Seasonal-Periodic Subgraph Mining in Temporal Networks

Qianzhen Zhang[1], Deke Guo[1], Xiang Zhao[1], Xinyi Li[1], Xi Wang[1]
[1]Science and Technology on Information Systems Engineering Laboratory, National University of Defense Technology, Changsha, China
Contact authors: *Deke Guo and Xiang Zhao*, Emails: {dekeguo, xiangzhao}@nudt.edu.cn

ABSTRACT

Seasonal periodicity is a frequent phenomenon for social interactions in temporal networks. A key property of this behavior is that it exhibits periodicity for multiple particular periods in temporal networks. Mining such *seasonal-periodic patterns* is significant since it can indicate interesting relationships between the individuals involved in the interactions. Unfortunately, most previous studies for periodic pattern mining ignore the *seasonal feature*. This motivates us to explore mining seasonal-periodic subgraphs, and the investigation presents a novel model, called maximal σ-periodic ω-seasonal k-subgraph. It represents a subgraph with size larger than k and that appears at least σ times periodically in at least ω particular periods on the temporal graph. Since seasonal-periodic patterns do not satisfy the anti-monotonic property, we propose a weak version of support measure with an anti-monotonic property to reduce the search space efficiently. Then, we present an effective mining algorithm to seek all maximal σ-periodic ω-seasonal k-subgraphs. Experimental results on real-life datasets show the effectiveness and efficiency of our approach.

CCS CONCEPTS

• **Mathematics of computing → Graph algorithms**; • **Information systems → Data mining**; • **Theory of computation → Graph algorithms analysis**.

KEYWORDS

seasonal-periodic pattern; subgraph mining; temporal graph

ACM Reference Format:
Qianzhen Zhang[1], Deke Guo[1], Xiang Zhao[1], Xinyi Li[1], Xi Wang[1]. 2020. Seasonal-Periodic Subgraph Mining in Temporal Networks. In *Proceedings of the 29th ACM International Conference on Information and Knowledge Management (CIKM '20), October 19–23, 2020, Virtual Event, Ireland*. ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3340531.3412091

1 INTRODUCTION

A temporal network is defined by a graph, where each edge is associated with temporal information. For example, Figure 1 gives a sample retweet graph. The edge $\langle v_3, v_5, \{3,4\}\rangle$ denotes that student v_3 retweets v_5 at timestamps 3 and 4. Other examples of temporal networks include face-to-face contact network, email communication network, scientific collaboration network, et cetera.

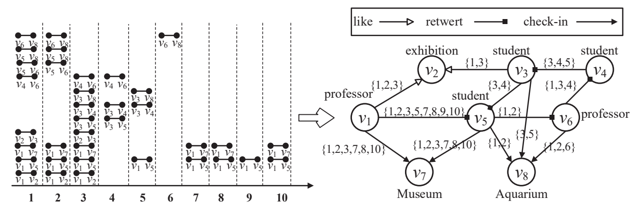


Figure 1: Sample temporal social graph

Seasonal periodicity is a frequently happening phenomenon for social interactions in temporal networks, e.g., weekly twitter retweets during summer and winter vacations. Mining such *seasonal-periodic* patterns is of great significance for many real-world applications since they often provide useful information pertaining seasonal or temporal associations and can indicate interesting relationships between the individuals involved in the interactions.

Motivating Example. In a scientific collaboration network (e.g., DBLP), each edge represents that two authors coauthored papers at corresponding timestamps. In this network, one may be interested in determining a short-term periodic collaboration relationship among some researchers, and the collaboration relationship will appear multiple times in different periods.

Although mining seasonal-periodic patterns is of great importance, previous work do not consider the *seasonal feature* (e.g., behaviors that appear in multiple periods) that exists in the temporal networks. Lahiri et al. [3] first investigated the problem of mining periodic subgraphs in dynamic social networks and analyzed the computational complexity of enumerating all periodic subgraphs. Radinsky et al. [7] developed a temporal model to predict the periodic actions. Qin et al. [6] proposed a σ-periodic k-clique model to find the cliques with size larger than k that appears at least σ times periodically in the temporal graph. Since the algorithms and models of these studies are designed for mining periodic behaviors in one period, they cannot discover the seasonal-periodic subgraphs.

In this work, we investigate a novel model, called maximal σ-periodic ω-seasonal k-subgraph, to characterize patterns that appear at least σ times periodically in at least ω periods with size larger than k on the temporal graph. We show that the traditional periodic subgraphs mining problem is a special case of our model when we set $\omega = 1$. Mining seasonal-periodical subgraphs is especially challenging since they do not satisfy the anti-monotonic property. Having an anti-monotonic support measure can allow the development of methods that effectively prune the search space; without an anti-monotonic measure exhaustive search is unavoidable [1]. To address this issue, we propose an efficient pruning strategy to reduce the search space while finding seasonal-periodic subgraphs in temporal networks.

CIKM '20, October 19–23, 2020, Virtual Event, Ireland

ACM ISBN 978-1-4503-6859-9/20/10...$15.00
https://doi.org/10.1145/3340531.3412091

Contributions. In short, we are the first to study the seasonal-periodic mining problem in temporal networks, and the major contributions we have made are summarized below: (1) We propose a novel maximal σ-periodic ω-seasonal k-subgraph model to characterize the seasonal-periodical patterns in temporal networks. (2) We define a *weak version of support* measure with anti-monotonic property, which can help to reduce the search space in the mining process. (3) We develop an effective algorithm for discovering maximal σ-periodic ω-seasonal k-subgraphs.

Comprehensive empirical study verifies the efficiency and effectiveness of the proposed algorithms and techniques.

Related Work. The studies of behavior mining in temporal networks are related to our work. Besides the aforementioned relevant research [3, 6, 7], Ma et al. [5] devised a dense subgraph mining algorithm to identify cohesive subgraphs in a temporal network. Li et al. [4] addressed the problem of mining periodic behaviors for moving objects. Kurashima et al. [2] modeled the periodic actions in real-world (e.g., eating, sleep, and exercise) to make predictions for future actions. Their work, however, does not consider the periodic or seasonal feature of subgraph, thus cannot be used for mining seasonal-periodic subgraphs.

2 PROBLEM FORMULATION

A temporal graph $\mathcal{G}_T$ is defined as $(\mathcal{V}, \mathcal{E})$ such that $\mathcal{V}$ is a set of vertices, $\mathcal{E}$ is a set of temporal edges. Each temporal edge $e = \langle v, v', \mathcal{T} \rangle$ encodes a link between v and v' that exists interaction in $\mathcal{T}$, where v, v' are vertices in $\mathcal{V}$, and $\mathcal{T}$ is a time window (a sequence of consecutive timestamps). We assume that each $t \in \mathcal{T}$ is an integer, because the timestamp is an integer in practice. As defined in [6], the *de-temporal* graph of $\mathcal{G}_T$ denoted by $G = (V, E)$ is a graph that ignores all the timestamps associated with the temporal edges, where $V = \mathcal{V}$ and $E = \{\langle v, v' \rangle | \langle v, v', \mathcal{T} \rangle \in \mathcal{E}\}$. A graph $S = \{V_S, E_S\}$ is referred to as an induced subgraph of G iff $V_S \in V$ and $E_S \in E$. For convenience, we use the notion $S \subseteq G$ ($S \subset G$ if $S \neq G$) to represent that S is a subgraph of G.

Given a timestamp $t_i \in \mathcal{T}$, a snapshot G_i of $\mathcal{G}_T$ at t_i is a graph induced by the set of all the edges associated with timestamp t_i. In the rest of this paper, we assume without loss of generality that all the timestamps are sorted in a chronological order, i.e., $t_1 < t_2 < \cdots < t_{|\mathcal{T}|}$.

Definition 2.1 (Time support set). Given a temporal graph $\mathcal{G}_T$ and a subgraph S, the time *support* set $T(S)$ is the set of all timestamps t_i such that $S \subseteq G_i$, where G_i is the snapshot of $\mathcal{G}_T$ at t_i.

Definition 2.2 (σ-periodic time support set). Given a temporal graph $\mathcal{G}_T$ and a parameter σ, a σ-periodic support set of a subgraph S, denoted by $\pi_\sigma(S)$, is a **maximal** continuous subset of $T(S)$ such that (1) $\pi_\sigma(S) = \{t_i, t_{i+1}, \cdots, t_j\}$ where $j - i \geq \sigma - 1$, (2) $t_{k+1} - t_k \leq p$ ($k \in [i, j-1]$) for a user-defined period threshold value p, and (3) there exists no other timestamp in which an inter-arrival time between t_i (resp. t_j) is no more than p.

Note that, our definition of *σ-periodic time support set* is different from the one in [6] from two aspects: (1) we consider a continuous subset in $T(S)$ rather than an arbitrary subset; and (2) we claim that an inter-arrival time of the subgraph S is periodic if it is no more than a user-defined period threshold rather than a fixed value (i.e., $t_{k+1} - t_k = p$). For example, consider a temporal graph in Figure 1. For the subgraph $S = \{\langle v_1, v_7 \rangle, \langle v_1, v_5 \rangle, \langle v_5, v_7 \rangle\}$, the time support set of S is $\{1, 2, 3, 7, 8, 10\}$. By Definition 2.2, the set $\{7, 8, 10\}$ is a 3-periodic time support set of S if we set $p = 2$. However, if we use the definition in [6], we will miss above periodic time support set.

According to Definition 2.2, we can discover more interesting behaviors pertaining to those rare patterns that have exhibited sufficient number of periodicity in a portion of the temporal graph. What's more, given a time support set $T(S)$ of a subgraph S, there may be more than one σ-periodic time support set in $T(S)$ to represent the ***seasonal feature*** of S.

Definition 2.3 (σ-periodic ω-seasonal support set). Given a temporal graph $\mathcal{G}_T$ and a parameter ω, the σ-periodic ω-seasonal support set, defined as $\Omega^\sigma_\omega(S)$, is a subset of $T(S)$ such that (1) $\Omega^\sigma_\omega(S) = \{\pi^1_\sigma(S), \pi^2_\sigma(S), \cdots, \pi^n_\sigma(S)\}$ where $n \geq \omega$; and (2) each $\pi^i_\sigma(S)$ ($i \in [1, n]$) is a σ-periodic time support set.

Definition 2.4 (σ-periodic ω-seasonal subgraph). Given a temporal graph $\mathcal{G}_T$, it's de-temporal G, and two parameters σ, ω, a subgraph S of G is called a σ-periodic ω-seasonal subgraph if there exists a σ-periodic ω-seasonal support set $\Omega^\sigma_\omega(S)$ for S.

Note that, many σ-periodic ω-seasonal subgraphs are small and may not be interesting to the users. Therefore, it is desirable to find large σ-periodic ω-seasonal subgraphs for practical applications. Then we propose a novel model, namely, maximal σ-periodic ω-seasonal k-subgraph, which is defined as follows.

Definition 2.5 (maximal σ-periodic ω-seasonal k-subgraph). A maximal σ-periodic ω-seasonal k-subgraph S is a subgraph of the de-temporal graph G such that (1) S is a σ-periodic ω-seasonal subgraph with $|S|$ [1] $> k$; and (2) there exists no other σ-periodic ω-seasonal k-subgraph S' such that S is a subgraph of S' via subgraph isomorphism [2], denoted as $S \subset S'$.

For brevity, in the rest of this paper, the maximal σ-periodic ω-seasonal k-subgraph is abbreviated as **MSPs**.

Problem Statement. Given a temporal graph $\mathcal{G}_T$, and parameters p, σ and ω, the goal of MSPs mining problem is to enumerate all the maximal σ-periodic ω-seasonal k-subgraphs in $\mathcal{G}_T$.

3 METHODOLOGY

We detect seasonal-periodical subgraphs by explicitly using the maximal σ-periodic ω-seasonal k-subgraph model. To this end, we introduce correlation measures for MSPs (Section 3.1), followed by the description of the MSPs discovery algorithm (Section 3.2).

3.1 Support and Variant

Recall that an anti-monotonicity of support measure is of crucial importance for finding user-interest-based subgraphs. The most intuitive support measure for a subgraph S is to count its number of σ-periodic time support sets in $T(S)$, denoted as $\mathrm{supp}(S)$. Intuitively, S is a σ-periodic ω-seasonal subgraph if $\mathrm{supp}(S) \geq \omega$. Unfortunately, such a measure is not anti-monotone since there are cases where a subgraph has less σ-periodic time support sets than its extension.

[1] The number of de-temporal edges of the subgraph S

[2] https://en.wikipedia.org/wiki/Subgraph isomorphism problem

Example 3.1. Consider a temporal graph in Figure 1. Suppose that $\sigma = 3$, $p = 2$. For the subgraph $S_1 = \{\langle v_1, v_5\rangle\}$, $\text{supp}(S_1) = 1$ since there exists only one σ-periodic time support set {1,2,3,5,7,8,9,10}. For the subgraph $S_2 = \{\langle v_1, v_7\rangle, \langle v_1, v_5\rangle, \langle v_5, v_7\rangle\}$, $\text{supp}(S_2) = 2$ since there exists two σ-periodic time support sets {1,2,3} and {7,8,10}. As a result, $\text{supp}(S)$ does not satisfy the anti-monotonicity.

Weak version of support. To address the above issue, we propose $\text{supp}^*(S)$, which is a weak version of $\text{supp}(S)$. In detail, $\text{supp}^*(S)$ is defined as follows.

$$\text{supp}^*(S) = \sum_{i=1}^{\text{supp}(S)} \left\lfloor \frac{|\pi_\sigma^i(S)|}{\sigma} \right\rfloor, \tag{1}$$

where $|\pi_\sigma^i(S)|$ denotes the number of timestamps in the i-th σ-periodic time support set in $T(S)$. Note that, $\text{supp}^*(S) \geq \text{supp}(S)$ always holds since $\left\lfloor \frac{\pi_\sigma^i(S)}{\sigma} \right\rfloor \geq 1$. Thus, $\text{supp}^*(S)$ can be set as the upper bound of $\text{supp}(S)$.

We show that $\text{supp}^*(S)$ is anti-monotonic.

LEMMA 3.2. *For any temporal graph $\mathcal{G}_T$, and any subgraph S_1 and S_2, $supp^*(S_1) \geq supp^*(S_2)$ if $S_1 \subseteq S_2$.*

Example 3.3. Continue to Example 3.1. For the subgraph S_1, $\text{supp}^*(S_1) = 2$; and for the subgraph S_2, $\text{supp}^*(S_2) = 2$. Thus, we have $\text{supp}^*(S_1) = \text{supp}^*(S_2)$ when $S_1 \subseteq S_2$.

3.2 Discovery Algorithm

Given a temporal graph $\mathcal{G}_T$ and a subgraph S, we will verify whether S is a σ-periodic ω-seasonal subgraph by checking $\text{supp}^*(S)$ as follows. First, we have the following observation:

- If the support of S satisfies $\text{supp}(S) \geq \omega$, there must be a $\text{supp}^*(S)$ such that $\text{supp}^*(S) \geq \omega$.

Consequently, if such a weak version of support does not exist, the subgraph S cannot be σ-periodic ω-seasonal subgraph because a weak version of support is a necessary condition for a support.

From the above observation, we design a MSPs mining algorithm, namely, IsMSPs, to enumerate all maximal σ-periodic ω-seasonal k-subgraphs.

Overview. Algorithm IsMSPs "cold-starts" MSPs discovery by initializing a MSPs *subgraph generation tree* SGT with valid subgraphs (i.e., with weak version of support greater or equal to ω) that carry a single-edge pattern. SGT is then constructed by expanding a parent subgraph with one neighbor at a time. At each level i, it discovers and stores all applicable extensions of size i that have not been previously considered in *candidateSet*. To exclude already generated extensions, we adopt the *DFScode* canonical form as in GSPAN [8]. Then, IsMSPs verifies the members in *candidateSet*, and fills the level-i part of tree SGT. It works as follows.

Subgraph verification. For each candidate member S_i in level-i ($i \leq k+1$) part of SGT, algorithm IsMSPs eliminates S_i if $\text{supp}^*(S_i) < \omega$ since according to the anti-monotonic property, its extensions including itself cannot be MSPs. For each candidate member S_j in level-j ($j > k+1$), we use an expansion flag f_{S_j} to indicate whether S_j can be further expanded. f_{S_j} is set to **false** if $\text{supp}^*(S_j) < \omega$ and will be removed from *candidateSet*, else f_{S_j} is set to **true**. The expansion flag will be used to check whether a subgraph is a MSPs.

Next, IsMSPs is recursively executed to further extend each subgraph in the *candidateSet*. The expansion process terminates when $candidateSet = \emptyset$. Finally, IsMSPs enumerates all the maximal σ-periodic ω-seasonal k-subgraphs as follows.

Bottom-up enumeration. Recall that each node $v \in$ SGT at level-i stores a subgraph with i edges. Firstly, for each leaf node of SGT, if the expansion flag of the subgraph stored in it is set to **false**, we mark it as visited. Then, we process nodes of SGT level-by-level in a bottom-up fashion till level-$(k+1)$. When processing a node v at level-i ($i > k$), let $v.N$ denote the number of visited neighbors of v, $v.C$ denote the number of child nodes of v in SGT, S_i denote the subgraph stored in v. Here, we say the subgraph S_i is a MSPs if $v.N = v.C$ and $\text{supp}(S_i) \geq \omega$. To save search space, if S_i is a MSPs, all the ancestors (i.e., parent, parent of parent, $\cdots$) in SGT of v will not be processed anymore. What's more, if $v.N = v.C$ and $\text{supp}(S_i) < \omega$, we mark v as visited and set $f_{S_i} = $ **false**.

4 EXPERIMENTS

In our experiments, we implement two algorithms to identify maximal σ-periodic ω-seasonal k-subgraphs: IsMSPs and IsMSPs-B. Specially, algorithm IsMSPs-B is a baseline algorithm that does not integrated any pruning techniques (Section 3.1). All algorithms are implemented in C++. All the experiments are conducted on a PC with an Intel i7 3.50GHz CPU and 32GB memory.

Table 1: Graph datasets

Dataset	$\lvert V\rvert$	$\lvert E\rvert$	$\lvert\mathcal{T}\rvert$	Time scale
HS	327	5,818	101	hour
LKML	26,885	159,996	96	month
DBLP	1,729,816	8,546,306	59	year

Datasets. We used three real-life temporal networks in the experiments. HS [3] is a temporal network of face-to-face contacts between students in a French high school; LKML [4] is a temporal email communication network of users; and DBLP [5] is a temporal collaboration network of authors. The detailed statistics of our datasets are summarized in Table 1.

Parameter settings. There are four parameters k, p, σ and ω in our algorithms. For the parameter k, we vary it from 3 to 5 with a default value of 3. For the parameter p, we vary it from 2 to 4 with a default value of 3. We also vary σ from 3 to 5 with a default value of 3 and vary ω from 1 to 3 with a default of 2. Unless otherwise specified, the value of the other parameter are set to its default value when varying a parameter.

4.1 Efficiency Test

Efficiency of MSPs mining. Figure 2 shows the processing time of IsMSPs-B and IsMSPs on different datasets with parameters $k = 4$, $p = 3$, $\sigma = 4$ and $\omega = 2$. Similar results can also be observed under the other parameter settings. Note that, IsMSPs is much faster than IsMSPs-B on all datasets. Especially, IsMSPs outperforms IsMSPs-B by up to 304.71 times on DBLP. These results indicate that our proposed weak version of support measure can help shrink the search space dramatically as the mining process progresses.

[3] http://www.sociopatterns.org/datasets/
[4] http://konect.uni-koblenz.de
[5] http://dblp.uni-trier.de/xml/

Table 2: Running time of IsMSPs with varying parameters (LKML)

Algorithms	k	p	$\omega = 1$			$\omega = 2$			$\omega = 3$		
			$\sigma = 3$	$\sigma = 4$	$\sigma = 5$	$\sigma = 3$	$\sigma = 4$	$\sigma = 5$	$\sigma = 3$	$\sigma = 4$	$\sigma = 5$
IsMSPs	3	2	264	95	79	132	78	64	82	47	31
		3	706	631	473	523	452	311	208	172	109
		4	954	804	662	681	761	517	291	233	182
	4	2	152	63	46	91	51	32	74	28	17
		3	547	352	236	316	213	174	102	32	20
		4	602	393	295	397	325	242	117	38	24
	5	2	93	44	21	52	31	10	21	8	3
		3	152	81	27	137	73	13	30	11	6
		4	214	95	30	181	87	19	35	13	8

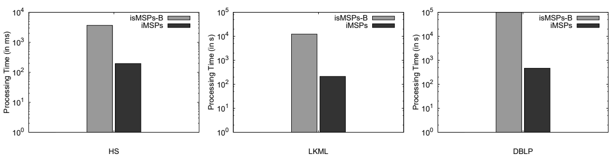


Figure 2: Processing time on various datasets

Effect of parameters. Table 2 gives the running time of IsMSPs with varying parameters on LKML. Similar results can also be observed on other datasets. Specially, the running time of IsMSPs decreases as k, σ or ω increases; and the running time increases as p increases. The reason could be that for a smaller σ, ω and for a larger p, the potential candidates in *candidateSet* will be larger and the construction of SGT will consume more time. Further, in the bottom-up enumeration process, a larger k will decrease the search space, and then reduce the running time correspondingly.

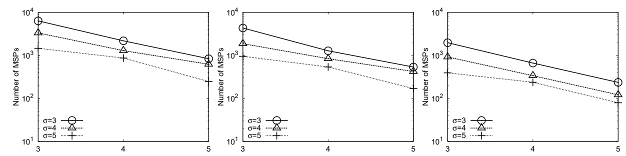
Scalability test. In this set of experiments, we measure the s-calability of IsMSPs on DBLP. Similar results can also be observed on the other datasets. We generate four temporal subgraphs by randomly picking 20%-80% of the de-temporal edges, and evaluate the running time of IsMSPs on these subgraphs. In general, the processing time increases smoothly in the increasing size of the dataset. The scalability suggests that IsMSPs is scalable when handling large temporal networks. The figures are omitted in the interest of space.

4.2 Effectiveness Test

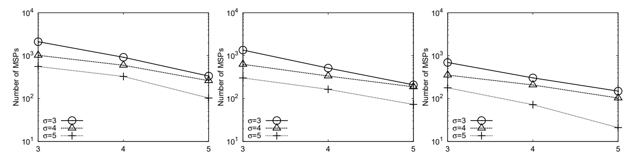
Number of MSPs with varying parameters. Figure 3 shows the number of MSPs with different parameters on LKML. The results on the other datasets are consistent. Especially, we observe that (1) the increase of k can decrease the number of MSPs at a fixed ω, p and σ; (2) the increase of ω can decrease the number of MSPs at a fixed k, p and σ; (3) the increase of σ can decrease the number of MSPs at a fixed k, p and ω; and (4) the decrease of p can decrease the number of MSPs at a fixed k, σ and ω. This is because with a larger k, σ, ω or with a smaller p, the seasonal-periodic subgraph constraint will be strong, thus the number of MSPs decreases.

5 CONCLUSION

In this paper, we have investigated a novel model, i.e., MSPs, to characterize the seasonal-periodical patterns in temporal networks. To find all MSPs, we first present a weak version of support measure that has an anti-monotonic property to reduce the search space. Then, we propose an effective enumeration algorithm to identify



(a) Vary k ($\omega = 1, p = 4$) (b) Vary k ($\omega = 2, p = 4$) (c) Vary k ($\omega = 3, p = 4$)



(d) Vary k ($\omega = 1, p = 2$) (e) Vary k ($\omega = 2, p = 2$) (f) Vary k ($\omega = 3, p = 2$)

Figure 3: MSPs discovered on LKML dataset

all MSPs. Comprehensive experiments demonstrate the efficiency, scalability and effectiveness of our algorithm.

ACKNOWLEDGMENTS

This work is partially supported by National Natural Science Foundation of China under Grant No.61872446, Natural Science Foundation of Hunan Province under Grant No.2019JJ20024, National key research and development program under Grant Nos. 2018YFB1800203 and 2018YFE0207600.

REFERENCES

[1] Mathias Fiedler and Christian Borgelt. 2007. Subgraph Support in a Single Large Graph. In *Workshops Proceedings of ICDM, October 28-31, 2007, USA*. 399–404. https://doi.org/10.1109/ICDMW.2007.74

[2] Takeshi Kurashima, Tim Althoff, and Jure Leskovec. 2018. Modeling Interdependent and Periodic Real-World Action Sequences. In *WWW, France, April 23-27*. 803–812. https://doi.org/10.1145/3178876.3186161

[3] Mayank Lahiri and Tanya Y. Berger-Wolf. 2010. Periodic subgraph mining in dynamic networks. *Knowl. Inf. Syst.* 24, 3 (2010), 467–497. https://doi.org/10.1007/s10115-009-0253-8

[4] Zhenhui Li, Bolin Ding, Jiawei Han, Roland Kays, and Peter Nye. 2010. Mining periodic behaviors for moving objects. In *SIGKDD, USA, July 25-28, 2010*. 1099–1108. https://doi.org/10.1145/1835804.1835942

[5] Shuai Ma, Renjun Hu, Luoshu Wang, Xuelian Lin, and Jinpeng Huai. 2017. Fast Computation of Dense Temporal Subgraphs. In *ICDE, USA, April 19-22, 2017*. 361–372. https://doi.org/10.1109/ICDE.2017.95

[6] Hongchao Qin, Rong-Hua Li, Guoren Wang, Lu Qin, Yurong Cheng, and Ye Yuan. 2019. Mining Periodic Cliques in Temporal Networks. In *ICDE, China, April 8-11, 2019*. 1130–1141. https://doi.org/10.1109/ICDE.2019.00104

[7] Kira Radinsky, Krysta M. Svore, Susan T. Dumais, Jaime Teevan, Alex Bocharov, and Eric Horvitz. 2012. Modeling and predicting behavioral dynamics on the web. In *WWW*. https://doi.org/10.1145/2187836.2187918

[8] Xifeng Yan and Jiawei Han. 2002. gSpan: Graph-Based Substructure Pattern Mining. In *ICDM, 9-12 December, Maebashi City, Japan*. 721–724. https://doi.org/10.1109/ICDM.2002.1184038