文章编号：2095-2163(2021)08-0189-05 中图分类号：TP301.6 文献标志码：A

# 时态图上图模式匹配研究综述

李发明，邹兆年，李建中
（哈尔滨工业大学 计算学部，哈尔滨 150001）

**摘 要：**图数据模型是一种通用且已经得到广泛应用的可以处理数据对象间复杂关系的数据模型。然而，现有的大多数研究只关注静态图数据的结构或者顶点和边上的属性，忽略了现实中数据的一个重要特征即时态信息。忽略了时态信息将导致错过很多的有价值的信息，甚至得到错误的信息。作为图研究领域中重要的研究内容之一，图模式匹配问题的研究也需要考虑时态信息。考虑到图模式匹配研究的重要性以及时态信息对数据的重要性，本文根据时态图的快照模型、边流模型和区间模型以及时态图数据的时序性、持续性和演化性对时态图上图模式匹配问题进行了全面地综述，并总结了现有工作的不足。

**关键词：**时态信息；时态图；图模式匹配

## Graph pattern matching on temporal graphs: A survey

LI Faming, ZOU Zhaonian, LI Jianzhong
(Faculty of Computing, Harbin Institute of Technology, Harbin 150001, China)

**【Abstract】** Graph data model is a universal and widely used data model, which can process the complex relationship between data objects. However, most of the existing works on static graphs only focus on the structure or properties of vertices and edges and ignore one important characteristic of data in real life-temporal information. It leads to a lot of valuable knowledge being missed or gets wrong results without considering the temporal information. As an important topic in the graph research, the graph pattern matching problem should take temporal information into account as well. Considering the importance of graph pattern matching and the importance of temporal information to the data, this paper thoroughly surveys graph pattern matching on temporal graphs based on the snapshot model, edge-stream model and interval model of the temporal graph, and the temporality, the evolvability and the durability of the temporal data. Finally, the shortcomings of existing works are summarized.

**【Key words】** temporal information; temporal graph; graph pattern matching

## 0 引 言

在众多图研究问题中，图模式匹配（graph pattern matching）问题一直占据着重要的地位。现有的研究一般根据子图同构（subgraph isomorphism）定义图模式匹配问题[1]。给定一个查询模式图 $Q$ 和数据图 $G$，图模式匹配问题是在 $G$ 中查找 $Q$ 的所有匹配。$Q$ 的一个匹配是满足如下条件的 $G$ 的一个子图 $H$：存在一个从 $Q$ 的顶点集到 $H$ 的顶点集的双射函数（bijective function），使得当且仅当 $(f(v), f(v'))$ 是 $H$ 中的一条边时，$(v,v')$ 是 $Q$ 中的一条边。如果图中顶点存在标签，则要求 $Q$ 中顶点 $v$ 的标签同 $H$ 中顶点 $f(v)$ 的标签相同。$H$ 则称为 $Q$ 在 $G$ 中的一个匹配。图模式匹配问题是很多研究的基础，例如，图数据库、知识图谱查询处理、图挖掘、计算机视觉等等。然而，现有的图模式匹配研究主要关注查询模式图的结构，常常忽略了图数据上的时态图信息。下面两个实际例子说明了时态信息在图模式匹配问题中的重要性。

（1）美国通讯公司 Verizon 每年都会公布安全事故报告，而这些安全事故中的攻击模式都带有时态信息，即这些模式都可以表示成带有时态信息的图模式。图 1 中给出了其中一种攻击模式，图中顶点表示服务器或者被攻击的终端，顶点之间的边表示服务器和终端之间的通信，边上的 $t$ 表示通信时间，图模式对通信时间要求是 $t_1 < t_2 < t_3 < t_4 < t_5$。监测这种常见的攻击模式将有利于识别恶意软件及其服务器。

（2）图 2 给出 3 个科研人员以合作的模式在同一个会议上发表论文的情况，其中顶点表示研究人员，顶点之间的边表示合作关系，图下面的文字表示会议的名称及合作的时间。了解不同科研人员之间的合作模式及持续时间将更好地发现研究团队及研究方向。

**基金项目：**国家自然科学基金重点子课题（61532015）；国家自然科学基金面上项目（61672189）。

**作者简介：**李发明（1988-），男，博士研究生，主要研究方向：图数据库、图数据挖掘；邹兆年（1979-），男，博士，教授，主要研究方向：图数据分析和数据库系统；李建中（1950-），男，教授，主要研究方向：数据库系统、无线传感器网络和数据质量。

**收稿日期：**2021-06-05

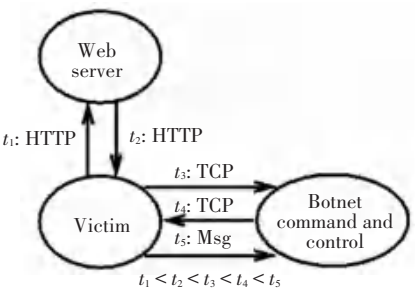


**图 1 攻击模式**

**Fig. 1 Cyber-attach pattern**

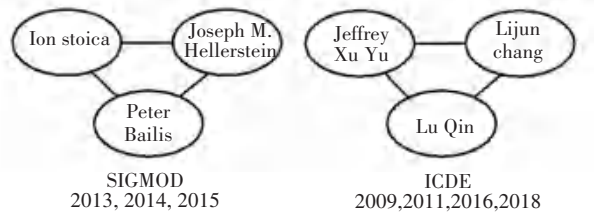


**图 2 长期合作模式**

**Fig. 2 Durable cooperation**

鉴于时态信息对于图数据查询、分析的重要性，而现有关于图模式匹配的研究又很少考虑时态信息，本文从时态图的不同模型和时态信息的不同特性出发，对时态图上的图模式匹配问题进行了全面地综述。

## 1 常见时态图模型

时态图数据（temporal graph data），也称为演化图数据（evolving graph data）、历史图数据（historical graph data），是在静态图数据基础上演变的一种包含时态信息的新型图数据[2]。时态图中顶点、边、顶点上的属性、边上属性等都可以随时间发生改变，一个典型的时态图如图 3 所示，边上的整数表示时间戳，即两个顶点之间发生联系的时间，例如：在社交网络中，该时间可以表示两个用户发生通信的时间；在交通网络中，该时间可以表示飞机从一个城市起飞并飞往另一个城市的时间；在银行转账网络中，该时间可以表示一个账户向另一个账户转账的时间等等。关联时间的边也被称为时态边。根据边上时间的表示方式和存储时态图方式的不同，常见的时态图模型有 3 个，即快照模型、边流模型和时间区间模型。

### 1.1 快照模型

快照模型（snapshot model）是时态图研究中一种常用的数据模型。该模型将一个时间区间（time interval，即一段时间）内的时态边映射到同一张静态图上，即一张快照[3]。如果图中某两点之间在该时间区间内存在多条时态边，则多条时态边只映射成两点之间的一条边。图 3 中的时态图在快照模型下的表示如图 4 所示，其中时间区间大小为 1。由于时间区间的大小设置为 1，所以图 4 中的快照数目为 4。需要注意的是不同大小的时间区间会导致同一个时态图转化为快照表示后快照数量不同、快照内边的数目不同。

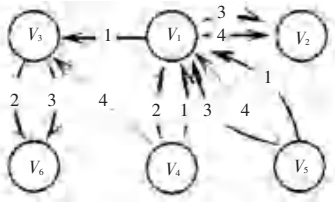


**图 3 时态图示例**

**Fig. 3 An Example of temporal graph**

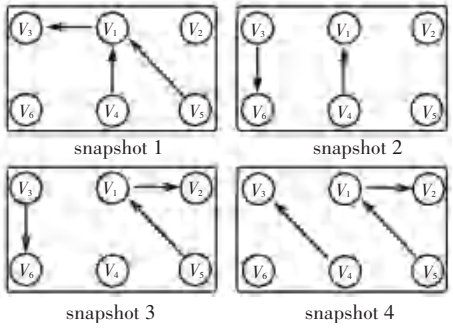


**图 4 快照模型**

**Fig. 4 Snapshot model**

### 1.2 边流模型

边流模型（edge-stream model）采用类似日志的形式将每条时态边单独表示，该模型详细地记录了每一条边每一次的变化。在通常情况下，所有的时态边根据边上的时态图信息升序排列。在边流模型中，一条边一般采用三元组表示，三元组中包含两个顶点及一个时刻，表示两个点之间建立联系的具体时间。图 3 中时态图的对应的边流表示，如图 5 所示。边流模型完整地记录了时态图所有的变化情况。

### 1.3 区间模型

区间模型（interval model）不同于以上两个模型表示离散时间的时态图，区间模型关注的是边上关联时间区间的情况，即边上的时态信息是连续的。时间区间表示两点之间关系持续的时长，时间区间一般使用开始时间和终止时间表示。一个适用区间模型表示的时态图，如图 6 所示。

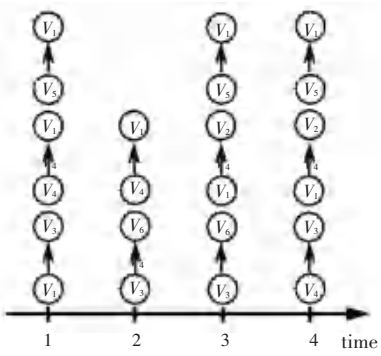


图 5　边流模型

Fig. 5　Edge-stream model

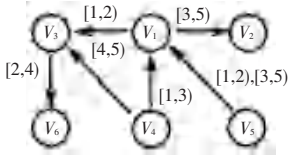


图 6　区间模型

Fig. 6　Interval model

## 2　时态图上图模式匹配相关工作

相比于静态图上大量关于图模式匹配的研究工作，时态图上图模式匹配的研究比较少。同静态图上的图模式匹配相比较，时态图上的图模式匹配问题除了需要考虑查询模式图引入的结构约束，还需要考虑定义在查询模式图的顶点或边上的时态约束。所以，时态图上的图模式匹配问题相较于静态图上的图模式匹配问题更加复杂[4]。根据时态数据的特点，本文总结出时态数据的 3 个重要性质，即时序性、持久性和演化性。基于这 3 个性质，对时态图上图模式匹配相关工作进行综述。

(1)时序性(Temporality)：由于数据对象本身关联时态信息，时态数据天然满足时序性，即根据数据对象上关联的时态信息，可以在数据全集上定义全序关系。在时态图数据上，时序性可以表现为任意两条边或者两个顶点在时态图中的出现存在先后顺序。

(2)演化性(Evolvability)：演化性表现为数据对象或数据对象间的关系发生变化，即数据对象或者数据对象间的关系随时间发生变化[5]。在时态图数据上，演化性可以表现为若干顶点的一个导出子图随时间变化为另一个导出子图[6]。

(3)持久性(Durability)：持久性表现为数据对象或者数据对象间关系不随时间发生变化，即数据对象或数据对象间的关系在多个时间不发生改变[7]。在时态图数据上，持久性可以表现为一个子图的结构在多个时间不发生改变[8-9]。

根据时态数据的 3 个性质，下面对时态图上图模式匹配的相关工作进行详细阐述。

### 2.1　时态图上基于时序性定义的图模式匹配

在时序图模式匹配的定义中，查询模式图中除了给定结构约束以外，还在边集上定义了先后顺序，即查询模式图中的边在时态数据图中的匹配边上的时间应该满足事先定义的顺序[10]。针对时序图模式匹配问题，已有的研究主要关注小的查询模式图[4,11-12]，即查询模式图中点的数目一般小于 6。该类方法主要采用遍历时态图方式搜索满足条件的匹配。除了要求匹配满足以上定义，Kosyfaki 等人的研究还对匹配的边上的权值和进行限制[13]；Li 等人和 Sun 等人针对任意大小的查询模式图进行研究，首先根据定义在边集上的偏序关系将一个查询模式图分解成若干个小的子查询模式图；其次，在输入的图流中查找这些子查询模式图的匹配，并将合格的匹配存储到索引中；最后，连接这些子查询模式图的匹配得到查询模式图的匹配[14-15]。Kumar 等人和潘敏佳等人主要关注时态图中一类特殊的结构—时态环，即起始顶点和结束顶点相同的一条简单的时态路径，都采用两阶段深度优先搜索的方法查找环结构[16-17]。以上的研究都是基于时态图的图流模型定义的图模式匹配。不同于之前采用的图流模型，Xu 等人在区间模型下定义图模式匹配问题，即在查询模式图和时态数据图的边上包含具体的时间区间[18]，该定义中要求对于任何查询模式图的匹配的边关联的时间区间同查询模式图中边上的时间区间必须重叠。Ma 等人首次根据图模拟和时态路径定义时态图上的图模式匹配基于连接的方式枚举匹配[19]。

### 2.2　时态图上基于持久性定义的图模式匹配

在持续图模式匹配的定义中，查询模式图中除了给定结构约束以外，还要求查询模式图的匹配在时态数据图中的多个时间出现。该定义是基于时态图的快照模型给出的，Semertzidis 等人使用位图(bitmap)位图构建索引，索引的大小同时态图中快照的个数成正比[8,20]。在搜索匹配的过程中，算法利用位图之间的位运算快速确定一个匹配持续的时间。

### 2.3　时态图上基于演化性定义的图模式匹配

在演化图模式匹配的定义中，查询模式图中除了给定结构约束以外，还要求查询模式图中的边在时态数据图中的匹配边在不同时间表现为不同的状态。该定义是基于时态图的快照模型给出的。Zufle 等人通过在查询模式图的边上指定时间集合限制其

匹配边上的时间集合定义时态图上的图模式匹配问题,并利用字符串编码子图结构加快匹配的搜索过程[21]。

## 3 现有工作的不足

通过以上综述可以看出,在时态图数据相关的众多研究问题中,时态图上的图模式匹配研究则刚刚开始。表 1 从定义、内存消耗、算法效率等方面给出现有时态图上图模式匹配工作的不足。

**表 1 时态图上图模式匹配总结**

**Tab. 1 Summary of graph pattern matching on temporal graphs**

| 现有研究工作 | 形式化定义 | 内存使用情况 | 算法效率 |
|---|---|---|---|
| 时序图模式匹配 | 已有形式化定义 | 内存消耗大 | 处理大规模时态图或者稠密的查询模式图时效率差 |
| 持续图模式匹配 | 已有形式化定义 | 内存消耗大 | 处理大规模时态图或者稠密的查询模式图时效率差 |
| 演化图模式匹配 | 没有形式化定义 | 无法考证 | 处理大规模时态图或者稠密的查询模式图时效率差 |

具体的说明如下。

(1)时序图模式匹配:现有时态图上的时序图模式匹配算法主要基于连接子查询的匹配的方式实现,这种方式会产生大量的中间结果,从而占用大量的内存。同时,验证连接后得到的结果,特别是那些不合格的匹配,会浪费大量的时间。基于以上原因导致现有方法在处理大规模时态图或稠密的查询模式图时时间效率差。

(2)持续图模式匹配:现有时态图上的时序图模式匹配算法主要利用位图构建索引降低验证匹配持续时间的代价。然而,索引占用的空间大小与时态图中的快照数目成正比。当时态图的规模变大或者时态图中快照数目较多时,索引将需要极大的存储空间,进而导致算法处理查询的时间效率变差。

(3)演化图模式匹配:在现有的工作中,没有发现时态图上演化图模式匹配的定义,只发现一个已有的工作可以处理演化图模式匹配问题,该工作需要对组成查询模式图的子图进行编码并构建索引,而索引的大小与时态图中快照的数目以及子图的匹配的数目成正比。当时态图的规模变大或者时态图中快照数目较多时,索引将耗费极大内存空间,进而导致算法处理查询的时间效率变差。

## 4 结束语

本文根据时态图的快照模型、边流模型和区间模型以及时态图数据的时序性、持续性和演化性对时态图上图模式匹配问题进行了综述。相比于静态图上图模式匹配问题丰硕的研究成果,时态图上图模式匹配问题的研究则刚刚开始。现有的时态图上图模式匹配研究在处理大规模时态图或者稠密的查询模式图时表现较差,而一些图模式匹配问题在时态图上还没有形式化定义。作为图研究领域一个重要的研究方向,时态图上图模式匹配问题从深度到广度、从理论到算法有待进一步的深入研究。

## 参考文献


[1] ULLMANN J R. An algorithm for subgraph isomorphism[J]. Journal of the ACM, 1976, 23(1): 31-42.

[2] HOLME P. Modern temporal network theory: A colloquium[J]. Physics of Condensed Matter, 2015, 88(9):1-30.

[3] KHURANA U, DESHPANDE A. Storing and analyzing historical graph data at scale[C]// International Conference on Extending Database Technology, 2016: 65-76.

[4] LIU P, BENSON A, CHARIKAR M. Sampling methods for counting temporal motifs[C]// Web Search and Data Mining. ACM, 2019: 294-302.

[5] DJAFRI N, FERNANDES A, PATON N W, et al. Spatio-temporal evolution: querying patterns of change in databases [C]//ACM SIGSPATIAL International workshop on Advances in Geographic Information System. ACM, 2002:35-41.

[6] BIMAL V, ALAN M, MEEYOUNG C, et al. On the evolution of user interactions in facebook[J]. Proceedings of the ACM Workshop on Online Social Networks. ACM, 2009, 39(4): 37-42.

[7] GAO J, AGARWAL P, YANG J. Durable top-k queries on temporal data[J]// Proceedings of the VLDB Endowment, 2018, 11(13): 2223-2235.

[8] SEMERTZIDIS K, PITOURA E. Durable graph pattern queries on historical graphs [C]// International Conference on Data Engineering. IEEE, 2016: 541-552.

[9] WANG R, JI W, SONG B. Durable relationship prediction and description using a large dynamic graph[J]. World Wide Web, 2018, 21(6): 1575-1600.

[10] TRIHINAS D, CHIROQUE L F, PALLIS G, et al. ATMoN: Adapting the "Temporality" in large-Scale dynamic networks [C]// International Conference on Distributed Computing Systems. IEEE, 2018: 400-410.

[11] PARANJAPE A, BENSON A, LESKOVEC J. Motifs in temporal networks[C]// Web Search and Data Mining. ACM, 2017: 601-610.


(下转封三)