\text{seg}_{(k_{c+1}, k_f)}$ with remained edges in $C_{(k_c, k_f)}$ in line 27 and update cycle truss numbers in $N\text{seg}_{(k_c, k_f)} \setminus N\text{seg}_{(k_{c+1}, k_f)}$ in line 28. In lines 30-32, we squeeze the search space for the maintenance of each edge. We update E_{up} and necessary variables in line 33. Note that OIn and ODel can be done in $O(1)$ based on the order maintenance algorithm [11].

7.2 Fully-Dynamic Search Maintenance

Based on the batch edges update algorithms above, we present the fully-dynamic MDSM query algorithm in Algorithm 5. For the given batch edges update which consists of batch edges insertion and deletion, we remove the insertion and deletion of the same edge first, then move deletions in the first half of the updates, and leave insertions in the second half of updates. In this way, we maintain \mathbb{I}_D by calling Algorithm 3 and Algorithm 4 in line 1. In lines 2-3, if the

Algorithm 3: Batch Edges Deletion Algorithm

```

Input:  $G, E^-, \mathbb{I}_D$ 
Output:  $\mathbb{I}'_D$ 
1  $G' \leftarrow G \setminus E^-; E_{up} \leftarrow E^-; Queue \leftarrow \emptyset; k_c \leftarrow 0;$ 
2 while  $!E_{up}.empty()$  do
3   for each edge  $e^- \in E_{up}$  do
4     Remove  $e^-$  from  $\leq_{k_f}$ ;
5     Update  $rem_{\leq}(\cdot)$  and  $ts(\cdot)$  for edges that form a  $\Delta$  with  $e^-$ ;
6     for each edge  $e$  with  $ts(e) < k_c(e)$  in line 5 do
7       if  $e \in Queue$  and  $e \notin E^-$  then  $Queue \leftarrow e;$ 
8   while  $!Queue.empty()$  do
9      $e \leftarrow Queue.top(); k'_c(e) \leftarrow \max\{k_c \mid |\Delta_e^{k_c}| \geq k_c\};$ 
10    for each  $\Delta$  that contains  $e$  in  $G'$  do
11      Let  $e'$  and  $e''$  be the other two edges of  $\Delta$ ;
12      Update  $ts(e')$  by checking  $k'_f(e), k_f(e'), k_f(e'')$ ;
13      if  $ts(e') < k_f(e')$  and  $e' \notin Queue$  then  $Queue \leftarrow e'$ ;
14      Repeat line 12-13 for  $e''$ ;
15    Append  $e$  into  $\leq'_{k_f}$ ;
16    Update  $rem_{\leq}(\cdot)$  and  $ts(\cdot)$  for edges that form triangles with  $e$ ;
17    Conduct maintenance from  $k_{cmax}$  similar to lines 8-16;
18    for each  $e \in \leq_{k_{f+1}}$  do
19       $k_c(e) \leftarrow \{\min(k_c^1, k_c^2) \mid e \in \text{seg}(k_c^1, k_{f_i} + 1), e \in \text{seg}'(k_c^2, k_f)\}$ 
20     $E_{up} \leftarrow E(H_{0,k_{f+1}}) \setminus E(H'_{0,k_{f+1}}); Queue \leftarrow \emptyset; k_f \leftarrow k_f + 1;$ 
21 return  $\mathbb{I}'_D;$ 

```

previous result H is empty, then we need to compute this new query from scratch. In lines 5-8, we consider all edges in the given query result $\mathcal{H}_{k_c, k_f, Q}(G)$ before applying edge updates, find those whose trussnesses can dominate or equal (k'_c, k'_f) and mark them as *kept*. After inspecting trussnesses, we check the connectivity of H in line 9. In line 10, we conduct searching based on H by the index-based query algorithm [28]. In lines 11-13, we check the containment of Q' of the result community H .

The optimization of leveraging all previous results. Algorithm 5 is only able to maintain the query result from the last one. When the results of two consecutive queries are not relevant, the extra checking procedures make the query maintenance algorithm worse than direct query algorithm. It is strongly motivated that making Algorithm 5 utilizes of all previous query results. The optimization can be extended based on Algorithm 5 as follows. With storing previous queries and their result subgraphs in advance, when we receive a new query Q_{i+1} , we can check the existence of query vertices in all previous result subgraphs $H_i, \forall i$ and obtain a candidate set of result subgraphs. Then we compute the difference of ΔG_{i+1} and each ΔG in the candidate set, where the difference is defined as the number of different deletions or insertions. Finally we select the previous query and its result subgraph with the smallest difference as the input of Algorithm 5.

7.3 Theoretical Analysis

Boundedness of Algorithm 3. The Algorithm 3 is bounded to the polynomial function of $|\text{CHANGED}|$, which is shown in Theorem 7.1. Our proposed index \mathbb{I}_D can handle batch edges deletion boundedly by Algorithm 3, meanwhile, it can handle batch edges insertion relative-boundedly, which is shown in Theorem 7.2.

THEOREM 7.1. *For Algorithm 3, the time complexity of is $O(\Delta k_f \cdot T_{\Delta}(\text{CHANGED}) \cdot \max_{e \in \text{CHANGED}} T_{\Delta}(e))$, where $T_{\Delta}(e)$ denotes the time of listing all triangles that contain e , and $O(m')$ for space.*

Algorithm 4: Batch Edges Insertion Algorithm

Input: G, E^+, \mathbb{I}_D
Output: \mathbb{I}'_D

```

1  $G' \leftarrow G \oplus E^+; E_{up} \leftarrow E^+; Queue \leftarrow \emptyset; k_f \leftarrow 0;$ 
    $Nseg_{(k_c, k_f)} \leftarrow E^+;$ 
2 while  $!E_{up}.empty()$  do
3    $k_c \leftarrow 0;$ 
4   while  $!Nseg_{(k_c, k_f)}.empty()$  do
5      $e^* \leftarrow \leq_{k_f}.top(); C_{(k_c, k_f)} \leftarrow Nseg_{(k_c, k_f)}$  with  $s(e)$ 
       correctly initialized for  $\forall e \in C_{(k_c, k_f)}$ ;
6     while  $\exists e \in C_{(k_c, k_f)}$  with  $s(e) \leq k_c$  do
7        $rem_{\leq}(e) \leftarrow s(e);$ 
8       for each  $\Delta$  that contains  $e$  do
9         Let  $e'$  and  $e''$  be other two edges of  $\Delta$ ;
10        if  $\Delta \notin G(C_{(k_c, k_f)} \cup E_{\leq e^*})$  then continue;
11        if  $e' \in C_{(k_c, k_f)}$  then  $s(e') \leftarrow s(e') - 1;$ 
12        if  $e'' \in C_{(k_c, k_f)}$  then  $s(e'') \leftarrow s(e'') - 1;$ 
13         $C_{(k_c, k_f)} \leftarrow C_{(k_c, k_f)} \setminus e; Olns(\leq_{k_f}, e, e^*);$ 
14    for each  $e \in C_{(k_c, k_f)}$  do
15      Update  $ext(\cdot)$  for edges after  $e^*$  inclusively in  $\leq_{k_f}$ ;
16    while  $e^* \neq nil$  do
17       $e^*_{next} \leftarrow$  the edge next to  $e^*$  in  $\leq_{k_f}$ ;
18      if  $ext(e^*) = 0$  then
19        Find the first  $e$  in  $\leq_{k_f}$  s.t.  $ext(e) > 0$ , else break;
20      else if  $ext(e^*) + rem_{\leq}(e^*) > k_c$  then
21        Add  $e^*$  to  $C_{(k_c, k_f)}$  with  $s(\cdot) = ext(\cdot) + rem_{\leq}(\cdot);$ 
22        Update  $ext(\cdot)$  for edges in  $E_{\leq e^*}; ODel(\leq, e^*);$ 
23      else
24        Update  $rem_{\leq}(e^*), ext(e^*)$  and  $s(\cdot)$  for  $C_{(k_c, k_f)}$ ;
25        Process all edges that are affected in  $C_{(k_c, k_f)}$ ;
26       $e^* \leftarrow e^*_{next};$ 
27     $Nseg_{(k_c+1, k_f)} \leftarrow C_{(k_c, k_f)}$ ;
28    for  $e \in Nseg_{(k_c, k_f)} \setminus Nseg_{(k_c+1, k_f)}$  do  $k'_c \leftarrow k_c;$ 
29     $k_c \leftarrow k_c + 1;$ 
30    Conduct maintenance from  $k_{cmax}$  similar to lines 4-29;
31    for each  $e \in \leq_{k_f}$  do
32       $k_c(e, k_f + 1) \leftarrow \min\{k_c(e, k_f + 1), k'_c(e, k_f)\};$ 
33     $E_{up} \leftarrow E(H'_{k_c+1, 0}) \setminus E(H_{k_c+1, 0}); Queue \leftarrow \emptyset; k_f \leftarrow k_f + 1;$ 
34 return  $\mathbb{I}'_D;$ 

```

Formulating AFF_{k_f} . Given the batch edges update ΔG , we consider two valid CD orders \leq_{k_f} and $\leq_{k'_f}^1$. We define the difference between them as $\text{dif}(\leq_{k_f}, \leq_{k'_f}^1) = \text{CHANGED} \cup \{e_1 \in G \mid \exists e_2 \in G, \text{s.t. } e_1 \leq_{k_f} e_2 \text{ and } e_2 \not\leq_{k'_f}^1 e_1\}$, where $e_1 \leq_{k_f} e_2$ means that e_1 appears before e_2 in the CD order \leq_{k_f} , also implies that the cycle truss number of e_1 is no bigger than e_2 's. Then the AFF_{k_f} of the CD order \leq_{k_f} is defined as, $AFF_{k_f} = \bigcap_{\leq_{k'_f}^1} \text{dif}(\leq_{k_f}, \leq_{k'_f}^1)$.

Relative boundedness of Algorithm 4. The MDM problem is unbounded for edge insertions, which is proved in Section 4. By formulating AFF_{k_f} and proposing the index \mathbb{I}_D , we show that the Algorithm 4 is relative bounded below.

THEOREM 7.2. *The time complexity of Algorithm 4 is $O(\Delta k_f \cdot \|AFF_{k_f}\| \|AFF_{k'_f}\|_1^2 \cdot \log(\|AFF_{k_f}\| \|AFF_{k'_f}\|_1^2))$ and $O(m')$ for space.*

THEOREM 7.3. *The time complexity of Algorithm 5 is $O(k_{fmax} \cdot m)$, the space complexity of Algorithm 5 is $O(m')$.*

Algorithm 5: MDSM Query Algorithm

Input: $G, Q, k_c, k_f, \mathcal{H}_{k_c, k_f, Q}(G), \mathbb{I}_D$ and $\Delta G, \Delta Q, k'_c, k'_f$
Output: $\mathcal{H}_{k'_c, k'_f, Q'}(G \oplus \Delta G)$

```

1  $H \leftarrow \mathcal{H}_{k_c, k_f, Q}$ ; Obtain  $\mathbb{I}'_D$  by Algorithm 3 and Algorithm 4;
2 if  $H$  is empty then
3    $H \leftarrow$  the result by the query algorithm [28] with
        $k'_c, k'_f, Q \oplus \Delta Q$  input;
4 else
5   for each  $e \in E(H)$  do
6     for each  $(k_{c_i}, k_{f_i}) \in \mathcal{ST}'(e)$  do
7       if  $(k_{c_i}, k_{f_i}) \leq (k'_c, k'_f)$  then Mark  $e$  as kept; break;
8     if  $e$  is NOT kept then  $H \leftarrow H \setminus e;$ 
9   if  $H$  is NOT connected then  $H \leftarrow \emptyset$ ; return  $H;$ 
10   $H \leftarrow$  the result by the query algorithm [28] that starts with
        $k'_c, k'_f, H;$ 
11  for each  $u \in V(H)$  do Mark  $u$  as visited;
12  for each  $q' \in \Delta Q$  do
13    if  $q'$  is NOT visited then  $H \leftarrow \emptyset$ ; return  $H;$ 
14 return  $H;$ 

```

8 EXPERIMENTS

In this section, we conduct all experiments on a Linux machine with an Intel Xeon Gold 6240R CPU and 1007GB main memory. Due to page limitations, we only show experimental results on selected datasets, complete results of all datasets are available here [34], similar performance trends are observed from other datasets.

Datasets. We use 7 real-world datasets which are publicly accessible, including Email (EM), Twitter (TW), BerkStan (BS), Wiki (WK), Pokec (PK) from SNAP [26], Edinburgh Associative Thesaurus (EAT) from Pajek [6] and DBpedia Links (DL) from KONECT [24]. The summary of datasets is shown in Table 2. Note that DL cannot be decomposed within 6×10^4 seconds, thus we sample subgraphs of it at different rates.

Algorithms. We evaluate the following algorithms in this paper. For the MDS problem: (1) the D-truss decomposition algorithm Dec (the decomposition algorithm [28] with $G \oplus \Delta G$ input). For the MDM problem: (1) the single deletion algorithm SDel (Algorithm 1), (2) the single insertion algorithm SIns (Algorithm 2), (3) the batch deletion algorithm with unit edge processing UBDel and with batch edge processing BDel (Algorithm 3) and (4) the batch insertion algorithm with unit edge processing UBIns and with batch edge processing BIns (Algorithm 4). For the MDSM problem: (1) ReQry (the query algorithm [28] with $G \oplus \Delta G$ input), (2) CoQry (the query algorithm [28] after being maintained by Algorithm 3 and Algorithm 4), (3) MtQry (Algorithm 5) and (4) OpQry (Algorithm 5 with the optimization). All algorithms are implemented in C++. In all experiments, the y -axis denotes the running time in logarithmic scale, and we end algorithms which run longer than 2×10^4 seconds.

8.1 Effectiveness Evaluation

EAT [6] is a word network which is built from the Edinburgh Associative Thesaurus (EAT). In the EAT network, the vertices denote the words, and the edge $\langle w_i, w_j \rangle$ from w_i to w_j means that the word w_i points to another word w_j if w_j appears as the response when w_i is received as the stimulus. For clearly illustrating the use case, we

Table 2: Summary of Datasets

Dataset	$ V $	$ E $	$\text{deg}_{\mathcal{B}max}^{\text{in}}$	$\text{deg}_{\mathcal{B}max}^{\text{out}}$	k_{cmax}	k_{fmax}
EM	1.0K	25.6K	211	333	14	21
EAT	23.1K	685K	1073	78	3	8
TW	81.3K	1.8M	3,383	1,205	161	199
BS	685K	7.6M	84,208	249	41	80
WK	1.8M	28.5M	238,040	3,907	36	37
PK	1.6M	30.6M	13,733	8,763	18	27
DL	18.2M	136.5M	612,308	8,105	-	-

sample the subgraph such that in Figure 6(a). Consider the following two queries for the word “drink”, one is $Query1 : \{Q = \text{“drink”}, k_c = 3, k_f = 0\}$, the other is $Query2 : \{Q = \text{“drink”}, k_c = 2, k_f = 4\}$. Here $Query1$ aims for finding the word community around “drink”, where every word can remind other 4 words with similar influences; while $Query2$ weakens the requirement of similar influences, but introduces the hierarchical reminding relationship that each word is participated in with other 5 words. The updated edges E^+ are marked red in Figure 6(a). Figure 6(b) and (c) show the query results before and after applying updated edges.

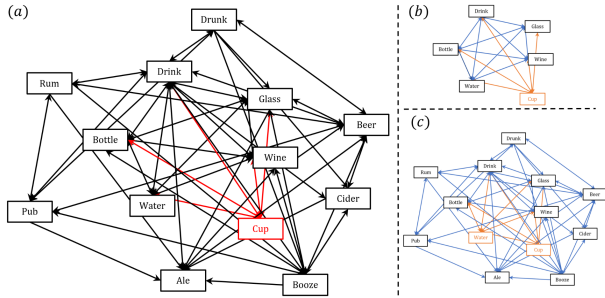


Figure 6: Effectiveness evaluation

$Query1$ desires a (3, 0)-truss of the word “drink”, before applying the updated edges, the result subgraph consists of “glass”, “bottle”, “wine” and “water”. After maintaining by Algorithm 4, the new (3, 0)-truss of “drink” also contains “cup”, which shares an equal relationship with all words in old result. Considering $Query2$, a (2, 4)-truss of the word “drink” is returned, many words are contained in the old result subgraph, form a mixed structure around “drink”. They are participated with the community of “drink” both in a relatively weak equal relationship and a relatively strong hierarchical relationship. Together with the result of $Query1$, we know that the words “glass”, “bottle”, “wine” forms a stronger equal relationship with “drink”. The word “drunk” is in the upstream of “drink”, which means that it always reminds people of “drink”, while the words “booze”, “cider”, “beer”, “rum”, “ale” are in the downstream of “drink”, which means that when people are given “drink” as the stimulus, these alcoholic words arise as a response. After being maintained, the new (2, 4)-truss of “drink” also contains “cup” and “water”, which enlarge the community with a relatively strong equal relationship.

8.2 Efficiency Evaluation

Exp-1: Index construction. We show the index construction time and index size in Figure 7(a). For all the datasets, the D-Index can be built within 2×10^4 seconds, and the index size is within 85 GB. Such construction costs are acceptable for real world applications.

Exp-2: Varying $|\Delta G|$. We conduct the experiment of investigating the impact of $|\Delta G|$ over EM and TW. Given a directed graph G , ΔG is sampled from $E(G)$ randomly, in this way, we set G as the original graph for edge deletions and set $G \setminus \Delta G$ as the original graph for edge insertions. We vary $|\Delta G|$ from 4% to 20% of the original graph size. The running times on each graph are shown in Figure 7 (b), (c), (d), (e) respectively. The results show that, (1) BIns performs better than SIns and UBIns consistently, and BIns runs faster than Dec with ΔG of around 12% for EM, and of around 10% for TW; (2) BDel also performs better than SDel and UBDel consistently, and BDel is more efficient than Dec with ΔG of around 20% for both datasets; (3) SIns and SDel perform better than UBIns and UBDel coordinately, which implies that single-update algorithms are more efficient than the D-Index because of index-related processing are costly than index-free operators under the single-update setting; (4) $|\Delta G|$ is usually very small relative to $|G|$ in real world [15], BIns runs at most 4.12x and 1.63x faster than Dec for EM and TW respectively, and BDel runs at most 6.89x and 2.94x faster than Dec for two datasets. We observe that BDel is more efficient than BIns, because BDel is bounded to CHANGED, while BIns is unbounded.

Exp-3: Varying $|G|$. The impact of $|G|$ is evaluated on 2 largest datasets PK and DL with 30.6M edges and 136.5M edges respectively. We also obtain 5 subgraphs for PK by randomly sampling from the original graph at rates of 20%, 40%, 60%, 80% and 100%. Note that $G_{20\%} \subseteq G_{40\%} \subseteq G_{60\%} \subseteq G_{80\%} \subseteq G_{100\%}$ is guaranteed. The results of PK are shown in Figure 7 (f) and (h). We sample DL similarly at rates of 20%, 25%, 30%, 35% and 40%, and results are shown in Figure 7 (g) and (i). In this experiment, we apply the same ΔG for these sampled subgraphs of each dataset, where $|\Delta G|$ is 1% of $|G|$. The usage of $|\Delta G|$ for insertions and deletions is exactly as same as Exp-2. We see that Dec runs super-linearly with the size of the graph for both PK and DL. Take Dec as the baseline method, (1) for edge insertions, BIns and SIns scale well and perform better than Dec as $|G|$ getting larger, but UBIns is not always scalable, especially for very large datasets such as DL, since too many low-level operations are executed in UBIns for each unit update and the time cost of these operations gets longer when $|G|$ gets larger; (2) for edge deletions, all algorithms perform well and BDel performs best in the experiment since all of them are bounded. Similar performance trends are observed from other datasets.

Exp-4: Vary $|\text{previous queries}|$. In this experiment, we generate 100 queries randomly, the queried trussness is guaranteed to be dominated by (k_{cmax}, k_{fmax}) . Note that generated queries are not necessarily valid, since the solution should also be efficient to judge invalid queries. ΔG is composed of a half of edge deletions and the other half of edge insertions. We set $|Q| = 4, \Delta Q$ as null. We vary the number of previous queries of 19, 39, 59, 79, 99. For ReQry, CoQry and MtQry, we report the average query time by each point; for CoQry, we report the exact query time at each point. In Figure 7 (j), we show the results on WK. When the number of previous queries gets larger, the query time changes slightly for ReQry, CoQry and

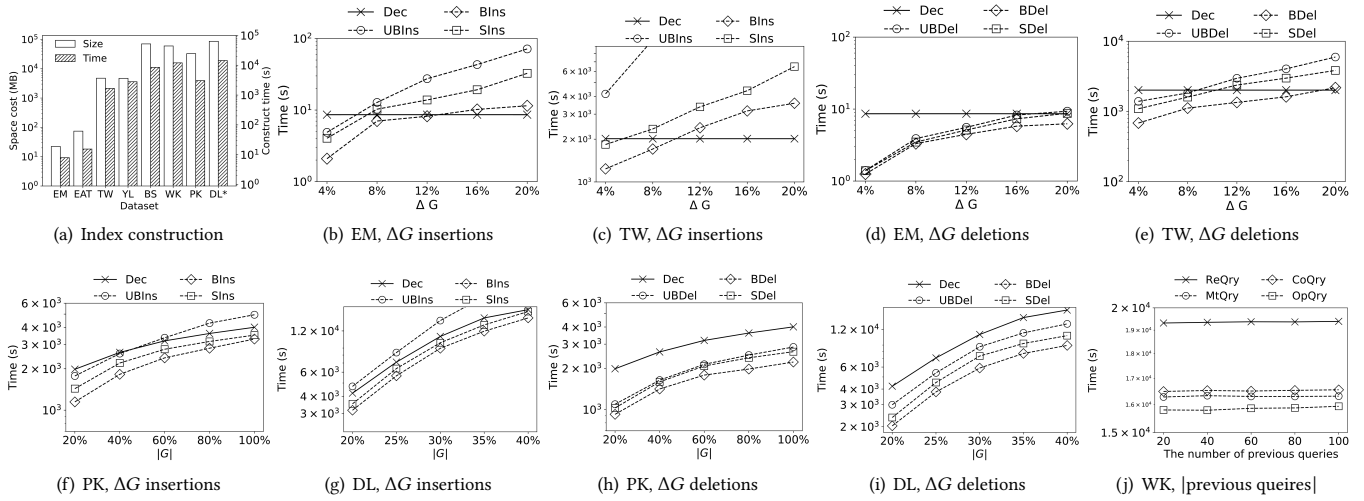


Figure 7: Exp-1, 2, 3 and 4 of Efficiency Evaluation

MtQry. OpQry is faster than other methods on almost all points. The time cost of OpQry decreases first but keeps increasing as the number of previous queries gets larger. Because with more previous results, the checking procedures of them take longer time.

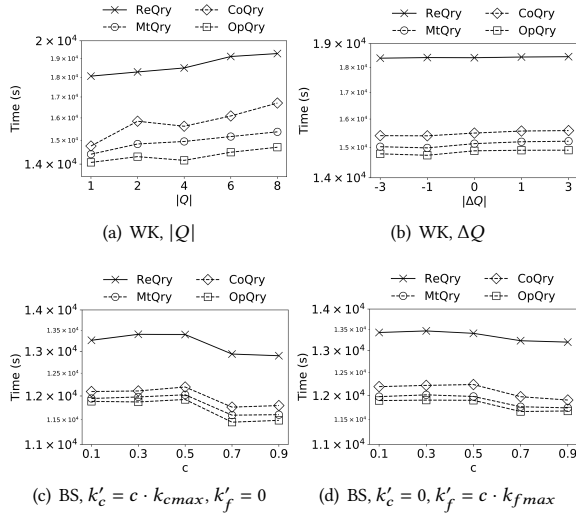


Figure 8: Exp-5 and 6 of Efficiency Evaluation

Exp-5: Varying $|Q|$ and $|\Delta Q|$. In Figure 8 (a), we evaluate the impact of $|Q|$ on WK. We set ΔQ as null and vary $|Q|$ of 1, 2, 4, 6 and 8. The query time becomes longer for all methods when $|Q|$ gets larger, because it costs longer for checking more query vertices. ReQry and CoQry cost more time than MtQry and OpQry when $|Q|$ gets larger, since the dominating cost of ReQry and CoQry is from the static decomposition from scratch. MtQry and OpQry show better scalability. In Figure 8 (b), we evaluate the impact of $|\Delta Q|$ on WK, where we set $|Q| = 4$. $|\Delta Q|$ is varied from -3 to 3 . The

impact of $|\Delta Q|$ is similar to that of $|Q|$, because when $|\Delta Q|$ gets larger, it needs to check more query vertices. In all tested cases, CoQry, MtQry and OpQry are much more efficient than ReQry. OpQry performs best among all competitors, since OpQry utilizes more previous results instead of only the latest one.

Exp-6: Varying (k'_c, k'_f) . In this experiment, Q and ΔG remain same as before varying (k'_c, k'_f) . We investigate the impact of k'_c and k'_f respectively in Figure 8 (c) and (d). Either k'_c or k'_f gets larger, all methods run faster because the size of the result gets smaller, OpQry is the most efficient one among all competitors with the same reason above.

9 CONCLUSIONS

In this paper, we propose a collection of incremental solutions for maximal D-truss search in dynamic graphs, including single-update algorithms and an order-based index that can handle batch updates efficiently. We conduct theoretical boundedness analysis of the maximal D-truss given edge insertions and deletions respectively. We further present a fully-dynamic query algorithm to help accelerate the query processing. Extensive experimental results on real-world graphs validate the effectiveness and efficiency of our solutions.

ACKNOWLEDGMENTS

Lei Chen is partially supported by National Science Foundation of China (NSFC) under Grant No. U22B2060, the Hong Kong RGC GRF Project 16209519, CRF Project C6030-18G, C2004-21GF, AOE Project AoE/E-603/18, RIF Project R6020-19, Theme-based project TRS T41-603/20R, China NSFC No. 61729201, Guangdong Basic and Applied Basic Research Foundation 2019B151530001, Hong Kong ITC ITF grants MHX/078/21 and PRP/004/22FX, Microsoft Research Asia Collaborative Research Grant, HKUST-Webank joint research lab grant and HKUST Global Strategic Partnership Fund (2021 SJTU-HKUST). Yue Wang is partially supported by China NSFC(No.62002235). Yue Wang is the corresponding author.

REFERENCES

- [1] Alessandro Acquisti and Ralph Gross. 2006. Imagined communities: Awareness, information sharing, and privacy on the Facebook. *PETS* (2006), 36–58.
- [2] Bowen Alpern, Roger Hoover, Barry K Rosen, Peter F Sweeney, and F Kenneth Zadeck. 1990. Incremental evaluation of computational circuits. *SODA* (1990), 32–42.
- [3] Sihem Amer-Yahia, Senjuti Basu Roy, Ashish Chawla, Gautam Das, and Cong Yu. 2009. Group Recommendation: Semantics and Efficiency. *PVLDB* 2, 1 (2009), 754–765.
- [4] Reid Andersen, Fan Chung, and Kevin Lang. 2006. Local graph partitioning using pagerank vectors. In *FOCS*. 475–486.
- [5] Jørgen Bang-Jensen and Gregory Z Gutin. 2008. *Digraphs: theory, algorithms and applications*. SSBM.
- [6] Vladimir Batagelj and Andrej Mrvar. 2006. Pajek datasets. <http://vlado.fmf.uni-lj.si/pub/networks/data/>.
- [7] Vladimir Batagelj and Matjaz Zaversnik. 2003. An $O(m)$ Algorithm for Cores Decomposition of Networks. *CoRR* cs.DS/0310049 (2003).
- [8] James Cheng, Yiping Ke, Shumo Chu, and M Tamer Özsu. 2011. Efficient core decomposition in massive networks. *ICDE* (2011), 51–62.
- [9] Jonathan Cohen. 2008. Trusses: Cohesive subgraphs for social network analysis. *National security agency technical report* 16, 3.1 (2008).
- [10] Laxman Dhulipala, Quanquan C Liu, Julian Shun, and Shangdi Yu. 2021. Parallel batch-dynamic k -clique counting. *APOCS* (2021), 129–143.
- [11] Paul Dietz and Daniel Sleator. 1987. Two algorithms for maintaining order in a list. In *STOC*. 365–372.
- [12] Danhao Ding, Hui Li, Zhipeng Huang, and Nikos Mamoulis. 2017. Efficient fault-tolerant group recommendation using alpha-beta-core. *CIKM* (2017), 2047–2050.
- [13] Dongsheng Duan, Yuhua Li, Ruixuan Li, and Zhengding Lu. 2012. Incremental K -clique clustering in dynamic social networks. *Artif. Intell. Rev.* 38, 2 (2012), 129–147.
- [14] Wenfei Fan and Chao Tian. 2022. Incremental graph computations: Doable and undoable. *TODS* 47, 2 (2022), 1–44.
- [15] Wenfei Fan, Chao Tian, Ruiqi Xu, Qiang Yin, Wenyan Yu, and Jingren Zhou. 2021. Incrementalizing graph algorithms. *SIGMOD* (2021), 459–471.
- [16] Yixiang Fang, Reynold Cheng, Xiaodong Li, Siqiang Luo, and Jiafeng Hu. 2017. Effective community search over large spatial graphs. *PVLDB* 10, 6 (2017), 709–720.
- [17] Yixiang Fang, Zhongran Wang, Reynold Cheng, Hongzhi Wang, and Jiafeng Hu. 2018. Effective and efficient community search over large directed graphs. *TKDE* 31, 11 (2018), 2093–2107.
- [18] Yixiang Fang, Haijun Zhang, Yunming Ye, and Xutao Li. 2014. Detecting hot topics from Twitter: A multiview approach. *J. Inf. Sci.* 40, 5 (2014), 578–593.
- [19] Christos Giatsidis, Dimitrios M Thilikos, and Michalis Vazirgiannis. 2013. D-cores: measuring collaboration of directed graphs based on degeneracy. *Knowl. Inf. Syst.* 35, 2 (2013), 311–343.
- [20] Roger Guimera and Luis A Nunes Amaral. 2005. Functional cartography of complex metabolic networks. *Nature* 433, 7028 (2005), 895–900.
- [21] Xin Huang, Hong Cheng, Lu Qin, Wentao Tian, and Jeffrey Xu Yu. 2014. Querying k -truss community in large and dynamic graphs. *SIGMOD* (2014), 1311–1322.
- [22] Xin Huang, Laks VS Lakshmanan, and Jianliang Xu. 2017. Community search over big graphs: Models, algorithms, and opportunities. *ICDE* (2017), 1451–1454.
- [23] Ann E Krause, Kenneth A Frank, Doran M Mason, Robert E Ulanowicz, and William W Taylor. 2003. Compartments revealed in food-web structure. *Nature* 426, 6964 (2003), 282–285.
- [24] Jérôme Kunegis. 2013. Konec: the koblenz network collection. *WWW* (2013), 1343–1350.
- [25] Kevin J Lang and Reid Andersen. 2007. Finding dense and isolated submarkets in a sponsored search spending graph. *CIKM* (2007), 613–622.
- [26] Jure Leskovec and Andrej Krevl. 2014. SNAP Datasets: Stanford Large Network Dataset Collection. <http://snap.stanford.edu/data>.
- [27] Zhe Lin, Fan Zhang, Xuemin Lin, Wenjie Zhang, and Zhihong Tian. 2021. Hierarchical core maintenance on large dynamic graphs. *PVLDB* 14, 5 (2021), 757–770.
- [28] Qing Liu, Minjun Zhao, Xin Huang, Jianliang Xu, and Yunjun Gao. 2020. Truss-based community search over large directed graphs. *SIGMOD* (2020), 2183–2197.
- [29] Feng Luo, James Z Wang, and Eric Promislow. 2008. Exploring local community structures in large networks. *WIAS* 6, 4 (2008), 387–400.
- [30] J-P Onnela, Anirban Chakraborti, Kimmo Kaski, Janos Kertesz, and Antti Kanto. 2003. Dynamics of market correlations: Taxonomy and portfolio analysis. *Phys. Rev. E* 68, 5 (2003), 056110.
- [31] 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.
- [32] Ganesan Ramalingam and Thomas Reps. 1996. On the computational complexity of dynamic graph problems. *Theor. Comput. Sci.* 158, 1-2 (1996), 233–277.
- [33] Erzsébet Ravasz, Anna Lisa Somera, Dale A Mongru, Zoltán N Oltvai, and A-L Barabási. 2002. Hierarchical organization of modularity in metabolic networks. *Science* 297, 5586 (2002), 1551–1555.
- [34] Repository. [n.d.]. <https://github.com/ExpCodeBase/DTrussMain/blob/main/full.pdf>.
- [35] Ahmet Erdem Saryüce, Buğra Gedik, Gabriela Jacques-Silva, Kun-Lung Wu, and Ümit V Çatalyürek. 2016. Incremental k -core decomposition: algorithms and evaluation. *VLDBJ* 25, 3 (2016), 425–447.
- [36] Binta Sun, T-H Hubert Chan, and Mauro Sozio. 2020. Fully dynamic approximate k -core decomposition in hypergraphs. *TKDD* 14, 4 (2020), 1–21.
- [37] Taro Takaguchi and Yuichi Yoshida. 2016. Cycle and flow trusses in directed networks. *R. Soc. Open Sci.* 3, 11 (2016), 160270.
- [38] Tim Teitelbaum and Thomas Reps. 1981. The Cornell program synthesizer: a syntax-directed programming environment. *Commun. ACM* 24, 9 (1981), 563–573.
- [39] Kai Wang, Wenjie Zhang, Xuemin Lin, Ying Zhang, Lu Qin, and Yuting Zhang. 2021. Efficient and effective community search on large-scale bipartite graphs. *ICDE* (2021), 85–96.
- [40] Tianbao Yang, Yun Chi, Shenghuo Zhu, Yihong Gong, and Rong Jin. 2010. Directed network community detection: A popularity and productivity link model. *ICDM* (2010), 742–753.
- [41] Long Yuan, Lu Qin, Xuemin Lin, Lijun Chang, and Wenjie Zhang. 2016. Diversified top- k clique search. *VLDBJ* 25, 2 (2016), 171–196.
- [42] Yikai Zhang and Jeffrey Xu Yu. 2019. Unboundedness and efficiency of truss maintenance in evolving graphs. *SIGMOD* (2019), 1024–1041.