



# Reliable Community Search in Dynamic Networks

Yifu Tang  
Deakin University  
Geelong, Australia  
tangyif@deakin.edu.au

Jianxin Li\*  
Deakin University  
Geelong, Australia  
jianxin.li@deakin.edu.au

Nur Al Hasan Haldar  
The University of Western Australia  
Perth, Australia  
nur.haldar@uwa.edu.au

Ziyu Guan  
Xidian University  
Xi'an, China  
zyguan@xidian.edu.cn

Jiajie Xu  
Soochow University  
Suzhou, China  
xujj@suda.edu.cn

Chengfei Liu  
Swinburne University of Technology  
Melbourne, Australia  
cliu@swin.edu.au

## ABSTRACT

Searching for local communities is an important research problem that supports advanced data analysis in various complex networks, such as social networks, collaboration networks, cellular networks, etc. The evolution of such networks over time has motivated several recent studies to identify local communities in dynamic networks. However, these studies only utilize the aggregation of disjoint structural information to measure the quality and ignore the reliability of the communities in a continuous time interval. To fill this research gap, we propose a novel  $(\theta, k)$ -core reliable community (CRC) model in the weighted dynamic networks, and define the problem of *most reliable community search* that couples the desirable properties of connection strength, cohesive structure continuity, and the maximal member engagement. To solve this problem, we first develop a novel edge filtering based online CRC search algorithm that can effectively filter out the trivial edge information from the networks while searching for a *reliable* community. Further, we propose an index structure, Weighted Core Forest-Index (WCF-index), and devise an index-based dynamic programming CRC search algorithm, that can prune a large number of insignificant intermediate results and support efficient query processing. Finally, we conduct extensive experiments systematically to demonstrate the efficiency and effectiveness of our proposed algorithms on eight real datasets under various experimental settings.

## PVLDB Reference Format:

Yifu Tang, Jianxin Li, Nur Al Hasan Haldar, Ziyu Guan, Jiajie Xu, and Chengfei Liu. Reliable Community Search in Dynamic Networks. PVLDB, 15(11): 2826 - 2838, 2022.  
doi:10.14778/3551793.3551834

## PVLDB Artifact Availability:

The source code, data, and/or other artifacts have been made available at <https://github.com/Cyril-Tang/CRC-query>.

\*Corresponding author

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](mailto:info@vldb.org). Copyright is held by the owner/author(s). Publication rights licensed to the VLDB Endowment.  
Proceedings of the VLDB Endowment, Vol. 15, No. 11 ISSN 2150-8097.  
doi:10.14778/3551793.3551834

## 1 INTRODUCTION

Local community search has attracted much attention in recent years and has shown its great success in different applications, e.g., personalized recommendation [19, 20], destination marketing [34]. In general, local community search aims to identify a densely connected structure with regard to a query vertex. Majority of the existing works [9, 29, 30] on local community search consider static network structure. For instance, Clauset et al. [9] and Luo et al. [30] proposed local modularity to measure the quality of the community in the static network. A random walk based community detection model from multiple static networks is proposed in [29]. Some existing works [7, 12] take into consideration that network structure may change over time and propose local community search in dynamic networks. For example, Bu et al. [7] provides a modularity-based criterion to find local communities in a dynamic network and update them in an incremental manner by monitoring the changes. In another work, DiTursi et al. [12] discovered the dynamic communities with optimal time intervals by minimizing *temporal conductance*, a well-known metric to measure the quality of a community. However, the above-mentioned works assess the community quality using their aggregated structural cohesiveness at independent timestamps and ignore the evolving structure of a community over time. In the dynamic network, the continuity of the community cohesiveness is an important factor in determining whether a community is reliable. For example, it is desirable to hire a team that continuously delivers high-quality outputs together over time in the collaboration network. Finding the user groups with a longer duration of reacting to the social event can help better analyze user behavior on social media. Existing works largely ignore the continuity of the community cohesiveness. In addition, they did not consider the edge weight, e.g., connection strength between a node pair, which incurs the new computational challenge to solve local community search in dynamic networks.

To fill this research gap, we propose a novel community model of  $(\theta, k)$ -core reliable community in dynamic networks where: (i) the community is a  $k$ -core with each edge weight no less than the weight threshold  $\theta$ , and (ii) spans over a period of time. The most reliable local community search aims to find the community with the maximum reliability score, which is defined by coupling temporal continuity and member engagement. In other words, this work jointly models the three important properties, i.e., connection strength, cohesiveness continuity, and member engagement, of a community in a dynamic network.

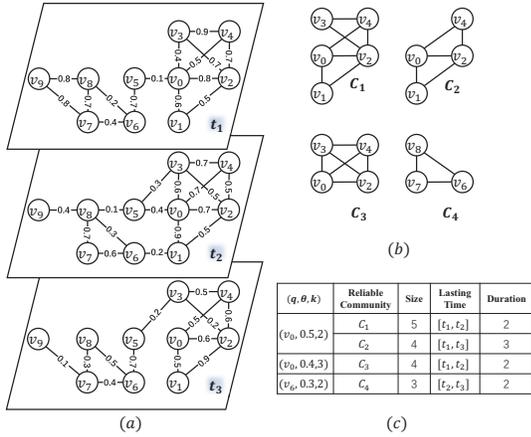


Figure 1: Interactions in a small graph of three timestamps.

**Example 1.1.** Figure 1 shows an illustrative example of a dynamic online social network of three timestamps  $t_1, t_2, t_3$ , where the weight of an edge represents the connection strength. Assume that a cosmetics company wants to post ads for a new product that can attract customers from local communities. The effective way is to search for a community from an existing customer, where the members might have continuous interaction or communication with each other with a minimum connection strength. Our proposed  $(\theta, k)$ -core reliable community search model can support such an application. Suppose,  $v_0$  is the target user and  $k=3, \theta=0.4$  are the query inputs. We can identify the community  $C_3=\{v_0, v_2, v_3, v_4\}$  as the optimal result, which has four members and spans for two continuous timestamps  $t_1$  and  $t_2$ . Similarly, it returns two communities  $C_1$  and  $C_2$  when  $k=2$  and  $\theta=0.5$ . Among these two communities,  $C_2$  can be more acceptable than  $C_1$  for the cosmetic company. This is because the members of  $C_2$  kept their close interaction for 3 continuous timestamps and the size of  $C_2$  is considerably similar to  $C_1$ . This enables the company to identify ads placeholder easily and cost-effectively that can maximize the impact of their ads.

Although there exist some similar works [22, 24, 25, 35] that identify meaningful communities over time, our proposed problem is more acceptable due to more sophisticated research challenges. In [22], Li et al. defined the persistent community search as the maximal  $k$ -core where each vertex's accumulated degree meets the  $k$ -core requirement within a time interval. They designed a novel temporal graph reduction algorithm and searched the maximum persistent community utilizing pruning and bounding techniques, which takes exponential complexity. Expanding from [22], Li et al. [24] studied a single query vertex based persistent community search by developing an enumeration-based subgraph search algorithm. In [35], Qin et al. proposed the stable communities by first selecting the centroid vertices where each centroid vertex has a certain number of neighbors with the desired similarity, and the star-shape of the centroid vertex and its neighbors exists frequently in a period of time; and then clustering the network vertices into stable groups based on the selected centroids. Similar to [35], Lin et al. [25] defined frequency-based dense subgraphs that satisfy the quasi-clique structure with at least  $\theta$  vertices and the degree of each

vertex exceeds a given threshold. The proposed searching algorithm takes exponential complexity. However, the persistent community in [22, 24] did not consider the weight of the edge and its time complexity is too high for dealing with large-scale networks. The frequency-based subgraph in [25, 35] ignored the continuity of the cohesive structure and failed to maximize member engagement.

To solve the proposed problem of the most reliable local community search, a naive idea is to enumerate all the possible community candidates and select the satisfied results by checking their edge weight and duration. However, this may incur an exponential time cost. To address the computational challenge, in this paper, we firstly propose an efficient eligible edge filtering online search algorithm that utilizes the minimum edge requirement of  $k$ -core to compute the reliability upper bound to prune a large number of edge sets without probing their corresponding  $(\theta, k)$ -core community candidates. To further accelerate the query processing, we develop a weighted core forest index by maintaining the standard  $\theta$ -threshold values and the  $(\theta, k)$ -core structural information of vertices, which supports efficient retrieval of  $k$ -core with regard to different thresholds and timestamps. Following this, we design an index-based dynamic programming algorithm, and derive the reliability upper bound of communities w.r.t. the time interval during the dynamic programming procedure to avoid probing the unsatisfactory community candidates. Besides, the index construction, maintenance, and compression are well presented in this paper.

The main contributions of this work are as below:

- We propose a novel problem of the most reliable community search that jointly considers community continuity, community size, and connection strength for online network analysis services.
- We develop an efficient online search algorithm by deriving and applying the properties of pruning the ineligible edges w.r.t. the given query conditions.
- We further present a weighted core forest index and develop an index-based dynamic programming algorithm to solve the most reliable community search problem in a more efficient way.
- We conduct extensive experiments to show the efficiency and effectiveness of the proposed algorithms and community model by using eight real-world datasets and comparing with three existing studies.

The remainder of this paper is organized as follows. First, we formalize the most reliable local community search problem in Section 2 and develop the online search algorithm in Section 3. Then, we introduce our index structure and the detailed index-based search algorithm in Section 4. The procedures of index construction, maintenance and compression are shown in Section 5. Experimental evaluation and results are discussed in Section 6. Finally, we discuss the related work in Section 7 and conclude the work in Section 8.

## 2 PRELIMINARIES AND PROBLEM DEFINITION

In this section, we first present the preliminaries and then formalize the problem of the most reliable local community search.

**Definition 2.1** (Dynamic Networks). A dynamic network  $\mathcal{G} = \{G_{t_1}, \dots, G_{t_T}\}$  is a sequence of time-variant weighted graph instances

$\{G_{t_1}, \dots, G_T\}$  s.t.,  $t_1 < t_2 < \dots < T$ , where each timestamped instance  $G_t = (V_t, E_t, W_t)$  contains a set of vertices  $V_t$ , a set of edges  $E_t$  with the weights  $W_t(e) \in (0, 1]$  for  $\forall e \in E_t$ .

In this work, we ignore the isolated vertices, so that vertex updates can be supported by edge insertions and deletions. The edge insertion with a new endpoint can represent the vertex addition and edge deletion isolated an endpoint reflects the vertex deletion. For simplicity, we assume all graph instances share a fixed set of vertex  $V$ , i.e.  $G_t = (V, E_t, W_t)$ . The edge weight is a widely-used network feature to represent the interaction frequency, similarity, or connection strength between vertices. In this work, we normalize the edge weight to be in  $(0, 1]$ .

For a graph instance  $G_t = (V, E_t, W_t)$ ,  $\deg(u, G_t)$  denotes the degree of a vertex  $u$  in  $G_t$ , which is the number of neighbors of  $u$  in  $G_t$ . Like [37], we consider  $k$ -core in  $G_t$  as a connected subgraph  $G_t^k = (V^k, E_t^k)$  where each vertex has the degree no less than  $k$ , i.e.  $\forall u \in V_t^k, \deg(u, G_t^k) \geq k$ .

**Definition 2.2** ( $(\theta, k)$ -core). Given a graph instance  $G_t = (V, E_t, W_t)$ , an integer  $k$ , and a threshold  $\theta$ , a connected subgraph  $G_t^{\theta, k} = (V_t^{\theta, k}, E_t^{\theta, k}, W_t^{\theta, k})$  is called a  $(\theta, k)$ -core of  $G_t$  if  $G_t^{\theta, k}$  is a  $k$ -core and each edge has the weight no less than  $\theta$ , i.e.  $\forall u \in V_t^{\theta, k}, \deg(u, G_t^{\theta, k}) \geq k$  and  $\forall e \in E_t^{\theta, k}, W_t^{\theta, k}(e) \geq \theta$ .

**Definition 2.3** (Time Interval based  $(\theta, k)$ -core Reliable Community (CRC)). Given a dynamic network  $\mathcal{G} = \{G_{t_1}, \dots, G_T\}$ , an integer  $k$ , a threshold  $\theta$ , and a time interval  $T_C = [t_s, t_e]$ , a  $(\theta, k)$ -core reliable community is a subgraph  $C = (V_C, E_C)$  that spans continuously from  $t_s$  to  $t_e$  and for each timestamp  $t_n \in T_C$ , the subgraph induced by  $E_C$  from the graph instance  $G_{t_n}$  is a  $(\theta, k)$ -core, i.e.  $\forall t_n \in T_C, G_{t_n}[E_C]$  is a  $(\theta, k)$ -core of  $G_{t_n}$ . In the remainder of this work,  $(\theta, k)$ -core reliable community is called CRC for brevity.

Based on Definition 2.3, a CRC maintains a cohesive structure with the required connection strength of a time interval in the dynamic network. Its reliability score can be measured by coupling the continuity and size of the community as below:

**Definition 2.4.** (Reliability Score of CRC) Given a dynamic network  $\mathcal{G} = \{G_{t_1}, G_{t_2}, \dots, G_T\}$ , a query time interval  $T_Q = [t_i, t_j]$ , and a CRC  $C = (V_C, E_C)$  with regard to the time interval  $T_C = [t_s, t_e]$ , where  $T_C \subseteq T_Q$ , the reliability score  $S_{rel}(C)$  is defined as the harmonic average of its normalized duration and size:

$$S_{rel}(C) = (1 + \alpha^2) \cdot \frac{\mathcal{N}(V) \cdot \mathcal{N}(T)}{(\alpha^2 \cdot \mathcal{N}(V)) + \mathcal{N}(T)} \quad (1)$$

where  $\mathcal{N}(V) = |V_C|/|V_{max}^k|$  and  $\mathcal{N}(T) = |T_C|/|T_Q|$  represent the normalized size and duration length respectively.  $V_{max}^k$  denotes the  $k$ -core with the maximum size in  $G_t$  ( $t \in T_Q$ ).

From Equation 1, we can infer that reliability score is monotonically related to size and duration of the community. Parameter  $\alpha$  means that community duration is  $\alpha$  times as much important as community size, so that higher  $\alpha$  tends to find the community with longer duration and smaller size. The default  $\alpha = 1$  means both community size and duration are important equivalently.

Based on the above definitions and reliability measurement, we formalize the most reliable local community search problem as below.

**PROBLEM 1 (MOST RELIABLE LOCAL COMMUNITY SEARCH).** Given a dynamic network  $\mathcal{G} = \{G_{t_1}, G_{t_2}, \dots, G_T\}$ , a query vertex  $q$ , a threshold  $\theta$ , a structural constraint integer  $k$ , and a query time interval  $T_Q = [t_i, t_j]$ , the problem of the Most Reliable Local Community Search is to find the CRC  $C = (V_C, E_C)$  and its continuous time interval  $T_C = [t_s, t_e]$ , satisfying

$$\arg \max_{C \subseteq V} S_{rel}(C) \quad (2)$$

subject to  $q \in V_C; \forall v \in V_C, \deg(v, C) \geq k; \forall e \in E_C \wedge \forall t \in T_C, W_t(e) \geq \theta$ ; and  $T_C \subseteq T_Q$ .

As shown in Figure 1, when the query time interval is  $[t_1, t_3]$ , we can obtain two reliable communities  $C_1$  and  $C_2$  w.r.t. the query input  $(v_0, 0.5, 2)$ . The maximal 2-core is composed of 10 vertices. When  $\alpha = 1$ , we have  $S_{rel}(C_1) = S_{rel}(C_2) = 0.57$ . When  $\alpha$  increases to 2, we have  $S_{rel}(C_1) = 0.63 < S_{rel}(C_2) = 0.77$ , and  $C_2$  with longer duration becomes the optimal result.

To solve the most reliable community search problem, a naive solution is to compute all the  $(\theta, k)$ -core at each timestamp, and then verify their longest duration in the dynamic network. After that, their reliability scores can be obtained by multiplying their size and the number of continuous timestamps. Finally, the most reliable community can be returned by selecting the ones with the maximum reliability scores. However, the operation of finding all the  $(\theta, k)$ -core needs to probe all the combinations of edges that form a connected subgraph with no less than  $k(k+1)/2$  edges and  $k+1$  vertices, i.e., satisfying the conditions of minimal  $k$ -core component. Thus, we can remark that the computational cost of finding the most reliable community is in exponential complexity.

### 3 ONLINE RELIABLE COMMUNITY SEARCH

To efficiently solve the problem of reliable local community search, in this section, we will present a novel *Eligible Edge Filtering (EEF)* based Online CRC Search algorithm. Different from the naive idea discussed in Section 2, *EEF* does not need to generate all the  $(\theta, k)$ -core candidates, which can greatly reduce the query time cost.

Given a graph instance  $G_{t_n} = (V, E_{t_n})$  and a weight threshold  $\theta$ , we can filter out “ineligible” edges whose weights are less than  $\theta$  and maintain only the “eligible” edges. We identify the set of eligible edges, denoted by  $E_{t_n, \theta}$ , i.e.  $E_{t_n, \theta} = \{e \in E_{t_n} | W_{t_n}(e) \geq \theta\}$  for the CRC construction.

**Definition 3.1** (Eligible Lasting Time of Edge). Given a graph instance  $G_{t_n} = (V, E_{t_n})$  at timestamp  $t_n \in [t_i, t_j]$ , and a threshold  $\theta$ , for an edge  $e \in E_{t_n, \theta}$ , its eligible lasting time  $\lambda_{t_n, \theta}(e)$  is measured by the length of the longest time interval  $[t_m, t_n]$  ( $t_i \leq t_m \leq t_n$ ) if  $W_{t_x}(e) \geq \theta$  for  $\forall t_x \in [t_m, t_n]$ .

Eligible lasting time calculates the number of continuous timestamps that the edge is “eligible” until the current timestamp. For example, in Figure 1, for edge  $e = (v_0, v_1)$ , we have  $\lambda_{t_1, 0.6}(e) = 1$ ,  $\lambda_{t_2, 0.6}(e) = 2$  and  $\lambda_{t_3, 0.6}(e) = 0$ . We can easily derive that the eligible lasting time of edges can be incrementally computed by accessing the dynamic network chronologically.

Based on the eligible time, we can easily identify common edges of multiple continuous graph instances, that can be utilized to construct CRC with varying duration. Different from the vertex-induced subgraph, the subgraph induced by an eligible edge set

provides the guarantee to meet the requirement of edge weight  $\theta$ . Therefore, the eligible edge set can be used to prune the unqualified  $k$ -core candidates by using the following property.

**PROPERTY 3.1 (MINIMUM  $k$ -core).** *Given an edge set  $E_{t_n, \theta}$  at timestamp  $t_n$  with regards to a threshold  $\theta$ ,  $E_{t_n, \theta}$  can be pruned without probing its induced communities if  $|E_{t_n, \theta}| < k(k+1)/2$ , i.e., the number of edges does not meet the density requirement of  $k$ -core.*

In addition, we can calculate the upper bound of the reliability score of CRCs constructed using  $E_{t_n, \theta}$ . Given  $G' = (V', E')$  as a  $k$ -core in the induced subgraph  $G_{t_n}[E_{t_n, \theta}]$ ,  $|E_{t_n, \theta}| \geq |E'| \geq (k \cdot |V'|)/2$  must hold because there are at least  $k$  edges for a vertex in the  $k$ -core. Therefore, if  $G'$  can form a CRC with its duration as  $d$ , then its vertex size satisfies that  $|V'| \leq 2|E_{t_n, \theta}|/k$ . Thus, we have  $\mathcal{N}(V') \leq \frac{2|E_{t_n, \theta}|/k}{|V_{max}^k|}$  and  $\mathcal{N}(T) = d/|T_Q|$  and we can determine the upper bound of reliability score of the CRC constructed by the given eligible edge set.

For simplicity of presentation, we omit the normalizers  $|V_{max}^k|$  of the maximum community size and  $|T_Q|$  of the query time interval in the following equations.

**PROPERTY 3.2 ( $S_{rel}$  UPPER BOUND OF CRC w.r.t.  $E_{t_n, \theta}$ ).** *Given an eligible edge set  $E_{t_n, \theta}$ , an integer  $k$ , the upper bound reliability score  $UBR_{t_n}^d$  of the CRC constructed using  $E_{t_n, \theta}$  whose duration is  $d$  is calculated as:*

$$UBR_{t_n}^d = (1 + \alpha^2) \cdot \frac{2|E_{t_n, \theta}|/k \cdot d}{(\alpha^2 \cdot 2|E_{t_n, \theta}|/k) + d} \quad (3)$$

The key idea of *EEF*-based Online CRC Search is to filter out edges in each graph instance  $G_{t_n}$  using the given threshold  $\theta$  while maintaining the lasting time of each edge by a timestamp. As shown in Algorithm 1, we first initialize  $C_{opt}$  and  $maxS$  to store the most reliable community and its reliability score (line 1). Then, for each timestamp  $t_n$ , we traverse the edges of  $G_{t_n}$  starting from the query vertex  $q$  in *Breadth-First Search* manner. In the meantime, vertices and edges that violating the degree and weigh constraints are pruned. During the traversing, the eligible time of edges is updated incrementally, and the eligible edges are added to the edge set  $E_{t_n, \theta}$ . Having  $E_{t_n, \theta}$ , we calculate its upper bound  $UBR_{t_n}^1$  of potential CRC whose duration is 1 (lines 2-10). Then, we visit each timestamp  $t_n$  in the descending order of  $UBR_{t_n}^1$ , which provides a best-first search strategy to exploit the CRCs. We utilize  $E_{t_n, \theta}$  to construct CRC with duration  $d$  iterating from 1 to  $|[t_i, t_n]|$  (lines 11-19). At each iteration, we select the edge set  $E'$  where each edge has the eligible time no less than  $d$  and update the upper bound w.r.t.  $d$ . Then we adopt Property 3.1 and Property 3.2 to prune the CRC construction if  $|E'|$  is too small or  $UBR_{t_n}^d$  cannot exceed  $maxS$ . After that, we can extract the CRC  $C$  (i.e., local maximal  $k$ -core) from the induced subgraph  $G_{t_n}[E']$  by finding the connected component containing  $q$  after the core decomposition, and then update  $C_{opt}$  and  $maxS$ . Finally, the algorithm returns  $C_{opt}$  as the optimal result.

The time complexity of Algorithm 1 can be analyzed as below. For each graph instance  $G_{t_n} = (V, E_{t_n})$ , it takes  $O(|V| + |E_{t_n}|)$  to run the *Breadth-First Search* that requires to visit very vertex and edge once. At the same time, the eligible time of each edge is obtained (lines 2-18). Then, it needs  $O(|[t_i, t_n]| \cdot |E_{t_n}|)$  to compute  $|[t_i, t_n]|$  number of CRCs where each CRC is obtained by a core

---

### Algorithm 1: *EEF*-based Online CRC Search

---

**Input:** A dynamic weighted graph  $\mathcal{G} = \{G_{t_1}, G_{t_2}, \dots\}$ , an integer  $k$ , a query vertex  $q$ , a query time period  $T_Q = [t_i, t_j]$ , and a threshold  $\theta$ .  
**Output:** The Most Reliable Community  $C_{opt}$

- 1  $C_{opt} \leftarrow \emptyset; maxS \leftarrow 0;$
- 2 **for**  $t_n \leftarrow t_i$  to  $t_j$  **do**
- 3     **while** traverse edges  $e = (v, w)$  starting from  $q$  in *BFS* **do**
- 4         **if**  $deg(v, G_{t_n}) \leq k$  **then** Remove vertex  $v$  from  $G_{t_n};$
- 5         **if**  $W_{t_n}(e) \geq \theta$  **then**
- 6              $E_{t_n, \theta} \leftarrow E_{t_n, \theta} \cup \{e\};$
- 7             **if**  $e \in E_{t_{n-1}, \theta}$  **then**
- 8                  $\lambda_{t_n, \theta}(e) \leftarrow \lambda_{t_{n-1}, \theta}(e) + 1$
- 9         **else** Remove edge  $e$  from  $G_{t_n};$
- 10     Calculate  $UBR_{t_n}^1$  by Equation 3
- 11 **for**  $t_n \in T_Q$  in descending order of  $UBR_{t_n}^1$  **do**
- 12     **for**  $d \leftarrow 1$  to  $|[t_i, t_n]|$  **do**
- 13          $E' \leftarrow \{e \in E_{t_n, \theta} | \lambda_{t_n, \theta}(e) \geq d\};$
- 14         Calculate  $UBR_{t_n}^d$  by Equation 3;
- 15         **if**  $|E'| \geq k(k+1)/2$  &  $UBR_{t_n}^d > maxS$  **then**
- 16              $C \leftarrow$  local maximal  $k$ -core in  $G_{t_n}[E'];$
- 17             **if**  $S_{rel}(C) \geq S_{rel}(C_{opt})$  **then**
- 18                  $maxS \leftarrow S_{rel}(C);$
- 19                  $C_{opt} \leftarrow C;$
- 20 **Return**  $C_{opt}.$

---

decomposition process that needs to consume  $O(|E_{t_n}|)$  [4] (lines 20-27). Therefore, for the query interval of  $|T_Q|$  timestamps, Algorithm 1 takes  $O(\sum_{t_n \in T_Q} (|V| + |E_{t_n}| + (|[t_i, t_n]| \cdot |E_{t_n}|)))$  in total, which can be rewritten as  $O(|T_Q| \cdot (|V| + |E_t|) + (|T_Q| \cdot |E_t|))$ , i.e.,  $O(|T_Q|^2 \cdot |E_t| + |T_Q| \cdot (|E_t| + |V|))$ , where  $|E_t|$  denotes the average number of edges of the graph instances.

## 4 INDEX BASED RELIABLE COMMUNITY SEARCH

To further accelerate the query processing, in this section, we first propose a forest index structure, called *Weighted Core Forest Index* (*WCF-Index*), to maintain the  $(\theta, k)$ -core vertices for each graph instance  $G_{t_n}$ . Then, we develop an index-based dynamic programming algorithm by using the proposed index and derive the reliable score upper bound with great pruning power to accelerate the query algorithm.

### 4.1 WCF-Index

The general idea of this index is to maintain the vertex candidates of the  $(\theta, k)$ -core with regards to the given  $\theta$  and  $k$  at each timestamp, from which we can work out the satisfied CRCs containing  $q$  with the different continuous time intervals.

**Definition 4.1** ( $\theta$ -threshold of a Vertex). Given a graph instance  $G_t = (V, E_t, W_t)$  at a timestamp  $t$  and an integer  $k$ , for a vertex  $u \in V$ , it may have a set of  $\theta$  values and their corresponding  $(\theta, k)$ -core

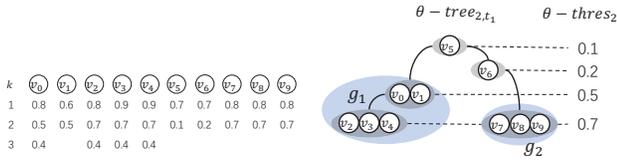


Figure 2:  $\theta$ -threshold of  $G_{t_1}$  Figure 3:  $\theta$ -tree of  $G_{t_1}$ ,  $k=2$

subgraphs containing  $u$ . Thus, we take the largest  $\theta$  value in the  $\theta$  set as the  $\theta$ -threshold of  $u$ , denoted as  $\theta\text{-thres}_k(u, G_t)$ .

**Example 4.1.** Figure 2 shows the  $\theta$ -threshold of vertices in  $G_{t_1}$  in Figure 1 (a) with regards to different  $k$  values, e.g.,  $\theta\text{-thres}_2(v_1, G_t)$  is 0.5 because  $(0.5, 2)$ -core (i.e.,  $\{v_0, v_1, v_2, v_3, v_4\}$ ) exists in  $G_{t_1}$ , but no one  $(\theta', 2)$ -core containing  $v_1$  exists if  $\theta' > 0.5$ .

According to the above definition and the example, we are able to justify whether a vertex  $v$  is contained in a  $(\theta, k)$ -core for given  $\theta$  and  $k$  if the  $\theta$ -threshold values of vertices are maintained. However, the  $\theta$ -threshold only implies the vertex candidatures of a  $(\theta, k)$ -core, but fails to reflect the structural connectivity of vertices. Therefore, it is highly desirable to design an index structure for maintaining the  $\theta$ -threshold and the structure information of vertices together.

Yang et al. in [44] proposed a forest-based index to query  $(k, \eta)$ -core in a static uncertain graph where  $k$  implies the degree constraint of the vertex and  $\eta$  implies the probability of the vertex to appear in the subgraph. By maintaining  $\eta\text{-tree}_k$  for each  $k$ , it can accelerate the search of all the  $(k, \eta)$ -core with custom  $\eta$  requirements. Motivated by  $\eta\text{-tree}_k$ , in this work, we extend the concept of  $\eta\text{-tree}_k$  to the dynamic weighted network to construct the  $\theta\text{-tree}_{k,t}$  for each  $k$  at time  $t$ , which can support quick retrieval of local maximal  $(\theta, k)$ -core from the indexed graph instance  $G_t$ .

**Definition 4.2** ( $\theta\text{-tree}_{k,t}$ ). Given a graph instance  $G_t$ , an integer  $k$ ,  $\theta\text{-tree}_{k,t}$  index is a tree structure, satisfying

- (1) **Node:** each tree node  $\mathbb{V}$  is a set of maximal connected vertices in  $G_t$  with same  $\theta$ -threshold value, denoted as  $\mathbb{V}, \theta$ , i.e.,  $\forall v \in \mathbb{V}, \theta\text{-thres}_k(v, G_t) = \mathbb{V}, \theta$ ;
- (2) **Parent-child relationship:** for a node  $\mathbb{W}$ ,  $N_{\mathbb{W}}(G_t)$  denotes the tree nodes that are connected to  $\mathbb{W}$  in  $G_t$  with  $\theta$ -threshold smaller than  $\mathbb{W}, \theta$ . The parent node  $\mathbb{V}$  of  $\mathbb{W}$  is the node with the largest  $\theta$ -threshold in  $N_{\mathbb{W}}(G_t)$ , i.e.  $\mathbb{V} = \text{argmax}_{\mathbb{V} \in N_{\mathbb{W}}(G_t)} \mathbb{V}, \theta$ .

**Example 4.2.** Figure 3 presents the constructed  $\theta\text{-tree}_{2,t_1}$  of  $G_{t_1}$  from  $\theta$ -threshold of  $G_{t_1}$  in Figure 2. If we search  $(0.5, 2)$ -core on  $\theta\text{-tree}_{2,t_1}$ , three tree nodes will be returned, i.e.,  $\{v_0, v_1\}$ ,  $\{v_2, v_3, v_4\}$ , and  $\{v_7, v_8, v_9\}$ . These tree nodes can induce two  $(0.5, 2)$ -core, i.e.,  $g_1$  and  $g_2$ .

$\theta$ -tree can be composed of several trees where each tree represents a connected component in the graph instance. We denote  $\mathcal{I}$  as the WCF-Index where  $\mathcal{I}[k][t]$  represents the  $\theta\text{-tree}_{k,t}$  of each  $k$  and  $t$  in the dynamic network.

**Remark 1.** In this work, we set the  $\theta$ -threshold as the standard values  $\{0, 0.1, 0.2, \dots, 0.9, 1\}$ . If the  $\theta$ -threshold of a vertex is not in the standard set, we will round it down to the nearest standard value. Accordingly, fetching  $(\theta, k)$ -core with non-standard  $\theta$

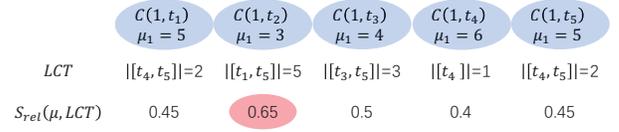


Figure 4: UBR of  $[t_1, t_5]$

value will also be processed as the nearest rounded down standard value. For instance, to fetch  $(\theta, k)$ -core with  $\theta = 0.55$ , the index accesses the tree nodes from  $\theta$ -threshold of 0.5 and then examines the  $\theta$ -threshold of vertices in the tree node  $\mathbb{V}$  if  $\mathbb{V}, \theta < 0.55$ . In the following discussion, we skip this process for simplicity.

## 4.2 Dynamic Programming based CRC search

To solve the most reliable community search problem, we need to compare CRC with different duration. In this section, we develop a dynamic programming algorithm based on the recursive relation of CRCs ending in consecutive timestamps and utilize the WCF-Index to search CRC with varying duration efficiently.

Assume that the lasting time interval of the CRC is fixed (so does the duration), then we only need to extract the CRC with the largest size. Given the duration of the CRC is  $d$  and the last timestamp it spans is  $t_n$ , we denote the maximal CRC w.r.t. the query input as  $C(d, t_n)$ . We can easily derive the following recursive relation between CRCs:

$$C(d, t_n) \subseteq C(d-1, t_{n-1}) \cap C(d-1, t_n) \quad (4)$$

The base situation is  $C(1, \cdot)$  that can be retrieved from WCF-Index. Based on Eq. 4, we can devise a DP algorithm to compute  $C(d, t_n)$ . More specifically, to get  $C(d, t_n)$ , we simply compute the intersection of  $C(d-1, t_{n-1})$  and  $C(d-1, t_n)$  and extract the local maximal  $(\theta, k)$ -core using core decomposition. The intermediate result of  $C(d, t_n)$  with varying  $d$  is maintained to support the adoption of the dynamic programming.

If at a timestamp  $t_a$ , the maximal community  $C(1, t_a)$  does not exist, i.e. for a given query, there is no such subgraph satisfying  $(\theta, k)$ -core constraint at time  $t_a$ , then it implies that further calculations depending on  $C(1, t_a)$  are unnecessary. In this work, these kinds of timestamps like  $t_a$  are called anchored timestamps of a query time interval. The anchored timestamps split the query interval  $[t_i, t_j]$  into several non-overlapping time intervals  $T_S = \{T_1, T_2, \dots\}$ . For each interval  $T_i \in T_S$ , we can compute the upper bound of the reliability score of the communities.

**PROPERTY 4.1** ( $S_{rel}$  UPPER BOUND OF CRC W.R.T.  $T_i$ ). Given a time interval  $T_i = [t_s, t_e]$  where  $C(1, t_n)$  exists for every  $t_n \in [t_s, t_e]$ , we can construct an array  $M = (\mu_s, \mu_{s+1}, \dots, \mu_e)$  to store the size of  $C(1, \cdot)$ , where  $\mu_n$  denotes the size of  $C(1, t_n)$ . The upper bound reliability score (UBR) in this time interval can be calculated by:

$$UBR_{T_i} = \max_{\mu_n \in M} \left( (1 + \alpha^2) \cdot \frac{\mu_n \cdot LCT(\mu_n, M)}{(\alpha^2 \cdot \mu_n) + LCT(\mu_n, M)} \right) \quad (5)$$

where  $LCT(\mu_n, M)$  stands for the length of the Longest Consecutive Timestamps  $S$  of  $\mu_n$  in array  $M$  such that  $\{i \in S | \mu_i \leq \mu\}$ .

**Example 4.3.** For a community  $C(1, t_n)$  of size  $\mu_n$ , the largest reliability score of a CRC constructed by  $C(1, t_n)$  is determined by the longest possible duration that  $\mu_n$  can remain. Hence, the upper bound score of a time interval is the maximum value among all the largest possible scores for each  $C(1, \cdot)$  size. Figure 4 shows an illustrative example of calculating the UBR of interval  $[t_1, t_5]$  where we assume  $|V_{max}^k| = 10$ . The size of  $C(1, t_3)$  is 4 and the longest continuous timestamps for this size is 3 ( $[t_3, t_5]$ ), so the largest possible score of the CRC constructed by  $C(1, t_3)$  is  $0.5/10 * 4 + 0.5/5 * 3 = 0.5$ . The largest possible reliability score is obtained by a CRC constructed by  $C(1, t_2)$  that contains three vertices and spans for five timestamps.

**Remark 2.** Similar to Property 2, we use  $UBR_{T_i}^d$  to denote the upper bound calculated by the size of  $C(d, \cdot)$ . However, it can determine the maximum reliability score of the community with duration longer than  $d$ . We can derive that

$$UBR_{T_i}^d = \max_{\mu_n \in M} \left( \frac{(1 + \alpha^2) \cdot \mu_n \cdot (d + LCT(\mu_n, M) - 1)}{(\alpha^2 \cdot \mu_n) + (d + LCT(\mu_n, M) - 1)} \right) \quad (6)$$

where  $\mu_n$  denotes the size of  $C(d, t_n)$  and  $(LCT(\mu_n, M) - 1)$  represents the additional lasting timestamps of size  $\mu_n$  on top of  $d$ . In the process of community search,  $UBR_{T_i}$  can be updated with different duration of CRC and provide sustainable pruning power. Having  $UBR_{T_i}$  calculated for each interval  $T_i$  and updated during the community exploration, we can skip exploring CRCs if  $UBR_{T_i}$  is no larger than the reliability score of intermediate community candidates we have obtained during the query processing.

Algorithm 2 presents the detailed dynamic programming procedure of the *WCF-Index* based CRC Search. We first initialize a table  $L_C$ ,  $maxS$ , and  $C_{opt}$  to store the extracted communities, the maximum reliability score, and the most reliable community (line 1). For each timestamp  $t_n \in [t_i, t_j]$ , we can obtain  $C(1, t_n)$  from the  $\theta$ -tree $_{k, t_n}$  index, and store it in  $L_C[1][n]$  (lines 2-6). In addition, we also determine whether  $t_n$  is an anchored timestamp. Then we split  $T_Q$  into several non-overlapping time intervals  $T_S = \{T_1, T_2, \dots\}$  with valid  $C(1, \cdot)$  by the anchored timestamps, and calculate their upper bound reliability score (lines 7-8). For each individual time interval  $T_i = [t_s, t_e]$ , if its upper bound is no larger than  $maxS$ , then the time interval is pruned (line 10). Otherwise, we initialize an array  $Q$  to store the size of CRC and compute the CRC with various duration  $d$  based on Eq. 4, and store the intermediate CRC  $C(d, t_x)$  in  $L_C[d][x]$  (lines 11-16). Then, we update  $maxS$  and  $C_{opt}$  and add the size of  $C(d, t_x)$  to  $Q$  for upper bound calculation (lines 17-20). Once  $C(d, \cdot)$  has been explored for every  $t_x \in [t_s, t_e]$ , we can update the  $UBR$  and determine whether it is necessary to explore communities with longer duration in this interval (line 21-22). Finally, the algorithm returns the most reliable community  $C_{opt}$  whose reliability score is the largest.

The time complexity of Algorithm 2 is dominated by the operation of finding CRCs with various duration (lines 10-23) as the  $C(1, \cdot)$  community can be queried from the index in constant time. In the worst case, there are up to  $|T_Q|^2$  subgraphs to be explored and the community construction takes  $O(|E_t|)$  complexity, where  $T_Q$  is the query interval and  $|E_t|$  is the average number of edges of graph instance  $G_t$ . The total complexity is  $O(|T_Q|^2 \cdot |E_t|)$ . Compared with the EEF-based Online CRC Search Algorithm, *WCF-Index* can avoid

---

#### Algorithm 2: *WCF-Index* based CRC Search

---

**Input:** A dynamic weighted graph  $\mathcal{G} = \{G_{t_1}, G_{t_2}, \dots, G_{t_T}\}$ , query time  $[t_i, t_j]$ , the *WCF-Index*  $\mathcal{I}$ , integer  $k$ , weight threshold  $\theta$ , query vertex  $q$   
**Output:** the most reliable community  $C_{opt}$

- 1  $L_C \leftarrow [[]]; maxS \leftarrow 0; C_{opt} \leftarrow \emptyset;$
- 2 **for**  $t_n \in [t_i, t_j]$  **do**
- 3     Extract  $C(1, t_n)$  from  $\mathcal{I}[k][t_n];$
- 4     **if**  $C(1, t_n)$  is  $\emptyset$  **then**
- 5         set  $t_n$  as anchored timestamp;
- 6     **else**  $L_C[1][n] \leftarrow C(1, t_n);$
- 7 Get all the consecutive timestamps  $T_S = \{T_1, T_2, \dots\}$  split by the anchored timestamp;
- 8 Calculate upper bound  $ub = \{UBR_{T_1}^1, UBR_{T_2}^1, \dots\}$  for each consecutive time sequence by Equation 5;
- 9 **for**  $T_i = [t_s, t_e] \in T_S$  by descending of  $UBR_{T_i}$  **do**
- 10     **if**  $UBR_{T_i}^1 \leq maxS$  **then continue** ;
- 11     **for**  $d \leftarrow 1$  **to**  $\lfloor [t_s, t_e] \rfloor$  **do**
- 12          $M \leftarrow [];$
- 13         **for**  $t_x \in [t_s, t_e]$  **do**
- 14             **if**  $d \leq \lfloor [t_s, t_x] \rfloor$  **then**
- 15                 **if**  $d > 1$  **then**
- 16                      $L_C[d][x] \leftarrow$  local maximal  $k$ -core in  $L_C[d-1][x-1] \cap L_C[d-1][t_x];$
- 17                      $maxS \leftarrow \max(S_{rel}(L_C[d][x]), maxS);$
- 18                     **if**  $S_{rel}(L_C[d][x]) \geq S_{rel}(C_{opt})$  **then**
- 19                          $C_{opt} \leftarrow L_C[d][x];$
- 20                      $M.append(|L_C[d][x]|);$
- 21             Calculate  $UBR_{T_i}^d$  by Equation 6;
- 22             **if**  $UBR_{T_i}^d \leq maxS$  **then break** ;
- 23 **Return**  $C_{opt}$

---

searching a large number of edges, which helps to reduce the time cost of computing  $C(1, \cdot)$ .

## 5 WCF INDEX CONSTRUCTION, MAINTENANCE, AND COMPRESSION

In this section, we describe the procedure of index construction, and propose index maintenance and compression strategies to support efficient query processing over dynamic weighted networks and reduce the time and space cost of the index.

### 5.1 WCF-Index Construction

The main idea of constructing *WCF-Index* is to build the  $\theta$ -tree $_{k, t}$  for each graph instance  $G_t$  for  $k \in [1, k_{max}]$ , where  $k_{max}$  denotes the maximum core number of the vertex in  $G_t$ . To obtain the  $\theta$ -tree $_{k, t}$ , we group the vertices by their  $\theta$ -threshold value and add the vertex groups as tree nodes into the  $\theta$ -tree according to the  $\theta$ -threshold and connectivity of the tree nodes, i.e., the vertex groups.

Algorithm 3 presents the procedure of building the *WCF-Index*  $\mathcal{I}$ . For each graph instance  $G_t$ , we build  $\theta$ -tree $_{k,t}$  for each available  $k$  by first computing  $\theta$ -threshold of vertices and then construct and insert tree nodes to the  $\theta$ -tree index. We first initialize two graphs  $G_{pre}$  and  $G_{cur}$  to store intermediate states of edge filtering (line 3). Then we iteratively pick  $\theta' \in \Theta$  in descending order. For each  $\theta'$ , we get the edge set  $E_{t,\theta'}$  whose weights are no less than  $\theta'$  and obtain the induced subgraph  $G_t[E_{t,\theta'}]$  as the current state  $G_{cur}$ . We can obtain a set of vertices  $V_{\theta'}$  whose core number in  $G_{cur}$  is increased with regards to the core number in the last state  $G_{pre}$ . This implies that for a vertex  $w \in V_{\theta'}$ , the  $\theta$ -threshold of  $w$  is  $\theta'$  w.r.t. its increased core number (lines 4-7). After that, we set the previous state  $G_{pre}$  to be  $G_{cur}$  and obtain the distinct values of the newly increased core numbers  $K$  (lines 8-9). Then, according to the newly identified  $\theta$ -threshold, we can construct and add tree nodes to the  $\theta$ -tree $_{k',t}$  for  $k' \in K$  (lines 10-21). To do that, we get  $\theta$ -tree $_{k',t}$  from  $\mathcal{I}[k'][t]$ , then we identify the groups of connected vertices whose  $\theta$ -threshold at  $k'$  is  $\theta'$  as the tree node  $\mathbb{X}$  (lines 11-14). To determine the position of  $\mathbb{X}$ , we find each tree node  $\mathbb{Y}$  that contains any neighbor  $v$  of  $G_t[\mathbb{X}]$  and its root  $\mathbb{Z}$ , so that  $\mathbb{X}, \mathbb{Y}, \mathbb{Z}$  are connected (lines 15-18). If  $\mathbb{Z}.\theta > \mathbb{X}.\theta$ ,  $\mathbb{X}$  is assigned as the parent of  $\mathbb{Z}$ , otherwise, their  $\theta$ -threshold are the same because the smaller  $\theta'$  has not been visited yet, so we need to merge  $\mathbb{X}$  to  $\mathbb{Z}$  (lines 19-21). After iterating all the standard threshold values, we can construct all the  $\theta$ -tree $_{k,t}$  completely for each possible  $k$  of each  $G_t$  and return the *WCF-Index*  $\mathcal{I}$ . The space cost of *WCF-Index* is  $O(\sum_{t_i \in T} \sum_{u \in V} core(u, G_{t_i}))$  as each vertex  $u$  appears  $core(u, G_{t_i})$  times in each graph instance.

**Example 5.1.** Consider the graph instance  $G_{t_1}$  in Figure 1, we calculate the  $\theta$ -threshold of its vertices by inducing  $G_{t_1}[E_{t_1,\theta}]$  with increasing  $\theta$ . Upon inducing  $G_{t_1}[E_{t_1,0.5}]$ , comparing to  $G_{t_1}[E_{t_1,0.6}]$ , we can observe the core number increase of  $v_0$  and  $v_1$  from 1 to 2, which implies  $\theta$ -thres $_2(v_0, G_{t_1}) = \theta$ -thres $_2(v_1, G_{t_1}) = 0.5$ . So we add tree node to  $\theta$ -tree $_2$  because the core number is increased to 2. One node  $\mathbb{X} = \{v_0, v_1\}$  is constructed as  $v_0$  and  $v_1$  are connected.  $\theta$ -tree $_{2,t_1}$  has two nodes  $\mathbb{Y}_1 = \{v_2, v_3, v_4\}$  and  $\mathbb{Y}_2 = \{v_7, v_8, v_9\}$  from previous steps and the node that contains neighbors of  $\mathbb{X}$  is  $\mathbb{Y}_1$ , whose root is itself.  $\mathbb{X}$  can thus be added as the parent of  $\mathbb{Y}_1$  since  $\mathbb{Y}_1.\theta > \mathbb{X}.\theta$ . By now,  $\theta$ -tree $_2$  contains three tree nodes. After inducing  $G_{t_1}[E_{t_1,0}]$  we can construct  $\theta$ -tree $_{2,t_1}$  as Figure 3.  $\theta$ -tree $_{1,t_1}$  and  $\theta$ -tree $_{3,t_1}$  will also be obtained.

## 5.2 WCF-Index Maintenance

In general, dynamic networks might have subtle changes, i.e., a small percentage of edges and vertices change or update, in two consecutive timestamps. It is time-consuming to simply compute the  $\theta$ -threshold for all vertices and re-construct the index. Therefore, in this section, it is highly desirable to develop an index maintenance strategy and update the index using the small number of changed edges and vertices only.

To explore the relationships of vertices with regards to different core numbers, we are motivated by the work in [36] that proposed an incremental core number update method for an evolving graph. It supports to locate a small set of vertices whose core number will be affected by using the below two concepts. The other works

---

### Algorithm 3: $\theta$ -Tree Construction

---

**Input:** A dynamic weighted graph  $\mathcal{G} = \{G_{t_1}, G_{t_2}, \dots, G_T\}$

**Output:** The *WCF-Index*  $\mathcal{I}$

```

1  $\mathcal{I} \leftarrow \emptyset$ ;
2 for  $G_t \in \mathcal{G}$  do
3    $G_{pre} \leftarrow \emptyset$ ;  $G_{cur} \leftarrow \emptyset$ ;
4   for  $\theta' \in \Theta$  in descending order do
5      $E_{t,\theta'} \leftarrow \{e \in E_t \mid W_t(e) \geq \theta'\}$ ;
6      $G_{cur} \leftarrow G_t[E_{t,\theta'}]$ ;
7      $V_{\theta'} \leftarrow \{u \in V \mid core(u, G_{cur}) > core(u, G_{pre})\}$ ;
8      $G_{pre} \leftarrow G_{cur}$ ;
9      $K \leftarrow$  set of values  $\{core(u, G_{cur}) \mid u \in V_{\theta'}\}$ ;
10    for each  $k' \in K$  do
11       $\theta$ -tree $_{k',t} \leftarrow \mathcal{I}[k'][t]$ ;
12       $H \leftarrow \{v \in V_{\theta'} \mid \theta$ -thres $_{k'}(v, G_t) = \theta'\}$ ;
13      for each connected vertex set  $\mathbb{X} \subseteq H$  do
14         $\mathbb{X}$  as a new tree node of  $\theta$ -tree $_{k',t}$ ;
15        for each  $v \in N(G_t[\mathbb{X}])$  do
16          if  $\theta$ -thres $(v, G_t) > \theta'$  then
17             $\mathbb{Y} \leftarrow$  get the node containing  $v$ ;
18             $\mathbb{Z} \leftarrow$  get the root of  $\mathbb{Y}$ ;
19            if  $\mathbb{Z}.\theta > \mathbb{X}.\theta$  then  $\mathbb{Z}.\text{parent} \leftarrow \mathbb{X}$ ;
20            else merge node  $\mathbb{X}$  to  $\mathbb{Z}$ ;
21       $\mathcal{I}[k'][t] \leftarrow \theta$ -tree $_{k',t}$ ;
22 Return  $\mathcal{I}$ 

```

---

[15, 26, 45] also follow the similar concepts of *subcore* and *purecore* in [36] to maintain the core numbers.

**Definition 5.1** (*subcore* in [36]). Given a graph  $G=(V, E)$  and a vertex  $u \in V$ , the *subcore* of  $u$  denoted as  $S_u$ , is a set of vertices having the same core number as  $u$  and connected with  $u$  via a path, where each vertex on the path has the same core number as  $u$ .

**Definition 5.2** (*purecore* in [36]). Given a graph  $G=(V, E)$  and a vertex  $u \in V$ , the *purecore* of  $u$  denoted as  $P_u$ , is a set of vertices where each vertex  $w \in P_u$  satisfies:

- (1) Condition 1: the core number  $core(w, G)$  of  $w$  is equal to the core number  $core(u, G)$  of  $u$ .
- (2) Condition 2:  $w$  has a set  $W$  of neighbors whose core numbers are no less than  $core(w, G)$ , and  $|W|$  is larger than  $core(u, G)$ .
- (3) Condition 3:  $w$  is connected to  $u$  via a path, where each vertex on the path satisfies the conditions (1) and (2).

Specifically, given two vertices  $u$  and  $v$  in a graph  $G=(V, E)$ , and  $core(u, G) \leq core(v, G)$ , if an edge  $(u, v)$  is removed from  $G$ , then only the vertices in the subcore set  $S_u$  may have their core number decreased; if an edge  $(u, v)$  is added to  $G$ , then only the vertices in the purecore set  $P_u$  may have their core number increased. Thus, we can extend the rules to the weighted graph, in which the updates include edge insertion, edge deletion, and edge weight change.

Considering that  $G_{t'_n}$  is obtained by inserting an edge  $(u, v)$  with weight  $\theta'$  to  $G_{t_n} = (V, E_{t_n}, W_{t_n})$ . For a vertex  $w \in V$ , if

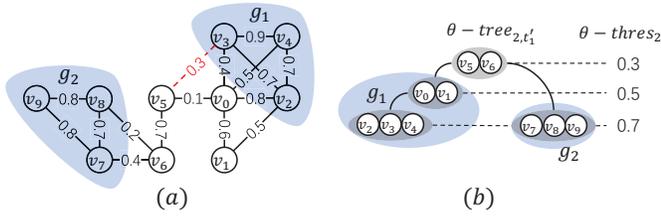


Figure 5:  $\theta$ -tree of  $G_{t'_1}$ ,  $k=2$

$\theta$ -thres $_k(w, G_{t_n}) = \theta''$  and  $\theta'' \geq \theta'$ , then  $\theta$ -thres $_k(w, G_{t'_n})$  will remain unchanged.

PROPERTY 5.1 (INSERTION OF AN EDGE). Given a graph  $G_{t_n}$  and  $\theta$ -tree $_{k,t_n}$  for each  $k \in [1, k_{max}]$ , and two vertices  $u$  and  $v$  such that  $\theta$ -thres $_k(u, G_{t_n}) \leq \theta$ -thres $_k(v, G_{t_n})$ , if an edge  $(u, v)$  is inserted with weight  $\theta'$ , then only the vertices  $\{w \in P_u | \theta$ -thres $_k(w, G_{t_n}) < \theta'\}$  may have their  $\theta$ -threshold increased.

PROPERTY 5.2 (DELETION OF AN EDGE). Given a graph  $G_{t_n}$  and  $\theta$ -tree $_{k,t_n}$  for each  $k \in [1, k_{max}]$ , and two vertices  $u$  and  $v$  such that  $\theta$ -thres $_k(u, G_{t_n}) \leq \theta$ -thres $_k(v, G_{t_n})$ , if an edge  $(u, v)$  is removed with weight  $\theta'$ , then only the vertices  $\{w \in S_u | \theta$ -thres $_k(w, G_{t_n}) \leq \theta'\}$  may have their  $\theta$ -threshold decreased.

The edge insertion and edge deletion can be treated as the update of edge weight.

PROPERTY 5.3 (UPDATE OF EDGE WEIGHT). Given a graph  $G_{t_n}$  and  $\theta$ -tree $_{k,t_n}$  for each  $k \in [1, k_{max}]$ , two vertices  $u$  and  $v$  such that  $\theta$ -thres $_k(u, G_{t_n}) \leq \theta$ -thres $_k(v, G_{t_n})$ , we have (1) if the weight of edge  $(u, v)$  increases from  $\theta_1$  to  $\theta_2$  (w.r.t.  $\theta_1 < \theta_2$ ), then only the vertices  $\{w \in P_u | \theta$ -thres $_k(w, G_{t_n}) \in [\theta_1, \theta_2)\}$  may have their  $\theta$ -threshold increased; (2) if the weight of edge  $(u, v)$  decreases from  $\theta_2$  to  $\theta_1$  (w.r.t.  $\theta_1 < \theta_2$ ), then only the vertices  $\{w \in S_u | \theta$ -thres $_k(w, G_{t_n}) \in (\theta_1, \theta_2]\}$  may have their  $\theta$ -threshold decreased.

**Example 5.2.** Figure 5(a) shows an updated graph instance of  $G_{t'_1}$  by adding an edge  $(v_3, v_5)$  with weight 0.3 to  $G_{t_1}$ . We can identify  $P_{v_3} = \{v_0, v_2, v_3, v_5, v_6, v_7, v_8\}$  in  $G_{t_1}$ . In addition, the  $\theta$ -threshold of  $v_5$  or  $v_6$  is less than 0.3 and other vertices'  $\theta$ -threshold is no less than 0.3. According to Property 5.1, only  $v_5$  and  $v_6$  may have their  $\theta$ -threshold increased. After recalculating  $\theta$ -thres $_2(v_5, G_{t'_1})$  and  $\theta$ -thres $_2(v_6, G_{t'_1})$ , we can update the tree index from Figure 3 to Figure 5(b).

### 5.3 WCF-Index Compression

Sometimes, the graph instances of some consecutive timestamps may be similar because the edge weight and graph structure change progressively over time. Besides that, one tree node usually contains multiple vertices as it gathers many connected vertices with the same threshold. It is likely to have much duplicate information across  $\theta$ -tree indices. Thus, we need to develop an index compression strategy in order to reduce the redundancy. The key idea is to utilize a virtual node to replace the tree node that contains multiple vertices and appears frequently. The actual vertices of the virtual nodes are stored in an auxiliary table.

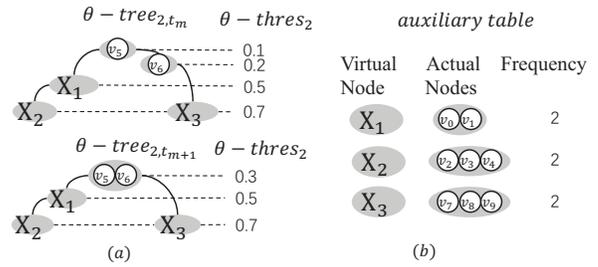


Figure 6: Compressed WCF-Index and auxiliary table

**Example 5.3.** Figure 6 shows an example of compressed  $\theta$ -tree $_{2,t_1}$  and  $\theta$ -tree $_{2,t'_1}$  where the nodes with vertices  $\{v_0, v_1\}$ ,  $\{v_2, v_3, v_4\}$ ,  $\{v_7, v_8, v_9\}$  are replaced by virtual nodes  $X_1, X_2, X_3$ .

Each virtual node can be regarded as an encoding of a unique vertex set. The space cost can be reduced if we use an auxiliary table to only maintain one copy of the tree nodes that frequently appear in WCF-Index, and keep a virtual id at the positions of these nodes in WCF-Index. To make the compression, we only select the tree nodes that can bring positive space gain.

**Definition 5.3** (Space Gain). The space gain is defined as the space saved from replacing the tree node  $\mathbb{X}$  as a virtual node  $X$ , that is

$$SG(\mathbb{X}) = f * (|\mathbb{X}| - 1) - |\mathbb{X}|$$

where  $f$  is the frequency of  $\mathbb{X}$  that appears in the WCF-Index, and  $|\mathbb{X}|$  is the size the vertex set.

By scanning and counting the frequency of tree nodes, we can calculate the space gains for all the tree nodes, and generate the compressed index easily. We do not provide the pseud codes in this paper due to the limited space.

## 6 EXPERIMENT

We conduct extensive experiments to evaluate the performance of our proposed algorithms, including EEF-based Online CRC Search in Algorithm 1, WCF-Index based CRC Search in Algorithm 2 and  $\theta$ -Tree Construction in Algorithm 3, denoted as EEF-CRC, WCF-CRC and WCF-Construct, respectively. We implement a baseline method based on maximal spanning core (SpanCore) [14], which calculates all the  $k$ -core subgraphs with different  $k$  value in various time intervals. We additionally remove the edges and calculate the reliability score of each candidate subgraph to get the optimal results. We compare the effectiveness of our proposed community model with PC [22] and SC [35]. We also evaluate the effectiveness of index maintenance (WCF-Maintain) and index compression. All the experiments are conducted on a Windows machine with an Intel i9-10900F CPU @ 2.80GHz and 32.0 GB DDR4-RAM.

### 6.1 Experimental Setup

**Datasets.** We conduct the experiments on eight real-world dynamic network datasets collected from SNAP<sup>1</sup> and Network Data Repository<sup>2</sup>. In BitcoinAlpha (BA) and BitcoinOTC (BO) datasets,

<sup>1</sup><https://snap.stanford.edu/data/>

<sup>2</sup><https://networkrepository.com>

**Table 1: Dataset Statistics**

	BitcoinAlpha	BitcoinOtc	Retweet	TAT	Email	Reddit	HepPh	StackOverFlow
$ V  / \widehat{ V }$	3,783 / 688	5,881 / 997	18,470 / 4,249	34,761 / 3,613	986 / 658	55,863 / 4,908	28,093 / 4,009	2,601,977 / 79,629
$ E  / \widehat{ E }$	24,186 / 1,497	35,592 / 2,247	61,157 / 5,554	171,403 / 5,122	332,334 / 2,619	571,927 / 8,556	4,596,803 / 78,535	63,497,050 / 325,080
$ T $	10	10	30	30	30	30	50	100
$\widehat{density}$	0.0065	0.0047	0.0006	0.0015	0.0121	0.0007	0.0099	0.0002
$\widehat{k}_{max}$	6.6	8.5	4.4	12.6	7.9	12.4	118.2	27.1
$\widehat{k}_{query}$	3.5	4.3	1.9	7.1	3.9	5.5	59.4	10.2

**Table 2: Parameters and default values**

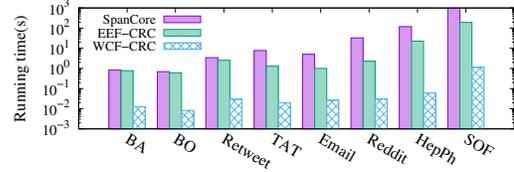
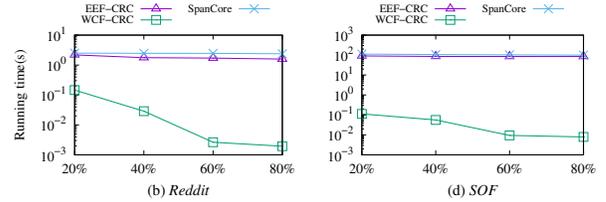
Parameter	Values	Description
$k$	20%, <b>40%</b> , 60%, 80%	% of $k_{max}$
$\theta$	0.0, 0.2, <b>0.4</b> , 0.6, 0.8	weight threshold
$t$	4, 8, <b>12</b> , 16, 20	Time span (snapshots)

the edge weight represents the rating between two users. In the remaining datasets, the edge weight is calculated from the interaction frequency. The edge weight of all the datasets is normalized to  $[0, 1]$  by min-max normalization. The statistics of the dataset are shown in Table 1. The number of vertices and edges are denoted as  $|V|$  and  $|E|$ , respectively. For each dataset, we first sort the edges by chronological order, and then divide them into  $|T|$  partitions, i.e.,  $|E|/|T|$  edges, where  $|T|$  is the target number of graph instances. It guarantees each graph instance contains meaningful  $k$ -core components. For example, the largest dataset *StackOverFlow* (SOF) is divided into  $|T| = 100$  snapshots and the medium-sized datasets, e.g., *TechAsTopology* (TAT), *Retweet*, etc. are divided into  $|T| = 30$  instances. We denote  $\widehat{|V|}$ ,  $\widehat{|E|}$ ,  $\widehat{density}$  and  $\widehat{k}_{max}$  as the average of vertex numbers, edge numbers, density, and the largest core numbers of the  $|T|$  graph instances, respectively.

**Parameters.** Table 2 shows the detailed setting of the parameters used in the experiments. To better fit the dataset and cover more meaningful situations, we vary the query parameter  $k$  as 20%, 40%, 60%, 80% of  $\widehat{k}_{max}$  for each dataset with the default value 40%. The threshold value varies from 0.0 to 0.8 with the default value of 0.4. The length of the query time interval was specified as 4, 8, 12, 16, 20 with the default value of 12. Their default values are marked in bold font. We also vary the  $\alpha$  parameter from 0 to 6 to show its effect on the returned community. We sample 100 query vertices whose core numbers are uniformly distributed in  $[1, k_{max}]$  for each dataset and report their average running time as the time cost. The average core number of the 100 query vertices is shown in Table 1 as  $\widehat{k}_{query}$ . In general,  $\widehat{k}_{query}$  is around 50% of  $k_{max}$ , which reflects the common scenario of query vertex.

## 6.2 Evaluation of Query Efficiency

In this section, we present the performance of SpanCore, EEF-CRC, and WCF-CRC under the default parameter settings. Figure 7 demonstrates the time cost when we run the three algorithms over eight datasets. Both of the proposed algorithms outperform the baseline algorithm *SpanCore*. *EEF-CRC* is slightly faster than *SpanCore*, as they both need to determine the core number of the vertices but *EEF-CRC* only requires the local information of the query vertex.


**Figure 7: CRC Query Time on All Datasets**

**Figure 8: CRC Query Time with Varying  $k$** 

*WCF-CRC* runs much faster than *EEF-CRC*. For instance, *WCF-CRC* reduces the time cost of *EEF-CRC* by about 89 times.

To show the impact of each parameter, we also evaluate the efficiency of the proposed algorithms by varying the values of parameters  $k$ ,  $\theta$ , and  $t$ , respectively. We utilize two representative datasets *Reddit* and *SOF* to demonstrate the experimental results.

**Varying  $k$ .** Figure 8 shows the average time cost of our proposed algorithms when  $k$  varies from 20% to 80% of the corresponding  $\widehat{k}_{max}$  values. *WCF-CRC* is significantly more efficient than the other two algorithms, and *SpanCore* consumes the most time in all settings. For instance, in *Reddit*, *SpanCore* takes 2.46s, *EEF-CRC* takes 1.76s, while *WCF-CRC* only needs 0.03s to complete the query processing in *Reddit* dataset where  $k$  is 40% of the average large core number (i.e.,  $k = 5$ ). With the increase of  $k$ , all the algorithms consume decreasing time. But *SpanCore* and *EEF-CRC* are less sensitive to  $k$  than *WCF-CRC* because they need to scan all the edges and compute the core numbers of the vertices induced by the edges, while *WCF-CRC* can directly retrieve the core numbers using *WCF-Index*.

**Varying threshold  $\theta$ .** Figure 9 shows the average time cost of our proposed algorithms when the threshold  $\theta$  varies from 0 to 0.8. *WCF-CRC* outperforms *SpanCore* and *EEF-CRC* significantly. For instance, in *Reddit*, when  $\theta = 0.2$ , *SpanCore* takes 3.33s, *EEF-CRC* takes 2.41s while *WCF-CRC* takes 0.12s. *WCF-CRC* is faster than the other two by more than 20 times. When  $\theta$  is given as 0.8, the three algorithms take 1.62s, 1.17s and 0.002s, respectively, i.e.,

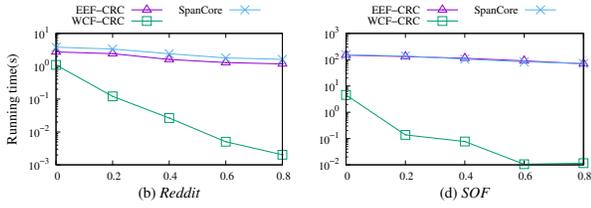


Figure 9: CRC Query Time with Varying  $\theta$

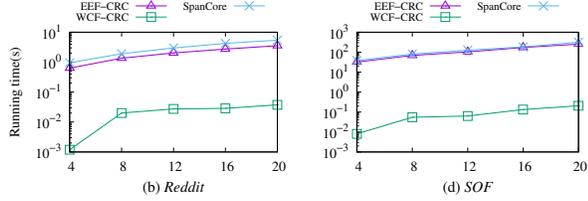


Figure 10: CRC Query Time with Varying  $|T_Q|$

*WCF-CRC* can reduce the time cost of *EEF-CRC* by about 500 times. With the increase of  $\theta$ , the speedup trend of *WCF-CRC* becomes significant because there are small number of tree nodes in *WCF-Index* when  $\theta$  is set as a large value, i.e., more vertices can be pruned. Similarly, *SpanCore* and *EEF-CRC* also consume less time because the significant number edges can be pruned with the higher *theta* threshold. However, scanning all the edges in the locally connected subgraph for each query vertex is inevitable.

Varying time span  $|T_Q|$ . Figure 10 shows the average time cost of the proposed algorithms when the time span  $|T_Q|$  varies from 4 to 20. All the algorithms consume higher time cost when the query time span  $|T_Q|$  increases. For instance, in *Reddit* dataset, *SpanCore* takes 0.96s, 1.91s, 2.98s, 4.24s, and 5.44s. *EEF-CRC* takes 0.63s, 1.35s, 2.04s, 2.73s and 3.55s, respectively. But, *WCF-CRC* only takes 0.0012s, 0.02s, 0.027s, 0.029s and 0.037s at the same settings. The growth of runtime is consistent with the time complexity in Section 3 and Section 4.2.

Efficiency of Upper Bound. Figure 11 shows the efficiency of the proposed algorithms with (i.e., *EEF-UB*, *WCF-UB*) and without (i.e., *EEF-Base*, *WCF-Base*) using the upper bound pruning strategy on the *Reddit* dataset with varying  $|T_Q|$ . We can find that the pruning capability of the upper bound can be accelerated with the increase of the query time interval. For instance, *EEF-UB* is faster than *EEF-Base* by 5% when  $|T_Q| = 4$ , but the acceleration can achieve by 10% when  $|T_Q| = 20$ . Compared to *WCF-Base*, the improvement of *WCF-UB* is not significant. Since the nature of the upper bound is to estimate the maximum reliability score of the community candidates for each potential time interval, there are more chances to prune more intermediate community candidates when query time interval is large.

### 6.3 Evaluation of Index Construction and Maintenance

In Figure 12, we report the runtime of *WCF-Construct* for all the graph instances of eight datasets. Generally, it takes around 10s

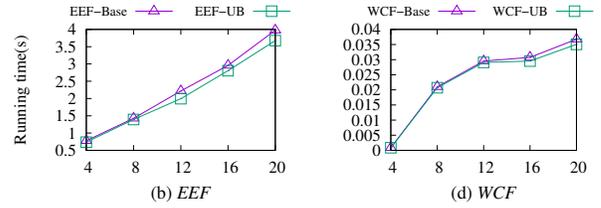


Figure 11: Upper Bound Evaluation with Varying  $|T_Q|$

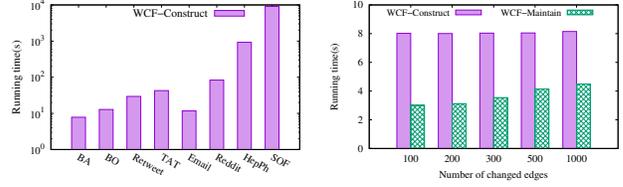


Figure 12: Index Construction Time on All Datasets

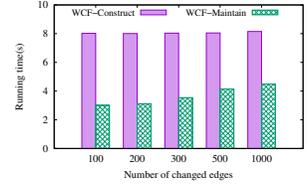


Figure 13: Index Construction Time by Maintenance

for datasets with small number of vertices like *BA*, *BO*, and *Email*. For the largest dataset like *SOF*, it needs 25h to complete the index construction.

We also evaluate the effectiveness of the index maintenance method. Taking the first graph instance of *Reddit* as the base, we randomly sample 100, 200, 300, 500, 1000 edges and mix the operation of edge insertion, deletion, and weight update to generate a synthetic instance. Figure 13 shows the time cost of *WCF-Construct* and *WCF-Maintain* on the synthetic graph, where the speed up of *WCF-Maintain* is significant. For instance, reconstructing the index takes 8s, but *WCF-Maintain* only takes 4s when 1000 edges are updated.

### 6.4 Evaluation of Scalability

We evaluate the scalability of proposed algorithms including *EEF-CRC*, *WCF-CRC* and *WCF-Construct* by using five graph instances from two datasets *Reddit* and *HepPh*. For each dataset, we generate four new datasets with different sizes by randomly sampling 20%, 40%, 60%, 80% edges from the dataset, respectively. The dataset itself is considered with the 100% data size. Figure 14 shows the time cost of *WCF-CRC* and *EEF-CRC* on the size-varying datasets. With the increase of the data size, we can find that the running time of *WCF-CRC* and *EEF-CRC* grow in a gentle trend, which implies that both algorithms are easily applicable to large-scale networks.

Furthermore, we also show the scalability of index construction in Figure 15. From this, we can observe a linear increasing trend of the construction time. For instance, the index constructing time of *Reddit* is 4.7s, 12.8s, 22.3s, 34.0s and 47.9s when the size of the dataset increases as 20%, 40%, 60%, 80% and 100%, respectively.

### 6.5 Evaluation of Index Size with Compression

We show the *WCF-Index* size of the eight datasets in Table 3. For each dataset, we take ten graph instances. The largest dataset *Stack-Overflow* takes 874,931kb and the smallest dataset *BitcoinAlpha* takes 333kb. The compressed size is the sum of the compressed

Table 3: Index Size & Compression (kb)

	BitcoinAlpha	BitcoinOtc	Retweet	TAT	Email	Reddit	HepPh	StackOverFlow
Raw Data	1,108	1,648	4,935	13,619	5,324	37,376	263,168	1,153,024
Original Index	333	645	1,569	7,047	1,139	45,216	252,561	874,931
Compressed	328	616	1,418	6,205	1,133	39,924	158,734	835,721
Auxiliary Table	3	11	120	430	3	834	11,406	5,361

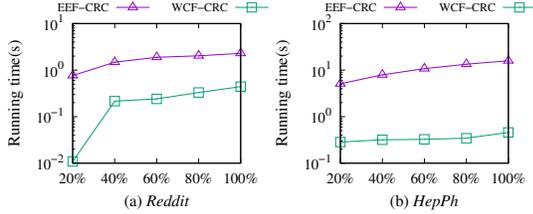


Figure 14: CRC Query Time with Different Sampling Ratios

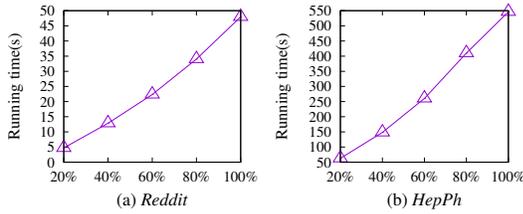


Figure 15: Index Construction Time with Different Sampling Ratios

Table 4: Case Study on *Reddit*

t	Community (ASS)	ASD	ASCore	ASCon
3	SC (65.4)	0.19	7.65	0.80
	PC (50.2)	0.17	5.5	0.85
	CRC (56)	<b>0.31</b>	<b>10.6</b>	<b>0.77</b>
4	SC (30.4)	0.31	6.5	0.89
	PC (49.8)	0.23	7.3	0.83
	CRC (41)	<b>0.39</b>	<b>10.4</b>	<b>0.83</b>
5	SC (23.2)	0.40	6.8	0.91
	PC (48)	0.26	7.9	0.84
	CRC (36)	<b>0.43</b>	<b>10.3</b>	<b>0.83</b>

index and the auxiliary table. A significant compression effectiveness can be observed in Table 3, e.g., the index size of *HepPh* can be compressed to 50% of the original size.

### 6.6 Evaluation of Query Effectiveness

To show the effectiveness of finding communities in dynamic or temporal networks, we compare our  $(\theta, k)$ -core reliable community (CRC) with the Persistent Community (PC) and Stable Community (SC) proposed by Li [22] and Qin [35], respectively. To do this, we select five graph instances of *Reddit* dataset and return the largest community  $C$  obtained by SC, PC, and our CRC. To show the quality of returned communities, we utilize three community quality metrics:

- **Average Snapshot Density (ASD)** measures how dense is the community and captures the intuition that a good community should be closely connected inside. The larger is the density, the closer the community is connected. Average snapshot density is calculated as the average density of the community in each snapshot:  $ASD = \sum_{i=1}^t density(G_{t_i}[C])/t$ .
- **Average Snapshot Core (ASCore)** captures the degree information of vertices and evaluates the closeness of the community. The larger is the core number, the more interactions each vertex will keep with others in the community. ASCore calculates the average value of the average core number of each vertex in each snapshot:  $ASCore = \sum_{i=1}^t (\sum_{v \in V} core(v, G_{t_i}[C])/|V|)/t$ .
- **Average Snapshot Conductance (ASCon)** measures how “well-knit” the graph is. The higher is the conductance, the easier the community can communicate with the vertices outside the community. In the local community detection task, the smaller conductance is desired as it implies the community is tightly self-encapsulated. Here, the average snapshot conductance is calculated as the average conductance of the community in each snapshot:  $ASD = \sum_{i=1}^t conductance(G_{t_i}[C])/t$ .

Table 4 shows the experimental results of evaluating the community quality on the *Reddit* dataset. The community is obtained with the same structural cohesiveness constraint (core number or number of neighbors equals 8). We vary the duration or frequency of the community ( $\tau$  in SC and PC,  $d$  in CRC), denoted by  $t$  in Table 4, to compare the community quality with different temporal features. The size of each community is also provided as the average snapshot size ASS in Table 4. It can be observed that CRC performs best in all the measurements. For example, when  $t = 4$ , CRC finds a community with the highest ASD of 0.39, the highest ASCore of 10.4 and the lowest ASCon of 0.83. When  $t = 4$  and 5, PC and CRC have similar ASCon score. PC generally finds the largest community at the cost of lower density and cohesiveness. CRC outperforms SC with the larger community size and the closer connection.

Community with varying  $\alpha$ . Figure 16 shows an example of obtained reliable community by querying the vertex *funny* in the *Reddit* dataset where  $\alpha$  varies from 0 to 6. With the increase of  $\alpha$ , the duration of the optimal CRC increases and the community size decreases. The progressive change of community duration shows that parameter  $\alpha$  is able to smoothly adjust the balance between community size and duration. Figure 17 shows the trending change of the selected three quality metrics ASD, ASCore and ASCon when  $\alpha$  increases. All the scores increase significant when  $\alpha$  varies from 0 to 2. After that, their trends become steady relatively.

## 7 RELATED WORKS

**Local Community Search in Static and Time-varying Networks.** Local community search has been studied in many existing

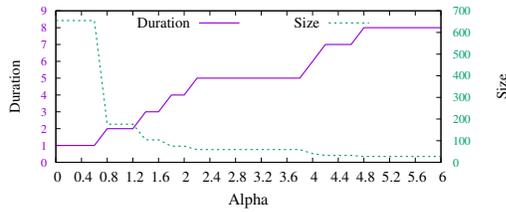


Figure 16: Community Size and Duration with Varying  $\alpha$

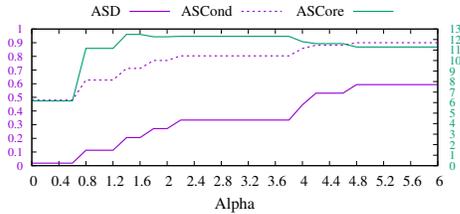


Figure 17: Community Quality with Varying  $\alpha$

works. In static networks, existing methods can be classified into two categories. The first method is based on random walk, which aims to assign scores to the vertex from the query vertex and identify the local community based on the scores. Wu et al. [43] used a single random walker, and Bian et al. [5] introduced multiple walkers to assign vertex scores based on the hitting probability. Bian [6] further proposed memory based multiple walker that records the entire visiting history and supports multiple local communities w.r.t. different query vertices simultaneously. Another method is based on capturing cohesiveness structures [17] such as  $k$ -truss [2, 16, 27],  $k$ -core [3, 11, 21] and  $k$ -clique [10, 39]. To deal with the changes of network data over time, Takaffoli et al. [41] explored local community mining in the dynamic social network by extending  $L$ -metric [8] to an incremental version. Luo et al. [31] divided the formation of the local community into three stages and designed different dynamical membership functions to construct the local community with better cohesiveness. DiTursi et al. [12] proposed PHASR method to detect local community in the dynamic networks, and Papadopoulos [33] expanded PHASR to fit the distributed processing standard of Apache Spark engine.

**K-core Community Search.** In this work, we consider the community size (number of vertices) and duration (continuity of the vertex engagement) as two important factors, so we use  $k$ -core model which is defined on vertex attributes rather than other classic models like  $k$ -truss that defined on edge attributes.  $K$ -core was firstly introduced by Seidman et al. [37] and becomes one of the most widely used measurements of graph cohesiveness. Seidman et al. [40] developed a greedy algorithm to discover the dense subgraph by iteratively removing the vertices with the minimum degrees. Batagelj et al. [4] proposed a linear core decomposition algorithm to compute the core number of all the vertices. Cui et al. [11] developed a local community search algorithm that starts from a query vertex  $q$  and spans iteratively to include the local optimal vertex into the community. Barbieri et al [3] proposed an index structure based on the nested feature of the core number

and improved the community search significantly. Following [3], Fang et al. [13] improved the efficiency of index construction. To include network dynamics, Li et al. [23] devised a core maintenance algorithm in large dynamic networks. Wu [42] proposed distributed algorithms based on the block-centric model to compute cores in the temporal graph. Galimberti et al. [14] identified all the maximal  $k$ -core with various time span and  $k$  value. Based on [14], Hung and Tseng [18] extended to the maximal lasting  $k$ -core. However, these works mainly focused on core decomposition in dynamic graphs and didn't consider the edge weight and query vertex like this work.

**Cohesive Subgraph Mining in Dynamic Networks.** Our work also relates to dense subgraph mining, which aims to identify the densely connected vertices in temporal or dynamic networks. Abdelhamid et al. [1] proposed an incremental approach called IncGM+ to extend the traditional Frequent Subgraph Mining (FSM) method into dynamic networks by only updating the "fringe" patterns. Ma et al. [32] proposed a fast computation algorithm to identify dense subgraphs in temporal graphs where edges have positive or negative weights. However, their method relies on the "evolving convergence phenomenon" that assumes weights of all edges are increasing or decreasing in the same direction which is too strict for the real world. And they didn't consider the community continuity. Semertzidis and Pitoura [38] proposed the problem of querying the frequent subgraph patterns in the directed dynamic networks and returns the top- $k$  durable matches. Liu et al. [28] considered the duration of the found dense subgraphs using an expectation-maximization method. They are unable to deal with edge weight, which is limited in real applications.

## 8 CONCLUSIONS AND FUTURE WORK

In this paper, we first discussed the necessity of reliable local community in dynamic networks and proposed the novel most reliable community search problem. Then, we developed an online  $(\theta, k)$ -core reliable community search approach by pruning the ineligible edges based on the given threshold and their lasting times. After that, we designed an effective  $WCF$ -Index to maintain the vertex candidates of  $(\theta, k)$ -core subgraphs, and developed an efficient index-based dynamic programming approach. Finally, the empirical evaluations on a variety of datasets and parameter settings illustrate the efficiency and effectiveness of the proposed approaches. In this work, we mainly focus on single query vertex situation. However, our proposed algorithms can be extended to support querying a set of vertices. For  $EEF$  algorithm, we can start the edge search from all query nodes simultaneously, and maintain a visited edge set to avoid repeat traverse. Then we can follow the  $EEF$  algorithm to return the valid local  $k$ -core. For  $WCF$ -Index query algorithm, multiple query nodes can be easily supported by filtering the intermediate result that does not contain all the query nodes. To explore the significance of multiple query nodes in community discovery, one potential research direction is to investigate the local engagement of the query nodes and identify the meaningful communities.

## ACKNOWLEDGMENTS

This work was mainly supported by the ARC Linkage Project under Grant No. LP180100750.

## REFERENCES

- [1] Ehab Abdelhamid, Mustafa Canim, Mohammad Sadoghi, Bishwaranjan Bhat-tacharjee, Yuan-Chi Chang, and Panos Kalnis. 2017. Incremental frequent sub-graph mining on large evolving graphs. *IEEE Transactions on Knowledge and Data Engineering* 29, 12 (2017), 2710–2723.
- [2] Esra Akbas and Peixiang Zhao. 2017. Truss-based community search: a truss-equivalence based indexing approach. *Proceedings of the VLDB Endowment* 10, 11 (2017), 1298–1309.
- [3] Nicola Barbieri, Francesco Bonchi, Edoardo Galimberti, and Francesco Gullo. 2015. Efficient and effective community search. *Data mining and knowledge discovery* 29, 5 (2015), 1406–1433.
- [4] Vladimir Batagelj and Matjaz Zaversnik. 2003. An  $O(m)$  algorithm for cores decomposition of networks. *arXiv preprint cs/0310049* (2003).
- [5] Yuchen Bian, Jingchao Ni, Wei Cheng, and Xiang Zhang. 2017. Many heads are better than one: Local community detection by the multi-walker chain. In *2017 IEEE International Conference on Data Mining (ICDM)*. IEEE, 21–30.
- [6] Yuchen Bian, Yaowei Yan, Wei Cheng, Wei Wang, Dongsheng Luo, and Xiang Zhang. 2018. On multi-query local community detection. In *2018 IEEE international conference on data mining (ICDM)*. IEEE, 9–18.
- [7] Zhan Bu, Zhiang Wu, Jie Cao, and Yichuan Jiang. 2015. Local community mining on distributed and dynamic networks from a multiagent perspective. *IEEE Transactions on cybernetics* 46, 4 (2015), 986–999.
- [8] Jiyang Chen, Osmar R Zaiane, and Randy Goebel. 2009. Detecting communities in large networks by iterative local expansion. In *2009 International Conference on Computational Aspects of Social Networks*. IEEE, 105–112.
- [9] Aaron Clauset. 2005. Finding local community structure in networks. *Physical review E* 72, 2 (2005), 026132.
- [10] Wanyun Cui, Yanghua Xiao, Haixun Wang, Yiqi Lu, and Wei Wang. 2013. Online search of overlapping communities. In *Proceedings of the 2013 ACM SIGMOD international conference on Management of data*. 277–288.
- [11] Wanyun Cui, Yanghua Xiao, Haixun Wang, and Wei Wang. 2014. Local search of communities in large graphs. In *Proceedings of the 2014 ACM SIGMOD international conference on Management of data*. 991–1002.
- [12] Daniel J DiTursi, Gaurav Ghosh, and Petko Bogdanov. 2017. Local community detection in dynamic networks. In *2017 IEEE International Conference on Data Mining (ICDM)*. IEEE, 847–852.
- [13] Yixiang Fang, Reynold Xiang, Siqiang Luo, and Jiafeng Hu. 2016. Effective community search for large attributed graphs. *Proceedings of the VLDB Endowment* 9, 12 (2016), 1233–1244.
- [14] Edoardo Galimberti, Alain Barrat, Francesco Bonchi, Ciro Cattuto, and Francesco Gullo. 2018. Mining (maximal) span-cores from temporal networks. In *Proceedings of the 27th ACM international Conference on Information and Knowledge Management*. 107–116.
- [15] Qiang-Sheng Hua, Yuliang Shi, Dongxiao Yu, Hai Jin, Jiguo Yu, Zhipen Cai, Xiuzhen Cheng, and Hanhua Chen. 2019. Faster parallel core maintenance algorithms in dynamic graphs. *IEEE Transactions on Parallel and Distributed Systems* 31, 6 (2019), 1287–1300.
- [16] Xin Huang, Hong Cheng, Lu Qin, Wentao Tian, and Jeffrey Xu Yu. 2014. Querying k-truss community in large and dynamic graphs. In *Proceedings of the 2014 ACM SIGMOD international conference on Management of data*. 1311–1322.
- [17] Xin Huang, Laks VS Lakshmanan, and Jianliang Xu. 2017. Community search over big graphs: Models, algorithms, and opportunities. In *2017 IEEE 33rd international conference on data engineering (ICDE)*. IEEE, 1451–1454.
- [18] Wei-Chun Hung and Chih-Ying Tseng. 2021. Maximum (L, K)-Lasting Cores in Temporal Social Networks. In *International Conference on Database Systems for Advanced Applications*. Springer, 336–352.
- [19] Roberto Interdonato and Andrea Tagarelli. 2017. Personalized recommendation of points-of-interest based on multilayer local community detection. In *International Conference on Social Informatics*. Springer, 552–571.
- [20] Chaoyi Li and Yangsen Zhang. 2020. A personalized recommendation algorithm based on large-scale real micro-blog data. *Neural Computing and Applications* 32, 15 (2020), 11245–11252.
- [21] Rong-Hua Li, Lu Qin, Jeffrey Xu Yu, and Rui Mao. 2015. Influential community search in large networks. *Proceedings of the VLDB Endowment* 8, 5 (2015), 509–520.
- [22] Rong-Hua Li, Jiao Su, Lu Qin, Jeffrey Xu Yu, and Qiangqiang Dai. 2018. Persistent community search in temporal networks. In *2018 IEEE 34th International Conference on Data Engineering (ICDE)*. IEEE, 797–808.
- [23] Rong-Hua Li, Jeffrey Xu Yu, and Rui Mao. 2013. Efficient core maintenance in large dynamic graphs. *IEEE Transactions on Knowledge and Data Engineering* 26, 10 (2013), 2453–2465.
- [24] Yuan Li, Jinsheng Liu, Huiqun Zhao, Jing Sun, Yuhai Zhao, and Guoren Wang. 2021. Efficient continual cohesive subgraph search in large temporal graphs. *World Wide Web* 24, 5 (2021), 1483–1509.
- [25] Longlong Lin, Pingpeng Yuan, Rong-Hua Li, Jifei Wang, Ling Liu, and Hai Jin. 2021. Mining stable quasi-cliques on temporal networks. *IEEE Transactions on Systems, Man, and Cybernetics: Systems* 52, 6 (2021), 3731–3745.
- [26] Bin Liu and Feiteng Zhang. 2020. Incremental algorithms of the core maintenance problem on edge-weighted graphs. *IEEE Access* 8 (2020), 63872–63884.
- [27] Qing Liu, Minjun Zhao, Xin Huang, Jianliang Xu, and Yunjun Gao. 2020. Truss-based community search over large directed graphs. In *Proceedings of the 2020 ACM SIGMOD International Conference on Management of Data*. 2183–2197.
- [28] Xuanming Liu, Tingjian Ge, and Yinghui Wu. 2019. Finding densest lasting subgraphs in dynamic graphs: A stochastic approach. In *2019 IEEE 35th International Conference on Data Engineering (ICDE)*. IEEE, 782–793.
- [29] Dongsheng Luo, Yuchen Bian, Yaowei Yan, Xiao Liu, Jun Huan, and Xiang Zhang. 2020. Local community detection in multiple networks. In *Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining*. 266–274.
- [30] Feng Luo, James Z Wang, and Eric Promislow. 2008. Exploring local community structures in large networks. *Web Intelligence and Agent Systems: An International Journal* 6, 4 (2008), 387–400.
- [31] Wenjian Luo, Daofu Zhang, Hao Jiang, Li Ni, and Yamin Hu. 2018. Local community detection with the dynamic membership function. *IEEE Transactions on Fuzzy Systems* 26, 5 (2018), 3136–3150.
- [32] Shuai Ma, Renjun Hu, Luoshu Wang, Xuelian Lin, and Jinpeng Huai. 2017. Fast computation of dense temporal subgraphs. In *2017 IEEE 33rd International Conference on Data Engineering (ICDE)*. IEEE, 361–372.
- [33] Apostolos N. Papadopoulos and Georgios Tzortzidis. 2020. Distributed Time-Based Local Community Detection. In *24th Pan-Hellenic Conference on Informatics*. 390–393.
- [34] Oun Joung Park, Xinran Y Lehto, and Alastair M Morrison. 2008. Collaboration between CVB and local community in destination marketing: CVB executives' perspective. *Journal of Hospitality & Leisure Marketing* 17, 3-4 (2008), 395–417.
- [35] Hongchao Qin, Rong-Hua Li, Guoren Wang, Xin Huang, Ye Yuan, and Jeffrey Xu Yu. 2022. Mining Stable Communities in Temporal Networks by Density-Based Clustering. *IEEE Transactions on Big Data* 8, 3 (2022), 671–684.
- [36] Ahmet Erdem Saryüce, Buğra Gedik, Gabriela Jacques-Silva, Kun-Lung Wu, and Ümit V Çatalyürek. 2016. Incremental k-core decomposition: algorithms and evaluation. *The VLDB Journal* 25, 3 (2016), 425–447.
- [37] Stephen B Seidman. 1983. Network structure and minimum degree. *Social networks* 5, 3 (1983), 269–287.
- [38] Konstantinos Semertzidis and Evaggelia Pitoura. 2018. Top-k Durable Graph Pattern Queries on Temporal Graphs. *IEEE Transactions on Knowledge and Data Engineering* 31, 1 (2018), 181–194.
- [39] Jing Shan, Derong Shen, Tiezheng Nie, Yue Kou, and Ge Yu. 2016. Searching overlapping communities for group query. *World Wide Web* 19, 6 (2016), 1179–1202.
- [40] Mauro Sozio and Aristides Gionis. 2010. The community-search problem and how to plan a successful cocktail party. In *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*. 939–948.
- [41] Mansoureh Takaffoli, Reihaneh Rabbany, and Osmar R Zaiane. 2013. Incremental local community identification in dynamic social networks. In *2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2013)*. IEEE, 90–94.
- [42] Huanhuan Wu, James Cheng, Yi Lu, Yiping Ke, Yuzhen Huang, Da Yan, and Hejun Wu. 2015. Core decomposition in large temporal graphs. In *2015 IEEE International Conference on Big Data (Big Data)*. IEEE, 649–658.
- [43] Yubao Wu, Ruoming Jin, Jing Li, and Xiang Zhang. 2015. Robust local community detection: on free rider effect and its elimination. *Proceedings of the VLDB Endowment* 8, 7 (2015), 798–809.
- [44] Bohua Yang, Dong Wen, Lu Qin, Ying Zhang, Lijun Chang, and Rong-Hua Li. 2019. Index-based optimal algorithm for computing k-cores in large uncertain graphs. In *2019 IEEE 35th International Conference on Data Engineering (ICDE)*. IEEE, 64–75.
- [45] Yikai Zhang, Jeffrey Xu Yu, Ying Zhang, and Lu Qin. 2017. A fast order-based approach for core maintenance. In *2017 IEEE 33rd International Conference on Data Engineering (ICDE)*. IEEE, 337–348.