Processing SPARQL Queries Over Linked Data—
A Distributed Graph-based Approach

Peng Peng, Lei Zou, M. Tamer Özsu, Lei Chen, Dongyan Zhao

1 Peking University, China;
2 University of Waterloo, Canada;
3 Hong Kong University of Science and Technology, China;

1 pku09pp,zoulei,zhaodongyan@pku.edu.cn, tamer.ozsu@uwaterloo.ca, leichen@cse.ust.hk

ABSTRACT

We propose techniques for processing SPARQL queries over linked data. We follow a graph-based approach where answering a query \( Q \) is equivalent to finding its matches over a distributed RDF data graph \( G \). We adopt a “partial evaluation and assembly” framework. Partial evaluation results of query \( Q \) over each repository—called local partial match—are found. In the assembly stage, we propose a centralized and a distributed assembly strategy. We analyze our algorithms both theoretically and the experimentally. Extensive experiments over both real and benchmark RDF repositories with billion triples demonstrate the high performance and scalability of our methods compared with that of the existing solutions.

1. INTRODUCTION

The semantic web data model, called the “Resource Description Framework”, or RDF, represents data as a collection of triples of the form (subject, property, object). A triple can be naturally seen as a pair of entities connected by a named relationship or an entity associated with a named attribute value. Linked Open Data (LOD), which is the focus of this paper, uses RDF to link web objects through their Uniform Resource Identifiers. LOD forms a directed graph where each vertex represents a Web document that contains data in the form of a set of RDF triples. These documents can be retrieved by looking up HTTP-scheme based URIs. A pair of document vertices is connected by a directed edge if any of the RDF triples in the source document contains a URI in its subject, predicate, or object position such that a lookup of this URI results in retrieving the target document. LOD forms a federated (or virtually integrated) database with the documents stored at autonomous sites some of which have the ability to run SPARQL queries over their own data (these are called “SPARQL endpoints”).

In this paper we address the issue of executing SPARQL queries over this type of federated RDF repositories. The approach we take to evaluating SPARQL queries is graph-based, which is consistent with RDF and SPARQL specifications. Consequently, answering a SPARQL query \( Q \) is equivalent to finding subgraph matches of query graph \( Q \) over an RDF graph \( G \), as employed, for example, in the gStore system [36]. We note that the problem in this context is different than graph subgraph matching in general graph processing systems since these depend on graph isomorphism while SPARQL query semantics are based on graph homomorphism.

There are several ways to address this issue [14]. One method is analogous to data warehousing where all of the data is downloaded into a single RDF database and queries are run over it using centralized techniques, such as Jena [12], RDF-3x [24], SW-Store [1], and gStore [36]. This is a reasonable approach, but at web scale, the feasibility of these approaches is a serious concern as is the freshness of query results.

A second approach that favors query result freshness is live querying [15]. In this approach, one starts with a set of seed LOD documents and follows the RDF links inside these until no further documents are reachable. These approaches provide up-to-date results, but have high latency that is hard to optimize.

The third alternative is to treat the interconnected RDF repositories as a virtually integrated distributed database, where each individual RDF repository is regarded as a fragment of this database. Each fragment \( F_i \) resides at its own site \( S_i \) and each query is executed at the relevant sites. Within this context, there are two approaches. One is to decompose queries and evaluate each subquery at the appropriate sites. The query decomposition strategy requires having full knowledge about the underlying schema of each fragment \( F_i \), which is not always feasible in RDF datasets since their fundamental characteristic is that they are self-describing without requiring an explicit schema. This is the approach followed by most existing works [18] [25] [27]. For example, DARQ [27] needs the service description that describes which triple patterns can be answered in each site, based on which, the SPARQL query is decomposed into several subqueries. The alternative approach is to use the “partial evaluation and assembly” strategy [17]. The basic idea of the partial evaluation is the following: given a function \( f(s,d) \), where \( s \) is the known input and \( d \) is yet unavailable input, the part of \( f \)'s computation that depends only on \( s \) generates a partial answer [8]. In our setting, each site \( S_i \) treats fragment \( F_i \) as the known input in the partial evaluation stage. The partial evaluation technique has been used in compiler optimization [17], querying XML trees [4], and evaluating reachability queries [8] and graph simulation [25] [9] over graphs. In this paper, we follow the partial evaluation and assembly strategy to evaluate SPARQL queries over a distributed and interconnected set of RDF repositories.

Because of interconnections between datasets, processing of queries requires special care. For example, consider the four RDF datasets at different sites in LOD shown in Figure 1 and its graph representation in Figure 2. Each entity in RDF is represented by a URI (uniform resource identifier), the prefix of which always denotes the location of the dataset. For example, “s1:dir1” has the prefix “s1”, meaning that the entity is located at site s1. There are cross-
There are two important issues to be addressed in this framework. The first is to compute the partial evaluation results at each site given a query graph $Q$ (i.e., the local partial match), which is intuitively the overlapping part between a crossing match and a fragment. This is discussed in Section 4. The second one is the assembly of these local partial matches to compute crossing matches. We consider two different strategies: centralized assembly, where all local partial matches are sent to a single site (Section 5.2); and distributed assembly, where the local partial matches are assembled at a number of sites in parallel (Section 5.3).

This is the first work that considers the LOD repositories (or similarly distributed and autonomous RDF repositories) as a virtually integrated distributed database, and proposes the partial evaluation and assembly framework for executing SPARQL queries. Within this context we make the following contributions.

1. We formally define the partial evaluation result for SPARQL queries and propose an efficient algorithm to find them. We prove theoretically the correctness and the optimality of our algorithm.

2. We propose centralized and distributed methods to assemble the local partial matches and experimentally study their performance advantages and bottlenecks in different settings.

2. RELATED WORK

There are two threads of related work: distributed SPARQL query processing and partial evaluation.

**Distributed SPARQL Query Processing.** As noted above, generally speaking, there are two different approaches to distributed RDF query processing: federated SPARQL query processing, and data partition-based query processing.

(1) Federated SPARQL Query Processing

Federated queries run SPARQL queries over multiple distributed and autonomous RDF datasets. A typical example is linked data, where different RDF repositories are interconnected, providing a virtually integrated distributed database.
A common technique is to precompute metadata for each individual RDF repository. Based on the metadata, the original SPARQL query is decomposed into several subqueries, where each subquery is sent to some RDF repositories that could answer the query. These subquery results are then joined together to answer the original SPARQL query. In DARQ \[27\], the metadata is called service description that describes which triple patterns (i.e., predicate) can be answered. Consider each triple pattern \( t \) in a SPARQL query \( Q \). If the triple pattern \( t \) matches exactly one data source, the triple will be added to the set of a subquery for this RDF data source. If \( t \) matches multiple data sources, the triple must be sent to all matching RDF data sources as separate subqueries. In \[29,12\], the metadata is called Q-Tree, based on which, the SPARQL is also decomposed into several parts.

FedX \[28\] follows the same approach, but it does not require preprocessing. However, for each triple pattern \( t \) in a SPARQL query \( Q \), FedX sends “SPARQL ASK” to collect the metadata on the fly. Based on the returned results, it annotates each triple pattern with its relevant sources. Each triple pattern is a basic query unit. In order to improve the query performance, several triple patterns in the same exclusive groups \[28\] can be sent in a single subquery to the respective source.

The problem studied in this paper belongs to federated SPARQL query approach, but, we adopt a different strategy, namely partial evaluation and assembly. Rather than decomposing a query, we send the whole SPARQL query to each RDF data source. The partial evaluation results from each RDF data source are merged together. The main benefits of our solution are twofold: (1) The number of remote requests is minimized. For example, for each triple pattern \( t \) in SPARQL \( Q \), FedX sends “SPARQL ASK” to each RDF data source. For each data source, the number of remote requests is \(|E(Q)|\) (the number of edges in SPARQL query graph \( Q \)). However, we only send a single SPARQL query to each RDF data source. (2) The intermediate result size is reduced. We propose several graph structure-based pruning techniques to reduce the intermediate result size. Experimental results show the our approach outperforms the existing ones. We note that our approach requires minor modifications to SPARQL endpoints to facilitate partial query evaluation.

(2) Data Partition-based Method

A number of approaches \[20\] \[15\] \[16\] \[35\] \[24\] \[22\] \[11\] start with a large RDF graph \( G \) (i.e., follow what we referred to as the data warehousing approach), and partition \( G \) into several fragments. At run time, a SPARQL query \( Q \) is decomposed into several subqueries such that each subquery can be answered locally at one site. Note that each of these works proposes its own data partitioning strategy, and different partitioning strategies result in different processing methods. Generally speaking, there are three major partitioning strategies: table-based, graph-based, and unit-based.

Table-based methods \[20\] \[15\] \[16\] \[35\] \[24\] \[22\] \[11\] partition the RDF data by considering it as a set of relational tables. Jena-HBase \[20\] is a distributed version of Jena \[32\] that maps Jena tables to HBase. A query \( Q \) is also translated to a query over HBase and is executed on the mapped tables. HadoopRDF \[16\] and P-Partition \[35\] extend SW-store \[11\] by further partitioning vertically partitioned tables. They also decompose the SPARQL query according to the predicates in the triple patterns.

Graph-based methods \[15\] \[15\] \[33\] partition the RDF graph using graph partitioning techniques. In GraphPartition \[15\], an RDF graph \( G \) is partitioned into \( n \) fragments, and each fragment is extended by including \( N \)-hop neighbors of boundary vertices. According to the partitioning strategy, the diameter of the graph corresponding to each decomposed subquery should not be larger than \( N \) to enable subquery processing at each local site. In EAGRE \[15\], the graph fragmentation is based on the predicates adjacent to the entities. If two entities’ incident predicates are similar, they are in one class and should be in the same fragment. At run time, query graph \( Q \)’s decomposition always depends on the predicates incident to each variable in \( Q \). Trinity.RDF \[33\] stores the whole RDF graph in a distributed graph system over a memory cloud (called Trinity \[30\]). Trinity.RDF uses the vertex in the RDF graph as the key and its adjacency list as the value. Since Trinity is a main memory system, it can randomly access all neighbors of a vertex. Hence, Trinity.RDF uses a graph exploration-based method to answer SPARQL queries.

Unit-based methods \[21\] \[22\] \[11\] decompose the RDF data into some partition units, and distribute these units among different sites. Kising Lee et. al. \[21\] \[22\] define the partition unit as a vertex and its neighbors, which they call a “vertex block”. The vertex blocks are distributed based on a set of heuristic rules. A query is partitioned into blocks that can be executed among all sites in parallel and without any communication. TriAD \[11\] distributes RDF-3X \[24\]. TriAD uses METIS \[19\] to divide the RDF graph into many partitions and the number of result partitions is much more than the number of sites. Each result partition is considered as a unit and distributed among different sites. Meanwhile, TriAD constructs a summary graph to maintain the partitioning information. During the query processing phase, TriAD determines a query plan based on this summary graph.

All of the above methods require partitioning and distributing the RDF data according to specific requirements of their approaches. In contrast, in LOD, the RDF are already distributed over different sites. Therefore, these data partition-based solutions cannot be used for processing queries over LOD unless LOD are collected in one central site and redistributed again, which is impractical.

Partial Evaluation. Partial evaluation has been used in many applications ranging from compiler optimization to distributed evaluation of functional programming languages \[17\]. Recently, partial evaluation has also been used for evaluating queries on distributed XML trees and graphs \[2\] \[4\] \[5\] \[8\]. In \[2\] \[4\] \[5\] \[8\], partial evaluation is used to evaluate some XPath queries on distributed XML. These works serialize XPath queries to a vector of subqueries, and find the partial results of all subqueries at each site by using a top-down \[2\] or bottom-up \[3\] traversal over the XML tree. Finally, all partial results are assembled together at the server site to form final results. Note that, since XML is a tree-based data structure, these works serialize XPath queries and traverse XML trees in a topological order. However, the RDF data and SPARQL queries are graphs rather than trees. Serializing the SPARQL queries and traversing the RDF graph in a topological order is not intuitive.

There are some prior works that consider partial evaluation on graphs. For example, Fan et al. \[8\] study reachability query processing over distributed graphs using the partial evaluation strategy. Partial evaluation-based graph simulation is well studied in Fan et al. \[2\] and Shuai et al. \[23\]. However, SPARQL query semantics is different than these (i.e., SPARQL is based on graph homomorphism \[25\]) and pose additional challenges. To the best of our knowledge, there is no prior work in applying partial evaluation to SPARQL query processing. As discussed in \[9\], graph simulation defines a relation between vertices in query graph \( Q \) (i.e. \( V(Q) \)) and that in the data graph \( G \) (i.e., \( V(G) \)), but, the graph homomorphism is a function (not a relation) between \( V(Q) \) and \( V(G) \). Thus, the solutions proposed in \[9\] \[25\] cannot be applied to the problem studied in this paper.

3. BACKGROUND
An RDF data set can be represented as a graph where subjects and objects are vertices and triples are labeled edges.

**Definition 3.1. (RDF Graph)** An RDF graph is denoted as \( G = (V, E, \Sigma) \), where \( V \) is a set of vertices that correspond to all subjects and objects in RDF data; \( E \subseteq V \times V \) is a set of directed edges that correspond to all triples in RDF data; \( \Sigma \) is a set of edge labels. For each edge \( e \in E \), its edge label is its corresponding property.

Similarly, a SPARQL query can also be represented as a query graph \( Q \). For simplicity, we ignore FILTER statements in SPARQL syntax and only consider BGP (basic graph pattern) in this paper.

**Definition 3.2. (SPARQL Query)** A SPARQL query is denoted as \( Q = (V^Q, E^Q, \Sigma^Q) \), where \( V^Q \subseteq V \cup V_{var} \) is a set of vertices, where \( V \) denotes all vertices in RDF graph \( G \) and \( V_{var} \) is a set of variables; \( E^Q \subseteq V^Q \times V^Q \) is a set of edges in \( Q \); Each edge \( e \) in \( E^Q \) either has an edge label in \( \Sigma \) (i.e., property) or the edge label is a variable.

In this paper, we assume that \( Q \) is a connected graph; otherwise, all connected components of \( Q \) are considered separately. Answering a SPARQL query is equivalent to finding all subgraph matches (Definition 3.3) of \( Q \) over RDF graph \( G \).

**Definition 3.3. (SPARQL Match)** Consider a RDF graph \( G \) and a connected query graph \( Q \) that has \( n \) vertices \( \{v_1, \ldots, v_n\} \). A subgraph \( M \) with \( m \) vertices \( \{u_1, \ldots, u_m\} \) (in \( G \)) is said to be a match of \( Q \) if and only if there exist a function \( f \) from \( \{v_1, \ldots, v_n\} \) to \( \{u_1, \ldots, u_m\} \) (\( n \geq m \)), such that the following conditions hold:

1. if \( v_i \) is not a variable, \( f(v_i) \) and \( v_i \) have the same URI or literal value (\( 1 \leq i \leq n \));
2. if \( v_i \) is a variable, there is no constraint over \( f(v_i) \) except that \( f(v_i) \in \{u_1, \ldots, u_m\} \);
3. if there exists an edge \( \overrightarrow{v_i v_j} \) in \( Q \), there also exists an edge \( f(v_i) \overrightarrow{f(v_i)} f(v_j) \) in \( G \); furthermore, \( f(v_i) f(v_j) \) has the same property as \( v_i v_j \) unless that the label of \( v_i v_j \) is a variable.

Vector \( f(v_1), \ldots, f(v_n) \) is a serialization of a SPARQL match. Note that we allow that \( f(v_i) = f(v_j) \) when \( 1 \leq i \neq j \leq n \). In other words, a match of SPARQL \( Q \) defines a graph homomorphism.

In the context of this paper, RDF graph \( G \) is distributed over different sites. We call \( G \) an integrated distributed RDF graph. Each constituent RDF repository is called a fragment of \( G \). For simplicity, we assume that each entity is only resident in a single site, i.e., we ignore replication (see condition (1) of Definition 3.4), but handling the more general case is not complicated.

**Definition 3.4. (Integrated Distributed RDF Graph)** Given \( k \) interconnected RDF repositories, each of them represented as an RDF graph \( F_i \) (\( i = 1, \ldots, k \)), the integrated distributed RDF graph \( G = (V, E, \Sigma) \) consists of a set of \( \{F_1, F_2, \ldots, F_k\} \). Each \( F_i \) is called a fragment of \( G \), which is specified by \( (V_i \cup V^F_i, E_i \cup E^F_i, \Sigma_i) \) (\( i = 1, \ldots, k \)) such that

1. \( \{V_1, \ldots, V_k\} \) is a partitioning of \( V \), i.e., \( V_i \cap V_j = 0, 1 \leq i, j \leq k, i \neq j \) and \( \bigcup_{i=1}^{k} V_i = V \);\n2. \( E_i \subseteq V_i \times V_i, i = 1, \ldots, k \);\n
In this paper, we assume that query graph \( Q \) is connected; otherwise, we can process each connected component of \( Q \) separately.

3. \( E_i^c \) is a set of crossing edges between \( F_i \) and other fragments, i.e., \( E_i^c = \bigcup_{u \in F_i \land u \neq E_i}(u \wedge u \in E) \cup \bigcup_{u \in F_i \land u \notin E_i}(u \wedge u \in E) \);
4. A vertex \( u' \in V^F_i \) if and only if vertex \( u' \) resides in other fragment \( F_j \) and \( u' \) is an endpoint of a crossing edge between fragment \( F_i \) and \( F_j \) (\( F_i \neq F_j \)), i.e., \( E_i = \bigcup_{u \in F_i \land u \not\in F_i}(u \wedge u \in E) \cup \bigcup_{u \in F_j \land u \not\in F_j}(u \wedge u \in E) \);
5. Vertices in \( V^F_i \) are called extended vertices of \( F_i \) and all vertices in \( V_i \) are called internal vertices of \( F_i \);
6. \( \Sigma_i \) is a set of edge labels in \( F_i \).

**Example 1.** Given four interconnected RDF repositories in Figure 5, the integrated RDF graph \( G \) consists of four constituent RDF fragment graphs \( F_1, F_2, F_3 \) and \( F_4 \) in Figure 7. The numbers besides the vertices are vertex IDs that are introduced for ease of presentation. In Figure 2 \{002, 001\} is a crossing edge between \( F_1 \) and \( F_2 \). As well, edges 004, 011, 001, 012 and 006, 008 are crossing edges between \( F_1 \) and \( F_3 \). Hence, \( V^F_1 = \{002, 006, 012, 004\} \) and \( E^F_1 = \{002, 001, 004, 011, 001, 012, 006, 008\} \).

**Definition 3.5. (Problem Statement)** Given an integrated distributed RDF graph \( G \) that consists of \( k \) fragments \( \{F_1, F_2, \ldots, F_k\} \), each located at a different computing node, and a SPARQL query graph \( Q \), our goal is to find all SPARQL matches of \( Q \) in \( G \).

Inner matches can be computed locally by a centralized RDF triple store, such as RDF-3X [24], SW-store [1] and gStore [36]. The main issue of answering SPARQL queries over the distributed RDF graph is finding crossing matches efficiently. That is the focus of this paper.

4. PARTIAL EVALUATION

We first discuss what is a local partial match (Section 4.1) and then discuss how to compute it efficiently (Section 4.2).

**4.1 Local Partial Match—Definition**

Recall that each site \( S_i \) receives the full query graph \( Q \) (i.e., there is no query decomposition). In order to answer a SPARQL query \( Q \), each site \( S_i \) computes the partial answers (called local partial matches) based on the known input \( F_i \). Intuitively, a local partial match \( P_M \) (in fragment \( F_i \)) is an overlapping part between an unknown crossing match \( M \) and fragment \( F_i \). An unknown crossing match \( M \) means that we do not know \( M \) at the partial evaluation stage. Moreover, the unknown crossing match \( M \) may or may not exist depending on the yet unavailable input \( d \), i.e., other fragments of \( G \). Based on the known input, i.e., fragment \( F_i \), we cannot judge whether or not \( M \) exists. For example, the subgraph induced by vertices 014, 007, 001 and 002 (shown in shared vertices and red edges) in Figure 5 is a local partial match between \( M \) and \( F_i \).

**Definition 4.1. (Local Partial Match)** Given a SPARQL query graph \( Q \) with \( n \) vertices \( \{v_1, \ldots, v_n\} \) and a connected subgraph \( P_M \) with \( m \) vertices \( \{u_1, \ldots, u_m\} \) (\( m \leq n \)) in a fragment \( F_i = (V_i \cup V^F_i, E_i \cup E^F_i, \Sigma_i) \). \( P_M \) is a local partial match in fragment \( F_i \) if and only if there exists a function \( f : \{v_1, \ldots, v_n\} \rightarrow \{u_1, \ldots, u_m\} \cup \{\text{NULL}\} \) such that the following conditions hold:

1. if \( v_i \) is not a variable, \( f(v_i) \) and \( v_i \) have the same URI or literal or \( f(v_i) = \text{NULL} \);
2. if \( v_i \) is a variable, \( f(v_i) \in \{u_1, \ldots, u_m\} \) or \( f(v_i) = \text{NULL} \).
3. If there exists an edge $v_i v_j$ in $Q$ (1 ≤ $i$ ≠ $j$ ≤ $n$), there also exists an edge $f(v_i) f(v_j)$ in $PM$ with property $p$ unless $f(v_i)$ and $f(v_j)$ are both in $V'_1$ or $f(v_i) = NULL$ $f(v_j) = NULL$. Furthermore, $f(v_i) f(v_j)$ has the same property as $v_i v_j$ unless the label of $v_i v_j$ is a variable.

4. $PM$ contains at least one crossing edge.

5. If $f(v_i) \in V'_2$ (i.e., $f(v_i)$ is an internal vertex in $F_1$) and $3 f(v_i) \in Q$ (or $v_i v_j \in Q$), there must exist $f(v_j) \neq NULL$ and $3 f(v_j) f(v_i) \in PM$ (or $3 f(v_i) f(v_j) \in PM$). Furthermore, if $\overline{v_i v_j}$ (or $\overline{v_i v_j}$) has a property $p$, $f(v_i) f(v_j)$ (or $f(v_j) f(v_i)$) has the same property $p$.

6. Any two vertices $v_i$ and $v_j$ (in query $Q$), where $f(v_i)$ and $f(v_j)$ are both internal vertices in $PM$, are weakly connected (see Definition 4.2) in $Q$.

Vector $[f(v_1), ..., f(v_n)]$ is a serialization of a local partial match.

Example 2. Given a SPARQL query $Q$ with six vertices in Figure 3, the subgraph induced by vertices 002, 001, 007 and 014 (shown in shaded circles and red edges) is a local partial match of $Q$ in fragment $F_1$. The function is $f((v_1, 002), (v_2, 001), (v_3, NULL), (v_4, 007), (v_5, NULL), (v_6, 014))$. Actually, there are five different local partial matches in $F_1$. We show them in Figure 4.

Definition 4.1 formally defines a local partial match, which is a subset of a complete SPARQL match. Therefore, some conditions in Definition 4.1 are analogous to SPARQL match with some subtle differences. In Definition 4.3, some vertices of query $Q$ are not matched in a local partial match. They are allowed to match a special value NULL. For example, $v_i$ and $v_j$ in Example 2 are matched to NULL. As mentioned earlier, a local partial match is the overlapping part of an unknown crossing match and a fragment $F_i$. Therefore, it must have a crossing edge, i.e., Condition 4.

The basic intuition of Condition 5 is that if vertex $v_i$ (in query $Q$) is matched to an internal vertex, all $v_i$’s neighbours should be matched in this local partial match as well. The following example illustrates the intuition.

Example 3. Let us recall the local partial match $PM_i^2$ of Fragment $F_1$ in Figure 3. An internal vertex 001 in fragment $F_1$ is matched to vertex $v_2$ in query $Q$. Assume that $PM$ is an overlapping part between a crossing match $M$ and fragment $F_1$. Obviously, $v_2$’s neighbors, such as $v_1$ and $v_4$, should also be matched in $M$. Furthermore, the matching vertices should be 001’s neighbors. Since 001 is an internal vertex in $F_1$, 001’s neighbors are also in fragment $F_1$.

According to the above analysis, if a subgraph $PM$ violates Condition 5, $PM$ cannot be a subgraph of a crossing match. In other words, we are not interested in these subgraphs when finding local partial matches, since they do not contribute to any crossing match.

Definition 4.2. Two vertices are weakly connected in a directed graph if and only if there exists a connected path between the two vertices if we replace all directed edges with undirected edges. The path is called a weakly connected path between the two vertices.

Condition 6 will be used to prove the correctness of our algorithm in Theorems 4.1 and 4.2. We will discuss them soon. The following example shows all local partial matches in the running example.

Example 4. Given a query $Q$ in Figure 5 and an RDF graph $G$ in Figure 2, Figure 2 shows all local partial matches and their serialization vectors in each fragment. A local partial match in fragment $F_i$ is denoted as $PM_i^2$, where the superscript distinguishes different local partial matches in the same fragment. Furthermore, we underline all extended vertices in serialization vectors.

The correctness of our method is stated in the following theorem.

1. The overlapping part between any crossing match $M$ and fragment $F_i$ (i = 1, ..., k) must be a local partial match.

2. Missing any local partial match may lead to result dismissal. Thus, our algorithm should find all local partial matches in each fragment.

3. It is impossible to find two local partial matches $M$ and $M'$ in fragment $F_i$, where $M'$ is a subgraph of $M$, i.e., each local partial match is maximal.

Theorem 4.1. Given any crossing match $M$ of SPARQL query $Q$ in an RDF graph $G$, if $M$ overlaps with some fragment $F_i$ (i = 1, ..., k), the overlapping part between $M$ and fragment $F_i$, denoted as $PM_i$, must be a local partial match in fragment $F_i$.

Let us recall Example 3. There are some local partial matches that do not contribute to any crossing match, such as $PM_i^2$ in Figure 3. We call these local partial matches false positives. However, the partial evaluation stage only depends on the known input. If we do not know the structures of other fragments, we cannot judge whether or not $PM_i^2$ is a false positive. Formally, we have the following theorem, which means that we have to find all local partial matches in each fragment $F_i$ in the partial evaluation stage.

Theorem 4.2. The results generated by the partial-evaluation-and-assembly algorithm satisfy no-false-negative requirement if and only if all local partial matches in each fragment are found in the partial evaluation stage.

Finally, we discuss another feature of a local partial match $PM_i$, in fragment $F_i$. Any $PM_i$ cannot be enlarged by introducing more vertices or edges to become a larger local partial match. The following theorem formalizes this.

Theorem 4.3. Given a query graph $Q$ and an RDF graph $G$, if $PM_i$ is a local partial match under function $f$ in fragment $F_i$ = $(V \cup V'_1, E \cup E'_i)$, there exists no local partial match $PM'_i$ under function $f'$ in $F_i$, where $f' \subseteq f$.

4.2 Computing Local Partial Matches

Given a SPARQL query $Q = (V^0, E^0)$ and a fragment $F_i = (V_i \cup V'_i, E_i \cup E'_i)$, the partial evaluation goal is to find all local partial matches (according to Definition 4.1) in $F_i$. The matching process consists of determining a function $f$ that associates vertices of $Q$ with vertices of $F_i$. The matches are expressed as a set of pairs $(v, u)$ (v ∈ Q and u ∈ $F_i$). A pair $(v, u)$ represents the matching of a vertex $v$ of query $Q$ with a vertex $u$ of fragment $F_i$. The set of vertex pairs $(v, u)$ constitutes function $f$ referred to in Definition 4.1.

A high-level description of finding local partial matches is outlined in Algorithm 1 and Function $\text{ComParMatch}$. According to Conditions 1 and 2 of Definition 4.1, each vertex $v$ in query graph $Q$ has a candidate list of vertices in fragment $F_i$. Since function $f$ is as a set of vertex pairs $(v, u)$ (v ∈ Q and $u \in F_i$), we start with an empty set. In each step, we introduce a candidate vertex pair $(v, u)$.

2The proofs of all theorems are in Appendix.
to expand the current function $f$, where vertex $u$ (in fragment $F_i$) is a candidate of vertex $v$ (in query $Q$).

Assume that we introduce a new candidate vertex pair $(v', u')$ into the current function $f$ to form another function $f'$. If $f'$ violates any condition except for Conditions 4 and 5 of Definition 4.1, the new function $f'$ cannot lead to a local partial match (Lines 6-7 in Function ComParMatch). If $f'$ satisfies all conditions except for Conditions 4 and 5, it means that $f'$ can be further expanded (Lines 8-9 in Function ComParMatch). If $f'$ satisfies all conditions, then $f'$ specifies a local partial match and it is reported (Lines 10-11 in Function ComParMatch).

Algorithm 1: Computing Local Partial Matches

Input: A fragment $F_i$ and a query graph $Q$.
Output: The set of all local maximal partial matches in $F_i$, denoted as $\Omega(F_i)$.

1. Select one vertex $v$ in $Q$.
2. for each candidate vertex $u$ with regard to $v$ do
   3. Initialize a function $f$ with $(v, u)$
   4. Call Function ComParMatch($f$)
5. Return $\Omega(F_i)$;

Function ComParMatch($f$)

1. if all vertices of query $Q$ have been matched in the function $f$ then
2. Return;
3. Select an unmatched $v'$ adjacent to a matched vertex $v$ in the function $f$
4. for each candidate vertex $u'$ with regard to $v'$ do
5. $f' \leftarrow f \cup (v', u')$
6. if $f'$ violates any condition (except for condition 4 and 5 of Definition 4.1) then
7. Continue
8. if $f'$ satisfies all conditions (except for condition 4 and 5 of Definition 4.1) then
9. Call ComParMatch($f'$)
10. if $f'$ satisfies all conditions of Definition 4.1 then
11. if $f$ specifies a local partial match $PM$ that will be inserted into the answer set $\Omega(F_i)$

As shown in Algorithm 1 in each step, we introduce a new candidate vertex pair $(v', u')$, where $v'$ is a vertex in query graph $Q$. The order of selecting query vertex can be arbitrarily defined. However, QuickSI [29] proposes several heuristics to select an optimized order that can speed up the matching process. These rules are also utilized in our experiments.

5. ASSEMBLY

Using Algorithm 1, each site $S_i$ finds all local partial matches in fragment $F_i$. The next step is to assemble partial matches to form crossing matches and compute the final result. We propose two assembly strategies: centralized and distributed (or parallel).

In centralized, all local partial matches are sent to a single site for assembly. For example, in a client/server system, all local partial matches may be sent to the server. In distributed/parallel, local partial matches are combined at a number of sites in parallel. In Section 5.1, we define a basic join operator for assembly. Then, we propose a centralized assembly algorithm in Section 5.2 using the join operator. In Section 5.3, we study how to assemble local partial matches in a distributed manner.

5.1 Join-based Assembly

We first define the conditions under which two partial matches are joinable. Obviously, crossing matches can only be formed by assembling partial matches from different fragments. If local partial matches from the same fragment could be assembled, this would result in a larger local partial match in the same fragment, which is contrary to Theorem 4.3.

Definition 5.1. (Joinable) Given a query graph $Q$ and two fragments $F_i$ and $F_j$ ($i \neq j$), let $PM_i$ and $PM_j$ be two local partial matches over fragments $F_i$ and $F_j$ under functions $f_i$ and $f_j$, respectively. $PM_i$ and $PM_j$ are joinable if and only if the following conditions hold: 1. There exist no internal vertices $u$ and $u'$ in $PM_i$ and $PM_j$, respectively, such that $f_i^{-1}(u) = f_j^{-1}(u')$; 2. There exists at least one crossing edge $uu'$ such that $u$ is an internal vertex and $u'$ is an extended vertex in $F_i$, while $u$ is an extended vertex and $u'$ is an internal vertex in $F_j$. Furthermore, $f_i^{-1}(u) = f_j^{-1}(u)$ and $f_i^{-1}(u') = f_j^{-1}(u')$.

The first condition says that there does not exist different vertices in separate partial matches that match the same query vertex. The second condition says that two local partial matches share at least one common crossing edge that corresponds to the same query edge.
Example 5. Let us recall query two local partial matches $PM_i$ and $PM_j$ in Figure 4. There do not exist two different vertices in the two local partial matches that match the same query vertex. Furthermore, they share a common crossing edge $002\rightarrow 001$, where 002 and 001 match query vertices $v_2$ and $v_1$ in the two local partial matches, respectively. Hence, they are joinable.

The join result of two joinable local partial matches is defined as follows.

**Definition 5.2.** Join Result. Given a query graph $Q$ and two fragments $F_i$ and $F_j$, $i \neq j$, let $PM_i$ and $PM_j$ be two joinable local partial matches of $Q$ over fragments $F_i$ and $F_j$, under functions $f_i$ and $f_j$, respectively. The join of $PM_i$ and $PM_j$ is defined under a new function $f$ (denoted as $PM = PM_i \bowtie_j PM_j$), which is defined as follows for any vertex $v$ in $Q$:

1. $f(v) = NULL \land f(v) = NULL$, $f(v) \leftarrow f_i(v)$
2. $f(v) = NULL \land f(v) \neq NULL$, $f(v) \leftarrow f_j(v)$
3. $f(v) \neq NULL \land f(v) \neq NULL$, $f(v) \leftarrow f_j(v)$ (In this case, $f_i(v) = f_j(v)$)
4. $f(v) = NULL \land f(v) = NULL$, $f(v) \leftarrow NULL$

5.2 Centralized Assembly

In centralized assembly, all local partial matches are sent to a final assembly site. We propose an iterative join algorithm (Algorithm 2) to find all crossing matches. In each iteration, a pair of local partial matches are joined. When the join is complete (i.e., a match has been found), the result is returned (Lines 12-13 in Algorithm 2); otherwise, it is joined with other local partial matches in the next iteration (Lines 14-15). There are $|V(Q)|$ iterations of Lines 4-16 in the worst case, since at each iteration only a single new matching vertex is introduced (worst case) and $Q$ has $|V(Q)|$ vertices. If no new intermediate results are generated at some iteration, the algorithm can stop early (Lines 5-6).

5.2.1 Partitioning-based Join Processing

The join space in Algorithm 2 is large, since we need to check if every pair of local partial matches $PM_i$ and $PM_j$ are joinable. This subsection proposes an optimized technique to reduce the join space.

The intuition of our method is as follows. We partition all local partial matches such that two local partial matches in the same partition cannot be joinable; we only consider joining local partial matches from different partitions. The following theorem specifies which local partial matches can be put in a partition.

**Theorem 5.1.** Given two local partial matches $PM_i$ and $PM_j$ from fragments $F_i$ and $F_j$ with functions $f_i$ and $f_j$, respectively, if there exists a query vertex $v$ where both $f(v)$ and $f_j(v)$ are internal vertices of fragments $F_i$ and $F_j$, respectively, $PM_i$ and $PM_j$ are not joinable.

Example 6. Figure 5 shows the serialization vectors (defined in Definition 4.2) of four local partial matches. For each local partial match, there is an internal vertex that matches $v_1$ in query graph. The underline indicates the extended vertex in the local partial match. According to Theorem 5.1, none of them are joinable.

$3f_j(v) = NULL$ means that vertex $v$ in query $Q$ is not matched in local partial match $PM_j$. It is formally defined in Definition 4.1 condition (2).

In this paper, we use “$\rightarrow$” to denote the assignment operator.

| Algorithm 2: Centralized Join-based Assembly |
|---------------------------------------------|
| **Input:** $\Omega(F_i)$, i.e., the set of local partial matches in each fragment $F_i$, $i = 1, \ldots, k$ |
| **Output:** All crossing matches set $RS$ |
| 1. Each fragment $F_i$ sends the set of local partial matches in each fragment $F_i$ (i.e., $\Omega(F_i)$) to a single site for the assembly |
| 2. Let $\Omega := \bigcup_{i=1}^{k} \Omega(F_i)$ |
| 3. Set $MS := \Omega$ |
| 4. for $N := 1$ to $|V(Q)|$ do |
| 5. if $|MS| = 0$ then |
| 6. Break; |
| 7. Set $MS' := \phi$ |
| 8. for each local partial match $PM$ in $MS$ do |
| 9. for each local partial match $PM'$ in $\Omega$ do |
| 10. if $PM$ and $PM'$ are joinable then |
| 11. Set $PM' = PM \bowtie_j PM'$ |
| 12. if $PM'$ is a complete match of $Q$ then |
| 13. put $PM'$ into $RS$ |
| 14. else |
| 15. put $PM'$ into $MS'$ |
| 16. $MS := MS'$ |
| 17. Return $RS$ |

Figure 5: Partitioning Local Partial Matches on $v_1$

**Definition 5.3.** Local Partial Match Partitioning. Consider a SPARQL query $Q$ with $n$ vertices $\{v_1, \ldots, v_n\}$. Let $\Omega$ denote all local partial matches. $\mathcal{P} = \{P_1, \ldots, P_n\}$ is a partitioning of $\Omega$ if and only if the following conditions hold: 1) Each partition $P_i$ ($i = 1, \ldots, n$) consists of a set of local partial matches, each of which has an internal vertex that matches $v_i$; 2) $P_i \cap P_j = \emptyset$, where $1 \leq i \neq j \leq n$; 3) $P_1 \cup \ldots \cup P_n = \Omega$.

Example 7. Let us consider all local partial matches of our running example in Figure 4. Figure 6 show an example partitioning.

As mentioned earlier, we only need to consider joining local partial matches from different partitions. Given a partitioning $\mathcal{P} = \{P_1, \ldots, P_n\}$, Algorithm 3 shows how to perform partitioning-based join of local partial matches. Note that different partitionings and the different join orders in the partitioning will impact the performance of Algorithm 3. In Algorithm 3, we assume that the partitioning $\mathcal{P} = \{P_1, \ldots, P_n\}$ is given, and that the join order is fixed from $P_1$ to $P_n$, i.e. the order in $\mathcal{P}$. Choosing a good partitioning and the optimal join order will be discussed in Section 5.2.5 and Section 5.2.6.

The basic idea of Algorithm 3 is to iterate the join process on each partition. First, we set $MS \leftarrow P_1$ (Line 1 in Algorithm 3). Then, we try to join local partial matches $PM$ in $MS$ with local partial matches $PM'$ in $P_i$ (the first loop of Line 3-13). If the join result is a complete match, it is inserted into the answer set $RS$ (Lines 8-9). If the join result is an intermediate result, we insert it into a temporary set $MS'$ (Lines 10-11). We also need to insert $PM'$ into $MS'$, since the local partial match $PM'$ (in $P_{i+1}$) will join local partial matches in the later partitions (Line 12). At the
end of the iteration, we insert all intermediate results (in MS') into MS, which will join local partial matches in the later partitions in the next iterations (Line 13). We iterate the above steps for each partition in the partitioning (Lines 3-13).

Algorithm 3: Partitioning-based Joining Local Partial Matches

Input: A partitioning $\mathcal{P} = \{P_1, \ldots, P_n\}$ of all local partial matches.

Output: All crossing matches set $RS$.

1. $MS \leftarrow P_v$.
2. for $i \leftarrow 2$ to $n$ do
   3. $MS' \leftarrow \emptyset$.
   4. for each partial match $PM$ in $MS$ do
      5. for each partial match $PM'$ in $P_v$ do
         6. if $PM$ and $PM'$ are joinable then
            7. Set $PM'' \leftarrow PM \Rightarrow PM'$.
            8. if $PM''$ is a complete match then
               9. Put $PM''$ into the answer set $RS$.
            else
               10. Put $PM''$ into $MS'$.
      11. Insert $MS'$ into $MS$.
12. Return $RS$.

5.2.2 Finding the Optimal Partitioning

Obviously, different partitionings lead to different join costs. Thus, we define a cost-based “best” partitioning in this section.

Definition 5.4. Join Cost. Given a query graph $Q$ with $n$ vertices $v_1, \ldots, v_n$ and a partitioning $\mathcal{P} = \{P_1, \ldots, P_n\}$ over all local partial matches $\Omega$, the join cost is defined as follows.

$$Cost(\Omega) = O(\prod_{i=1}^{n-1} (|P_i| + 1))$$

where $|P_i|$ is the number of local partial matches in $P_i$ and $1$ is introduced to avoid the “0 element” in the product.

Definition 5.4 assumes that each pair of local partial matches (from different partitions) are joinable, which is the worst case, so that we can quantify the worst-case performance. For example, the cost of the partitioning in Figure 6 is $6 \times 3 \times 4 = 72$. Of course, more sophisticated and more realistic cost functions can be used instead, but, determining the most appropriate cost function is a major research issue in itself and outside the scope of this paper. We leave it as an interesting future research to study.

The key issue is how to find the optimal partitioning (i.e., the partitioning with the minimal join cost). A naive enumeration method has $O(n^{2n})$ time complexity, where $n$ is the number of vertices in query $Q$ and $\Omega$ denotes all local partial matches. Obviously, the naive enumeration is not efficient, since $|\Omega|$ is very large in practice. Therefore, we need to find better solutions for finding the optimal partitioning.

Actually, we can reduce a 0-1 integer planning problem (a classical NP-complete problem) to finding the optimal partitioning. Thus, finding the optimal partitioning is a NP-complete problem.

Theorem 5.2. Finding the optimal partitioning (i.e., the join cost defined in Definition 5.4 is minimal) is a NP-complete problem.

Even that finding the optimal partitioning is a NP-complete problem, we can still find a much faster algorithm with time complexity $O(2^n \times |\Omega|)$. Note that this is a fixed-parameter tractable algorithm since $n$ is small in practice. For example, $n$ is no larger than 10 in most benchmark SPARQL queries (such as LUBM benchmarks [10]).

Our algorithm is based on the following characteristic of the optimal partitioning. Consider a query graph $Q$ with $n$ vertices $v_1, \ldots, v_n$. Let $U_{ij}$ ($i = 1, \ldots, n$) denote all local partial matches (in $\Omega$) that have internal vertices matching $v_j$. Unlike the partitioning defined in Definition 5.3, $U_{ij}$ and $U_{ji}$ (1 $\leq i \neq j \leq n$) may have overlaps. For example, $PM_2$ in Figure 4 contains an internal vertex 002 that matches $v_1$, thus, $PM_2$ is in $U_{1j}$, $PM_2$ also has internal vertex 010 that matches $v_2$, thus, $PM_2$ is also in $U_{2j}$. However, the partitioning defined in Definition 5.3 does not allow overlapping partitions.

If a partitioning $\mathcal{P}_{opt} = \{P_1, \ldots, P_n\}$ is optimal, the following theorem holds.

Theorem 5.3. Given a query graph $Q$ with $n$ vertices $v_1, \ldots, v_n$ and a set of all local partial matches $\Omega$, let $U_{ij}$ ($i = 1, \ldots, n$) be all local partial matches (in $\Omega$) that have internal vertices matching $v_j$. For the optimal partitioning $\mathcal{P}_{opt} = \{P_1, \ldots, P_n\}$ where $P_n$ has the largest partition size (i.e., the number of local partial matches in $P_n$ is maximum) in $\mathcal{P}_{opt}$, $P_n = U_{ij}$.

Based on Theorem 5.3, we propose a dynamic programming algorithm. The intuition of the algorithm is as follows. Assume that the optimal partitioning is $\mathcal{P}_{opt} = \{P_1, P_2, \ldots, P_n\}$. We record the partitions in $\mathcal{P}_{opt}$ in non-descending order of sizes, i.e., $\mathcal{P}_{opt} = \{P_1, \ldots, P_n\}$, $|P_i| \geq |P_{i+1}| \geq \ldots \geq |P_n|$. According to Theorem 5.3, $P_n = U_{ij}$. Let us denote by $\Omega_{U_{ij}}$ all local partial matches (in $\Omega$) that do not contain internal vertices matching $v_k$, i.e., $\Omega_{U_{ij}} = \Omega - U_{ij}$. It is straightforward to have the following optimal substructure in Equation 2.

$$\begin{align*}
\text{Cost}(\Omega_{opt}) &= |P_{v_j}| \times \text{Cost}(\Omega_{U_{ij}}) \\
&= |U_{ij}| \times \text{Cost}(\Omega_{U_{ij}}) \\
\end{align*}$$

Since we do not know which vertex is $v_{ij}$, we introduce the following optimal substructure that is used in our dynamic programming algorithm.

$$\begin{align*}
\text{Cost}(\Omega_{ij}) &= \text{MIN}_{1 \leq l < j} (|P_{v_j}| \times \text{Cost}(\Omega_{ij-l})) \\
&= \text{MIN}_{1 \leq l < n} (|U_{ij}| \times \text{Cost}(\Omega_{U_{ij-l}})) \\
\end{align*}$$

Obviously, it is easy to design a naive dynamic programming algorithm based on Equation 3. However, it can be further optimized by recording

An algorithm is called fixed-parameter tractable for a problem of size $l$, with respect to a parameter $n$, means it can be solved in time $O(f(n)g(l))$, where $f(n)$ can be any function but $g(l)$ must be polynomial.

A problem is said to have optimal substructure if an optimal solution can be constructed efficiently from optimal solutions of its subproblems [6]. This property is often used in dynamic programming formulations.
some intermediate results. Based on Equation [3] we can prove the following equation.

\[
\text{Cost}(\Omega_{\text{opt}}) = \text{MIN}_{\text{edges}} \{ |P_i| \times |P_j| \times \text{Cost}(\Omega_{\text{opt}}) \} = \text{MIN}_{\text{edges}} \{ |U_i| \times |U_j| \times \text{Cost}(\Omega_{\text{opt}}) \}
\]

(4)

where \( \Omega_{\text{opt}} \) denotes all local partial matches that do not contain internal vertices matching \( v_i \) or \( v_j \), and \( U_i \) denotes all local partial matches (in \( \Omega \)) that contain internal vertices matching vertex \( v_j \).

However, if Equation (4) is used naively in the dynamic programming formulation, it would result in repeated computations. For example, \( \text{Cost}(\Omega_{\text{opt}}) \) will be computed twice in both \( |U_i| \times |U_j| \times \text{Cost}(\Omega_{\text{opt}}) \) and \( |U_j| \times |U_i| \times \text{Cost}(\Omega_{\text{opt}}) \). To avoid this, we introduce a map that records \( \Omega \), which have been already calculated (Line 16 in Function OptComCost), so that subsequent uses of \( \text{Cost}(\Omega) \) can be serviced directly from searching the map (Lines 8-10 in Function ComCost).

We can prove that there are \( 2^n \) items in the map (worst case), where \( n = |V(Q)| \). Thus, the time complexity of the algorithm is \( (2^n \times |Q|) \). Although the time complexity is still exponential, (i.e., \(|V(Q)|\)) is small in practice. For example, \( n \) is no larger than 10 in most benchmark queries such as LUBM benchmarks [10].

5.2.3 Join Order

When we determine the optimal partitioning of local partial matches, the join order is also determined. If the optimal partitioning is \( \mathcal{P}_{\text{opt}} = \{P_{i_1}, ..., P_{i_k}\} \) and \( |P_{i_1}| \geq |P_{i_2}| \geq ... \geq |P_{i_k}| \), then the join order must be \( P_{i_1} \bowtie P_{i_2} \bowtie ... \bowtie P_{i_k} \). The reasons are as follows.

First, changing the join order often cannot prune any intermediate results. Let us recall the example optimal partitioning \( \{P_{i_1}, P_{i_2}, P_{i_3}\} \) shown in Figure 6. The join order should be \( P_{i_1} \bowtie P_{i_2} \bowtie P_{i_3} \), and any changes in the join order can not prune intermediate results. For example, if we first join \( P_{i_1} \) with \( P_{i_2} \), we can not prune the local partial matches in \( P_{i_2} \) that can not join with any local partial matches in \( P_{i_1} \). This is because there may be some local partial matches \( P_{i_3} \) that have an internal vertex matching \( v_i \) and can join with local partial matches in \( P_{i_1} \). In other words, the results of \( P_{i_1} \bowtie P_{i_2} \bowtie P_{i_3} \) is not smaller than \( P_{i_2} \bowtie P_{i_3} \). Similarly, any other changes of the join order of the partitioning have no effects.

Second, in some special cases, the join order may have an effect on the performance. Given a partitioning \( \mathcal{P}_{\text{opt}} = \{P_{i_1}, ..., P_{i_k}\} \) and \( |P_{i_1}| \geq |P_{i_2}| \geq ... \geq |P_{i_k}| \), if the set of the first \( n' \) vertices, \( \{v_{i_1}, v_{i_2}, ..., v_{i_{n'}}\} \), is a vertex cut of the query graph, the join order for the remaining \( n - n' \) partitions has an effect. In the worst case, if the query graph is a complete graph, the join order has an effect on the performance.

5.3 Distributed Assembly

An alternative to centralized assembly is to assemble the local partial matches in a distributed fashion. We adopt Bulk Synchronous Parallel (BSP) model [31] to design our synchronous algorithm for distributed assembly. A BSP system consists of processors connected by a communication network. A BSP computation proceeds in a series of global supersteps. A superstep consists of three components: local computation, communication and barrier synchronisation. In the following, we discuss how we apply this strategy to distributed assembly.

Local Computation. Each processor performs some computation based on the data stored in the local memory. The computations on different processors are independent in the sense that different processors perform the computation in parallel.

Consider the \( m \)-th superstep. For each fragment \( F_i \), let \( \Delta^m_m(F_i) \) denote all received intermediate results in the \( m \)-th superstep and \( \Omega^m_m(F_i) \) denote all local partial matches and the intermediate results generated the first \((m - 1)\) supersteps. In the \( m \)-th superstep, we join local partial matches in \( \Delta^m_m(F_i) \) with local partial matches in \( \Omega^m_m(F_j) \) by Algorithm 4. The basic idea of Algorithm 4 is similar to Algorithm 2. It is an iterative algorithm. For each intermediate result \( PM \), we check if it can join with some local partial match \( PM' = PM \bowtie PM' \) is a complete crossing match, it is returned. If the join result \( PM'' = PM' \bowtie PM'' \) is an intermediate result, we will check if \( PM'' \) can further join with another local partial match in \( \Omega^m_m(F_j) \cup \Delta^m_m(F_j) \) in the next iteration. We also insert the intermediate result \( PM'' \) into \( \Delta^m_m(F_j) \) that will be sent to other fragments in the communication step discussed below. Of course, we can also use the partitioning-based solution (in Section 5.2.1) to optimize join processing, but we do not discuss that due to space limitation.

Algorithm 4: Local Computation in Each Fragment \( F_i \)

\[
\text{Input: } \Omega^m_m(F_i), \text{ the local partial matches in fragment } F_i
\]
\[
\text{Output: } RS, \text{ the crossing matches found at this superstep; } \Delta^m_m(F_i), \text{ the intermediate results that will be sent}
\]
\begin{enumerate}
  \item Let \( \Omega = \Omega^m_m(F_i) \cup \Delta^m_m(F_i) \)
  \item Set \( MS = \Delta^m_m(F_i) \)
  \item for \( N = 1 \) to \(|V(Q)|\) do
    \begin{enumerate}
      \item if \(|MS| = 0\) then
        \begin{enumerate}
          \item Break;
        \end{enumerate}
      \end{enumerate}
      \begin{enumerate}
        \item Set \( MS' = \phi \)
        \item for each local partial match \( PM \) in \( MS \) do
          \begin{enumerate}
            \item if \( PM \bowtie PM' \) are joinable then
              \begin{enumerate}
                \item \( PM'' = PM' \bowtie PM'' \)
                \item if \( PM'' \bowtie PM'' \) is a SPARQL match then
                  \begin{enumerate}
                    \item Put \( PM'' \bowtie PM'' \) into the answer set \( RS \)
                  \end{enumerate}
                \end{enumerate}
            \end{enumerate}
          \end{enumerate}
      \end{enumerate}
      \begin{enumerate}
        \item Insert \( MS' \) into \( \Delta^m_m(F_i) \)
        \item Clear \( MS \) and \( MS' \leftarrow MS' \)
      \end{enumerate}
    \end{enumerate}
  \end{enumerate}
\]
\[
\Omega^{m+1}_m(F_i) = \Omega^m_m(F_i) \cup \Delta^m_m(F_i)
\]
\[
\text{Return } RS \text{ and } \Delta^m_m(F_i)
\]

Communication. All processors exchange data among themselves. Consider the \( m \)-th superstep. The straightforward communication strategy works as follows. If an intermediate result \( PM \) in \( \Delta^m_m(F_i) \) shares a crossing edge with fragment \( F_j \), \( PM \) will be sent to site \( S_j \) from \( S_i \) (assuming fragments \( F_i \) and \( F_j \) are stored in sites \( S_i \) and \( S_j \), respectively).

However, the above communication strategy may generate duplicate results. For example, as shown in Figure 4 we can assemble \( PM_i \) (at site \( S_i \)) and \( PM'_i \) (at site \( S'_i \)) to form a complete crossing match. According to the straightforward communication strategy, \( PM'_i \) will be sent to \( S_j \) from \( S_i \). Then, we can get \( PM_i'', PM'_i \bowtie PM''_i \bowtie PM'_i \) at \( S_j \). Analogously, we also send \( PM'_i \) from \( S_j \) to \( S_j \) and assemble them at site \( S_i \). In other words, we obtain the join result \( PM_i'', PM'_i \bowtie PM''_i \bowtie PM'_i \) at both sites \( S_i \) and \( S_j \). This wastes resources and increases total evaluation time.

To avoid duplicate result computation, we introduce a ‘‘divide-and-conquer’’ approach. We define a total order of fragment \( F_i \) \((i = 1, ..., k)\) in a non-descending order of \( |Q(F_i)| \), i.e., the number of local partial matches in fragment \( F_i \) found at the partial evaluation stage. Formally, the definition is as follows.

**Definition 5.5.** Given any two fragments \( F_i \) and \( F_j \), \( F_i < F_j \) if and only if \(|Q(F_i)| \leq |Q(F_j)| \) \((1 \leq i, j \leq n)\).
Without loss of generality, we assume that $F_1 < F_2 < ... < F_n$ in the remainder. The basic idea of the divide-and-conquer approach is as follows. Assume that a crossing match $M$ is formed by joining local partial matches that are from different fragments $F_1, ..., F_n$, where $F_1 < F_2 < ... < F_n$, and $1 ≤ i_1, ..., i_m ≤ n$. The crossing match should only be generated at fragment site $S_{i_m}$ rather than other fragment sites.

For example, at site $S_2$, we generate crossing matches by joining local partial matches from $F_1$ and $F_2$. The crossing matches generated at $S_2$ should not contain any local partial match from $F_3$ or even larger fragments (such as $F_4, ..., F_n$). Similarly, at site $S_3$, we should generate crossing matches by joining local partial matches from $F_3$ and fragments smaller than $F_3$. The crossing matches should not contain any local partial match from $F_4$ or even larger fragments (such as $F_5, ..., F_n$). As mentioned earlier, we define the total order for fragments $F_i$ according to the non-descending order of $|\Omega(F_i)|$. The basic intuition is the “load balancing” of each site.

The above “divide-and-conquer” framework can avoid duplicate results, since each crossing match can be only generated at a single site according to the “dividing search space”. To enable the “divide-and-conquer” framework, we need to introduce some constraints over data communication. The transmission of local partial matches from each site $Si$ to $Sj$ is allowed only if $Fi < Fj$. Let us consider an intermediate result PM in $\Omega_m(F_i)$. Assume $PM$ is generated by joining intermediate results from $m$ different fragments $F_1, ..., F_m$, where $F_1 < F_2 < ... < F_m$. We send $PM$ to another fragment $F_i$ if and only if the two conditions hold: 1) $F_i > F_m$; and 2) $F_i$ shares common crossing edges with at least one fragment of $F_1, ..., F_m$.

Barrier Synchronization. All communication in the $m$-th superstep should finish before entering into the $(m + 1)$-th superstep.

We now discuss the initial state (i.e., 0-th superstep) and the system termination condition.

Initial State. In the 0-th superstep, each fragment $Fi$ has only local partial matches in $Fi$, i.e., $\Omega_0$. Since it is impossible to assemble local partial matches in the same fragment, the 0-th superstep requires no local computation. It enters the communication stage directly. Each site $Si$ sends $\Omega_0$ to other fragments according to the communication strategy that has been discussed before.

System Termination Condition. A key problem in the BSP algorithm is the number of the supersteps to terminate the system. In order to facilitate the analysis, we propose utilizing a fragmentation graph topology graph.

Definition 5.6. (Fragmentation Topology Graph) Given a fragmentation $\mathcal{F} = \{F_1, F_2, ..., F_n\}$ over an RDF graph $G$, the corresponding fragmentation topology graph $T$ is defined as follows: 1) Each node in $T$ is a fragment $F_i$, $i = 1, ..., k$; 2) There is an edge between nodes $F_i$ and $F_j$ in $T$, $1 ≤ i ≠ j ≤ n$, if and only if there is at least one crossing edge between $F_i$ and $F_j$ in RDF graph $G$.

Let $Dia(T)$ be the diameter of $T$. We need at most $Dia(T)$ supersteps to transfer the local partial matches in one fragment $Fi$ to any other fragment $Fj$. Hence, the number of the supersteps in the BSP-based algorithm is $Dia(T)$.

6. EXPERIMENTS

We evaluate our method over both real and synthetic RDF datasets, and compare our approach with the state-of-the-art distributed RDF systems, including both federated SPARQL query systems (DARQ [27], FedEx [28] and Q-Tree [26]) and RDF data partition-based methods (GraphPartition [15] and EAGRE [33]).

Setting. We use one synthetic dataset with different sizes, and two real datasets in our experiments. Table 1 summarizes the statistics of these datasets. All sample queries are shown in Appendix.

1) LUBM [10] is a benchmark that adopts an ontology for the university domain, and can generate synthetic OWL data scalable to an arbitrary size. We vary the university number from 1000 to 10000. The number of triples is 133 million to 1.33 billion. On these experiments, we partition the LUBM datasets according to the university identifiers. We use the 7 benchmark queries in [2] to evaluate our method.

2) BTC 2012 (http://km.aifb.kit.edu/projects/btc-2012/) is a dataset that serves as basis for submissions to the Billion Triples Track of the Semantic Web Challenge. After eliminating all redundant triples, this dataset contains about 1 billion triples. We use METIS [19], a popular min-cut based graph partition algorithm, to partition it into several fragments. In addition, we use 7 queries in [34].

3) LOD is a dataset generated from ten datasets in Linked Open Data to further evaluate our methods in a real federated database system environment. We download ten datasets (DBpedia, LinkedMDB, Yago2, FreeBase, NYTimes, MusicBrainz, Diseasome, Geonames, Timbl, LinkedGeoData). By randomly selecting one as the primary and others as the copies of the primary, we merge entities in different datasets connected using predefined property “owl:sameAs”. Each dataset maps to a fragment. We also define six sample queries across different fragments.

| Dataset | Number of Triples | RDF N3 File Size(KB) | Number of Entities |
|---------|-------------------|-----------------------|--------------------|
| LUBM 1000 | 133,553,834 | 35,136,798 | 21,715,108 |
| LUBM 5000 | 667,292,653 | 76,522,983 | 108,498,627 |
| LUBM 10000 | 1,344,281,167 | 153,256,699 | 217,006,852 |
| BTC | 1,086,184,911 | 238,970,296 | 183,233,054 |
| LOD | 1,055,264,649 | 169,175,000 | 146,228,189 |

Table 1: Datasets

We conduct all experiments on a cluster of 10 machines running Linux, each of which has one CPU with four cores of 3.06GHz, 16GB memory and 500GB disk storage. At each site, we install gStore [15] to find inner matches, since it supports the graph-based SPARQL evaluation paradigm. We revise gStore to find all local partial matches in each fragment as discussed in Section 4. All implementations are in standard C++. We use MPICH-3.0.4 library for communication.

Exp 1. Evaluating Each Stage’s Performance. In this experiment, we study the performance of our system in each stage (i.e., partial evaluation and assembly process) with regard to different queries in LUBM1000. We report the running time of each stage (i.e., partial evaluation and assembly) and the number of local partial matches, inner matches, crossing matches, with regard to different queries in Table 2. We also compare the centralized and distributed assembly strategies. Note that we classify SPARQL queries into three categories according to query graphs’ structures: star, snowflake (several stars linked by a path) and complex (a combination of the above with complex structure).

Partial Evaluation: Table 2 shows that if there are some consistent entities (such as “http://www.Department.University0.edu” in $Q_6$ of LUBM) in the query, the partial evaluation is much faster than others. Our partial evaluation algorithm (Algorithm 1) is based on a state transformation, while the consistent entities can reduce the search space. Furthermore, the running time also depends on the number of inner matches and local partial matches, as shown in Table 2. More inner matches and local partial matches lead to higher running time in the partial evaluation stage.

Assembly: In this experiment, we compare centralized and distributed assembly approaches. Obviously, there is no assembly process for a star query. Thus, we only study the performances of snowflake and complex queries. We find that distributed assembly can beat the centralized one when there are lots of local partial
matches and crossing matches. The reason is as follows: in centralized assembly, all local partial matches need to be sent to the server where they are assembled. Obviously, if there are lots of local partial matches, the server becomes the bottleneck. However, in distributed assembly, we can take advantage of the parallelization to speed up both the network communication and assembly. For example, in Q2, there are 167621 local partial matches. It takes a long time to transfer the local partial matches to the server and assemble them in the server in centralized assembly. So, distributed assembly outperforms the centralized one by six times. However, if the number of local partial matches and the number of crossing matches are small, the barrier synchronisation cost dominates the total cost in distributed assembly. In this case, the advantage of distributed assembly is not clear. A quantitative comparison between distributed assembly and centralized one needs more statistics about the network communication, CPU and other parameters. A sophisticated quantitative study is beyond the scope of this paper and is left as future work.

**Exp 2: Evaluating Optimizations in Assembly** In this experiment, we evaluate two different optimization techniques in the assembly: partition-based join strategy (Section 5.1) and the divide-and-conquer approach in the distributed assembly (Section 5.3). Star queries Q2, Q4, and Q6 do not need the assembly process. Query Q3 does not have any match in RDF graph G. Therefore, we only use the three benchmark queries Q2, Q3, and Q6 in our experiments.

**Partitioning-based Join.** First, we compare partitioning-based join (i.e., Algorithm 3) with naive join processing (i.e., Algorithm 2) in Table 3 which shows that the partitioning-based strategy can greatly reduce the join cost. Second, we evaluate the effectiveness of our cost model. Note that the join order depends on the partitioning strategy, which is based on our cost model as discussed in Section 5.2. In other words, once the partitioning is given, the join order is fixed. So, we use the cost model to find the optimal partitioning and report the running time of the assembly process in Table 3. We find that the assembly with optimal partitioning is faster than that with the random partitioning, which confirms the effectiveness of our cost model.

| Table 3: Running Time of Partitioning-based Join VS. Naive Join (in ms) |
|-------------------------|-------------------------|-------------------------|
| Partitioning-based Join | Partitioning-based Join | Naive Join |
| Based on the Optimal Partitioning | Based on the Random Partitioning | |
| Q2 | 53 | 55 | 3180 |
| Q3 | 164 | 177 | 15382 |
| Q6 | 211670 | 236012 | 10876991 |

| Table 2: Evaluation of Each Stage |
|----------------------------------|
| Time(in ms) | # of LPMs | # of IMs | Centralized Assembly | Distributed Assembly | Total Assembly |
|-------|----------|---------|----------------------|---------------------|------------------|
| Star Q2 | 1818 | 0 | 1081187 | 0 | 0 | 0 | 1818 | 1818 | 1081187 |
| Q3 | 82 | 0 | 10 | 0 | 0 | 0 | 82 | 82 | 10 |
| Q6 | 8 | 0 | 10 | 0 | 0 | 0 | 8 | 8 | 10 |
| Snowflake Q2 | 158 | 6707 | 110 | 164 | 125 | 15 | 322 | 283 | 125 |
| Complex Q2 | 52548 | 3033 | 2524 | 53 | 60 | 4 | 52601 | 52608 | 2528 |
| Q6 | 920 | 3358 | 0 | 36 | 48 | 0 | 956 | 968 | 0 |
| Q6 | 3945 | 107021 | 42479 | 211670 | 35856 | 1709 | 215615 | 39801 | 44190 |

1. # of Q means that the query involves some consistent entities.
2. # of LPMs means the number of local partial matches.
3. # of IMs means the number of inner matches.
4. # of CMs means the number of crossing matches.
5. # of Matches means the number of matches.

**Exp 3: Scalability Test.** In this experiment, we vary the RDF dataset size from 133 million triples (LUBM 1000) to 1.3 billion triples (LUBM 10000) to study the scalability of our methods. Table 5 shows the performance of different queries of both centralized and distributed assembly. From LUBM1000 to LUBM10000, the data size increases 10 times and this causes the query response times to increase 10-20 times. The increase is also affected by the query graph's shape. For star queries, the query response time increases proportional to the datasize, as shown in Table 5. For snowflake/complex queries, the query response times may grow faster than the data size, as shown in Table 5. Especially for Q2, the query response time increases 20 times as the datasize increases 10 times. This is because the complex query graph shape causes more complex operations in query processing, such as joining and assembly. However, even for complex queries, the growth of the query response times are approximatively linear with the RDF graph size.

Note that, as discussed in Experiment 1, the assembly times of star queries (Q2, Q3, and Q6) are 0. Therefore, the query response times for star queries in centralized and distributed assembly are the same. In contrast, for other shapes of queries, some local partial matches and crossing matches result in some differences between the performances of centralized and distributed assembly. Here, Q2 is a special case. Although it is a complex query, there are few local partial matches for Q2. Furthermore, the crossing match number is 0 in Q2 (in Table 2). Therefore, the assembly times for Q2 are so small that the query response times in both centralized and distributed assembly are the almost same.

**Exp 4: Comparing with State-of-the-Art Algorithms.** In this experiment, we compare our approach with some state-of-the-art distributed RDF systems, including both federated SPARQL query systems (DARQ [27], FedX [28] and Q-Tree [26]) and RDF data partition-based methods (GraphPartition [15] and EAGRE [24]).

Comparing with Federated SPARQL queries. For further evaluating our methods, we download ten datasets (DBPedia, Linked-MDB, Yago2, FreeBase, NYTimes, MusicBrainz, Diseaseome, Geonames, Timbl, LinkedGeoData) from LOD to integrate a real LOD
dataset. Each dataset maps to a fragment and is stored at a site. We also define six queries across different fragments. Two of these six queries (Q₁ and Q₂) are stars, two (Q₃ and Q₄) are snowflakes and two (Q₅ and Q₆) are complex queries. Furthermore, three queries (Q₁, Q₃, and Q₄) contain selective triple patterns, while the others (Q₂, Q₅, and Q₆) do not. We report these six sample queries in Appendix.

Table 5: Scalability Test (in ms)

| Query | LUBM 1000 | LUBM 5000 | LUBM 10000 |
|-------|-----------|-----------|------------|
| Q₁    | 52601     | 274320    | 563037     |
| Q₂    | 52608     | 286320    | 588031     |
| Q₃    | 1818      | 10557     | 22647      |
| Q₄    | 1818      | 10557     | 22647      |
| Q₅    | 956       | 5196      | 10239      |
| Q₆    | 968       | 5261      | 10368      |

We compare our method with FedX and DARQ in the LOD dataset in Figure 7 which shows that our method outperforms FedX and DARQ by an order of magnitude for queries Q₁, Q₂, and Q₆.

There are two reasons for this phenomena. First, both GraphPartition and EAGRE are based on query decomposition. A query is decomposed into a set of subqueries and each subquery is performed locally on each fragment. They do not consider the whole query graph structure to prune the intermediate results. However, in our method, each fragment receives the whole query graph, and can utilize the query graph structure to prune more false positive intermediate results. Second, both GraphPartition and EAGRE are implemented on Hadoop while our system is running with MPI, a lightweight parallel library. According to our experiments, starting Hadoop jobs always spends 20-30 seconds.

In particular, for star queries (such as Q₁, Q₂, and Q₃ of LUBM), since their diameters are not larger than 2, GraphPartition can answer all these queries entirely without data communication in Hadoop. Therefore, the query response time of GraphPartition for these queries is as fast as our methods. However, for snowflake/complex queries whose diameters are larger than 2, GraphPartition needs to decompose the subqueries and join the intermediate results. Then, the query response time grows greatly. The performance differences also depend on the intermediate result sizes. Q₁ has more intermediate results, thus, the performance difference is not clear comparing with GraphPartition and EAGRE.

Finally, our method does not depend any specific graph partitioning strategy (which is known to be NP-hard). However, the query algorithms of GraphPartition and EAGRE depend on their own graph partitions. Therefore, neither GraphPartition nor EAGRE can be used directly in the virtually integrated distributed RDF graphs, such as LOD data, since there is no possibility of changing the partitioning of the data.
existing distributed SPARQL query algorithms, since they require re-partitioning the whole RDF graph as their own methods require. In contrast, in LOD, the RDF data are already distributed over different sites, thus, re-partitioning and redistributing are usually not possible. Our experiments show that the proposed algorithms perform very well. It does require minor modifications to SPARQL endpoints to evaluate partial queries.

There are a number of extensions we are currently working on. An important one is handling data at sites that do not have query capability (i.e., they are not SPARQL endpoints). Data at these sites need to be moved for processing that will affect the algorithm and cost functions. Furthermore, similarity search over LOD is also an ongoing work. Users may not specify the exact SPARQL query statements, because they have no full knowledge of the underlying schema of RDF graph in LOD.

8. REFERENCES

[1] D. J. Abadi, A. Marcus, S. Madden, and K. Hollenbach. Sw-store: a vertically partitioned dbms for semantic web data management. VLDB J., 18(2):385–406, 2009.

[2] M. Atre, V. Chaoji, M. J. Zaki, and J. A. Hendler. Matrix “bit” loaded: a scalable lightweight join query processor for RDF data. In WWW, pages 41–50, 2010.

[3] P. Buneman, G. Cong, W. Fan, and A. Kementsietsidis. Using partial evaluation in distributed query evaluation. In VLDB, 2006.

[4] G. Cong, W. Fan, and A. Kementsietsidis. Distributed query evaluation with performance guarantees. In SIGMOD Conference, pages 509–520, 2007.

[5] G. Cong, W. Fan, A. Kementsietsidis, J. Li, and X. Liu. Partial evaluation for distributed xpath query processing and beyond. ACM Trans. Database Syst., 37(4):32, 2012.

[6] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein. Introduction to Algorithms (3. ed.). MIT Press, 2009.

[7] R. G. Downey, M. R. Fellows, A. Vardy, and G. Whittle. The parameterized complexity of some fundamental problems in coding theory. SIAM J. Comput., 29(2):545–570, 1999.

[8] W. Fan, X. Wang, and Y. Wu. Performance guarantees for distributed reachability queries. PVLDB, 5(11):1304–1315, 2012.

[9] W. Fan, X. Wang, Y. Wu, and D. Deng. Distributed graph simulation: Impossibility and possibility. PVLDB, 7(12):1083–1094, 2014.

[10] Y. Guo, Z. Pan, and J. Hefflin. Lubm: A benchmark for owl knowledge base systems. J. Web Sem., 3(2):158–182, 2005.

[11] S. Gurajada, S. Seufert, I. Miliaraki, and M. Theobald. Triad: a distributed rdf graph engine for web scale RDF data. PVLDB, 6(4):265–276, 2013.

[12] Y. Guo, Z. Pan, and J. Heflin. Lubm: A benchmark for owl knowledge base systems. J. Web Sem., 3(2):158–182, 2005.

[13] S. Gurajada, S. Seufert, I. Miliaraki, and M. Theobald. Triad: a distributed shared-nothing RDF engine based on asynchronous message passing. In SIGMOD Conference, pages 289–300, 2014.

[14] A. Harth, K. Hose, M. Karnstedt, A. Polleres, K.-U. Satller, and J. Umbrich. Data summaries for on-demand queries over linked data. In WWW, pages 411–420, 2010.

[15] O. Hartig, C. Bizer, and J.-C. Freytag. Executing sparql queries over the web of linked data. In Proc. Int. Conf. on Semantic Web, pages 293–309, 2009.

[16] O. Hartig and M. T. Özas. Linked data query processing (Tutorial). In Proc. 30th IEEE Int. Conf. on Data Engineering, pages 1286–1289, 2014.

[17] J. Huang, D. J. Abadi, and K. Ren. Scalable sparql querying of large RDF graphs. PVLDB, 4(11):1123–1134, 2011.

[18] M. F. Husain, J. P. McGlothlin, M. M. Masud, L. R. Khan, and B. M. Thuraisingham. Heuristics-based query processing for large RDF graphs using cloud computing. IEEE Trans. Knowl. Data Eng., 23(9):1312–1327, 2011.

[19] N. D. Jones. An introduction to partial evaluation. ACM Comput. Surv., 28(3):480–503, 1996.

[20] Z. Kaoudi and I. Manolescu. RDF in the clouds: A survey. VLDB J., 2014.

[21] G. Karypis and V. Kumar. Analysis of multilevel graph partitioning. In SC, 1995.

[22] V. Khadilkar, M. Kantarciglu, B. M. Thuraisingham, and P. Castagna. Jena-HBase: A distributed, scalable and efficient RDF triple store. In International Semantic Web Conference (Posters & Demos), 2012.

[23] K. Lee and L. Liu. Scaling queries over big RDF graphs with semantic hash partitioning. PVLDB, 6(14):1894–1905, 2013.

[24] K. Lee, L. Liu, Y. Tang, Q. Zhang, and Y. Zhou. Efficient and customizable data partitioning framework for distributed big RDF data processing in the cloud. In IEEE CLOUD, pages 327–334, 2013.

[25] J. Huang, D. J. Abadi, and K. Ren. Scalable sparql querying of large RDF graphs. In WWW 2012, Lyon, France, April 16-20, 2012, pages 949–958, 2012.

[26] T. Neumann and G. Weikum. RDF-3x: A risc-style engine for rdf. PVLDB, 1(1):647–659, 2008.

[27] J. Pérez, M. Arenas, and C. Gutierrez. Semantics and complexity of SPARQL. ACM Trans. Database Syst., 34(3), 2009.

[28] F. Prasser, A. Kerper, and K. A. Kuhn. Efficient distributed query processing for autonomous RDF databases. In EDBT, pages 372–383, 2012.

[29] B. Quilitz and U. Leser. Querying distributed RDF data sources with SPARQL. In The Semantic Web: Research and Applications, 5th European Semantic Web Conference, ESWC 2008, Tenerife, Canary Islands, Spain, June 1–5, 2008, Proceedings, pages 524–538, 2008.

[30] A. Schwarte, P. Haase, K. Hose, R. Schenkel, and M. Schmidt. Fedx: A federation layer for distributed query processing on linked open data. In The Semantic Web: Research and Applications - 8th Extended Semantic Web Conference, ESWC 2011, Heraklion, Crete, Greece, May 29 - June 2, 2011, Proceedings, Part II, pages 481–486, 2011.

[31] H. Shang, Y. Zhang, X. Lin, and J. X. Yu. Taming verification hardness: an efficient algorithm for testing subgraph isomorphism. PVLDB, 1(1):364–375, 2008.

[32] B. Shao, H. Wang, and Y. Li. Trinity: a distributed graph engine on a memory cloud. In SIGMOD Conference, pages 505–516, 2013.

[33] L. G. Valiant. A bridging model for parallel computation. Commun. ACM, 33(8):103–111, 1990.

[34] K. Wilkinson, C. Sayers, H. A. Kuno, and D. Reynolds. Efficient rdf storage and retrieval in jena2. In 5WDB, pages 131–150, 2003.

[35] K. Zeng, J. Yang, H. Wang, B. Shao, and Z. Wang. A distributed graph engine for web scale RDF data. PVLDB, 6(4):265–276, 2013.

[36] X. Zhang, L. Chen, Y. Tong, and M. Wang. Eagle: Towards scalable i/o efficient sparql query evaluation on the cloud. In ICDE, pages 565–576, 2013.

[37] X. Zhang, L. Chen, and M. Wang. Towards efficient join processing over large rdf graph using mapreduce. In SSDBM, pages 250–259, 2012.

[38] L. Zou, M. T. Özasu, L. Chen, X. Shen, R. Huang, and D. Zhao. Answering pattern match queries in large graph databases via graph embedding. VLDB J., 2013.
APPENDIX

A. PROOFS

Theorem 4.1: Given any crossing match \( M \) of SPARQL query \( Q \) in an RDF graph \( G \), if \( M \) overlaps with some fragment \( F_i \) (\( i = 1, \ldots, k \)), the overlapping part between \( M \) and fragment \( F_i \), denoted as \( PM \), must be a local partial match in fragment \( F_i \).

Proof: Since \( M \) is a SPARQL query match and the overlapping part is a subset of the SPARQL match, it is easy to know that Conditions 1-6 of Definition 4.1 hold.

Since \( M \) is a crossing match, the overlapping part between \( M \) and \( F_i \) must contain a crossing edge, thus, Condition 7 holds.

Since the overlapping part \( PM \) is a subgraph of match \( M \), according to Claim 1, we can prove that Condition 8 holds.

If the overlapping part between \( M \) and fragment \( F_i \) has several weakly connected components, we consider each weakly connected component individually. Obviously, each weakly connected component satisfies Condition 9.

To summarize, the overlapping part between \( M \) and fragment \( F_i \) satisfies all conditions in Definition 4.1. Thus, Theorem 4.1 holds.

Theorem 4.2: The results generated by the partial-evaluation-and-assembly algorithm satisfy no-false-negative requirement if and only if all local partial matches in each fragment are found in the partial evaluation stage.

Proof: In two parts:

1) The “If” part: (proven by contradiction).

Assume that all local partial matches are found in each fragment \( F_i \) but a cross match \( M \) is missed in the answer set.

Since \( M \) is a crossing match, suppose that \( M \) overlaps with \( m \) fragments \( F_1, \ldots, F_m \). According to Theorem 4.1, the overlapping part between \( M \) and \( F_i \) (\( i = 1, \ldots, m \)) must be a local partial match \( PM_i \) in \( F_i \). According to the assumption, these local partial matches have been found in the partial evaluation stage. Obviously, we can assemble these partial matches \( PM_i (i = 1, \ldots, m) \) to form the complete cross match \( M \).

In other words, \( M \) would not be missed if all local partial matches are found. This contradicts the assumption.

2) The “Only If” part: (proved by contradiction).

We assume that a local partial match \( PM_i \) in fragment \( F_i \) is missed and the answer set still satisfy no-false-negative requirement.

Suppose that \( PM_i \) matches a part of \( Q \), denoted as \( Q' \). Assume that there exists another local partial match \( PM_j \) in \( F_j \) that matches a complementary graph of \( Q' \), denoted as \( Q'' \), i.e., \( Q'' \cup Q' = Q \). In this case, we can obtain a complete match \( M \) by assembling the two local partial matches. If \( PM_j \) in \( F_j \) is missed, then match \( M \) is missed. In other words, it cannot satisfy the no-false-negative requirement. This also contradicts the assumption.

Theorem 5.1: Given two local partial matches \( PM_i \) and \( PM_j \) from fragments \( F_i \) and \( F_j \) with functions \( f_i \) and \( f_j \), respectively, if there exists a query vertex \( v \) where both \( f_i(v) \) and \( f_j(v) \) are internal vertices of fragments \( F_i \) and \( F_j \), respectively, \( PM_i \) and \( PM_j \) cannot be joinable.

Proof: If \( f_i(v) \neq f_j(v) \), then a vertex \( v \) in query \( Q \) matches two different vertices in \( PM_i \) and \( PM_j \), respectively. Obviously, \( PM_i \) and \( PM_j \) cannot be joinable.

If \( f_i(v) = f_j(v) \), since \( f_i(v) \) and \( f_j(v) \) are both internal vertices, both \( PM_i \) and \( PM_j \) are from the same fragment. As mentioned earlier, it is impossible to assemble two local partial matches from the same fragment (see the first paragraph of Section 5.1). Thus, \( PM_i \) and \( PM_j \) cannot be joinable.

Theorem 5.2: Finding the optimal partition is NP-complete problem.

Proof: We can reduce a 0-1 integer planning problem to finding the optimal partition. We build a bipartite graph \( B \), which contains two vertex groups \( B_1 \) and \( B_2 \). Each vertex \( a_i \) in \( B_1 \) corresponds to a local partial match \( PM_i \) in \( \Omega \), \( j = 1, \ldots, n \). Each vertex \( b_i \) in \( B_2 \) corresponds to a query vertex \( v \), \( i = 0, \ldots, n \). We introduce an edge between \( a_i \) and \( b_j \) if and only if \( PM_i \) has an internal vertex that is matching query vertex \( b_j \). Let a variable \( x_{ij} \) denote the edge label of the edge \( a_i b_j \). Let us recall all local partial matches in Figure 4.

Figure 9 shows an example bipartite graph.

We define the 0-1 integer planning problem as follow: \( \min \prod_{i=0}^{n} (\sum_{j=0}^{n} x_{ij} + 1) \) s.t. \( \sum_{j=0}^{n} x_{ij} = 1 \)

The above equation means that each local partial match should be assigned to only one query vertex.

It is easy to know the equivalence between the 0-1 integer planning and finding the optimal partition. The former is a classical NP-complete problem. Thus, we know the theorem holds.

Theorem 5.3: Given a query graph \( Q \) with \( n \) vertices \( \{v_1, \ldots, v_n\} \) and a set of all local partial matches \( \Omega \), let \( U_{v_i} (i = 1, \ldots, n) \) be all local partial matches (in \( \Omega \)) that have internal vertices matching \( v_i \). For the optimal partitioning \( \mathcal{P}_{\text{opt}} = \{P_{v_1}, \ldots, P_{v_n}\} \) where \( P_{v_i} \) has the largest partition size (i.e., the number of local partial matches in \( P_{v_i} \) is maximum) in \( \mathcal{P}_{\text{opt}}, U_{v_{v_i}} = U_{v_i} \).

Proof: (by contradiction) Assume that \( P_{v_i} \neq U_{v_i} \) in the optimal partitioning \( \mathcal{P}_{\text{opt}} = \{P_{v_1}, \ldots, P_{v_n}\} \). Then, there exists a local

\[ a(P_{v_i}) = a(P_{v_i}) \]

\[ a(P_{v_i}) = a(P_{v_i}) \]

\[ a(P_{v_i}) = a(P_{v_i}) \]

\[ a(P_{v_i}) = a(P_{v_i}) \]
partial match $PM \notin P_v$, and $PM \in U_{v_n}$. We assume that $PM \in P_j$, $j \neq n$. The cost of $\mathcal{P}_{opt}$ is:

$$\text{Cost}(\mathcal{P}_{opt}) = \sum_{1 \leq j \leq m} (|P_j| + 1)$$

Since $PM \in U_{v_n}$, $PM$ has an internal vertex matching $v_n$. Hence, we can also put $PM$ into $P_{v_n}$. Then, we get a new partitioning $\mathcal{P}' = \{P_1, ..., P_{v_n} - \{PM\}, ..., P_v, U \cup \{PM\}\}$. The cost of the new partitioning is:

$$\text{Cost}(\mathcal{P}') = \sum_{1 \leq j \leq m} (|P_j| + 1)$$

Let $C = \sum_{1 \leq j \leq m} (|P_j| + 1)$, which exists in both Equations 5 and 6. Obviously, $C > 0$.

$$\text{Cost}(\mathcal{P}_{opt}) - \text{Cost}(\mathcal{P}') = C \times (|P_{v_n}| + 1) - C \times (|P_{v_n}| + 2) = C \times (-1)$$

Because $P_{v_n}$ is the largest partition in $\mathcal{P}_{opt}$, $|P_{v_n}| + 1 - |P_{v_n}| > 0$. Furthermore, $C > 0$. Hence, $\text{Cost}(\mathcal{P}_{opt}) - \text{Cost}(\mathcal{P}') > 0$, meaning that the optimal partitioning has larger cost. Obviously, this cannot happen.

Therefore, in the optimal partitioning $\mathcal{P}_{opt}$, we cannot find a local partial match $PM$, where $|P_{v_n}|$ is the largest. $PM \notin P_{v_n}$ and $PM \in U_{v_n}$. In other words, $P_{v_n} = U_{v_n}$ in the optimal partitioning.

B. QUERIES IN EXPERIMENTS

B.1 LUBM Queries

Q1: Select ?x, ?y, ?z where { ?x rdf:type ub:GraduateStudent. ?y rdf:type ub:University. ?z rdf:type ub:Department. ?x ub:memberOf ?y. ?x ub:underOrganizationOf ?y. ?x ub:undergraduateDegreeFrom ?y.}

Q2: Select ?x, ?y, ?z where { ?x rdf:type ub:Course. ?x ub:memberOf ?y. ?x ub:underOrganizationOf ?y. }

Q3: Select ?x where { ?x rdf:type ub:FullProfessor. ?x ub:memberOf ?y. ?x ub:underOrganizationOf ?y. ?x ub:ResearchGroup. }

Q4: Select ?x, ?y, ?z where { ?x rdf:type ub:Student. ?y rdf:type ub:Department. ?x ub:memberOf ?y. ?x ub:underOrganizationOf ?y. ?x rdf:type ub:Student. ?y rdf:type ub:Course. ?z rdf:type ub:Advisor. }

B.2 BTC Queries

Q1: Select ?a, ?l, ?t, ?o where { ?a ["Eiffel Tower" @en. ?a <http://www.w3.org/2003/01/geo/wgs84_pos#lat> ?l. ?a <http://www. w3.org/2003/01/geo/wgs84_pos#long> ?t. ?a <http://dbpedia. ontology/location> <http://dbpedia.org/resource/France>.]

Q2: Select ?b, ?p, ?bn where { ?a [] "Tim Berners-Lee" @en. ?a <http://dbpedia.org/resource/placeOfBirth> ?b. ?a <http://dbpedia.org/resource/name> ?bn.}

Q3: Select ?t, ?l, ?o where { ?a <http://dbpedia.org/resource/placeOfBirth> ?t. ?a <http://dbpedia.org/resource/MainInterest> ?o. ?a <http://dbpedia.org/resource/MainInterest> ?o. ?a <http://dbpedia.org/resource/placeOfInterest> ?o. ?a <http://dbpedia.org/resource/countries/US>.}

Q4: Select ?a, ?b, ?l where { ?a <http://dbpedia.org/resource/countries/US>. ?a <http://dbpedia.org/resource/countries/US>. ?a <http://dbpedia.org/resource/countries/US>. ?a <http://dbpedia.org/resource/countries/US>.}

Q5: Select ?a, ?b where { ?a <http://dbpedia.org/resource/countries/US>. ?a <http://dbpedia.org/resource/countries/US>. ?a <http://dbpedia.org/resource/countries/US>. ?a <http://dbpedia.org/resource/countries/US>.}

Q6: Select ?a, ?b, ?l where { ?a <http://dbpedia.org/resource/countries/US>. ?a <http://dbpedia.org/resource/countries/US>. ?a <http://dbpedia.org/resource/countries/US>. ?a <http://dbpedia.org/resource/countries/US>.}

B.3 LOD Queries

Q1: Select ?p where { ?p <http://www.mpii.de/yago/resource/hasOfficialLanguage> <http://www.mpii.de/yago/resource/English_language>. ?p <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> <http://dbpedia.org/ontology/Country>.}

Q2: Select ?p where { ?p <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> <http://www.mpii.de/yago/resource/wordnet_scientist_110560637>. ?p <http://dbpedia.org/ontology/nationality> <http://dbpedia.org/resource/United_States>.}

Q3: Select ?a where { ?a <http://www.mpii.de/yago/resource/hasBornIn> <http://www.mpii.de/yago/resource/Chicago>. ?a <http://www.w3.org/2003/01/geo/wgs84_pos#lat> ?l. ?a <http://www. w3.org/2003/01/geo/wgs84_pos#long> ?t. ?a <http://dbpedia.o nontology/Place>. ?a <http://www.mpii.de/yago/resource/mainInterest> ?o. ?a <http://www.mpii.de/yago/resource/mainInterest> ?o. ?a <http://www.mpii.de/yago/resource/mainInterest> ?o. ?a <http://www.mpii.de/yago/resource/mainInterest> ?o.}

Q4: Select ?a where { ?a <http://dbpedia.org/resource/placeOfResidence> ?a. ?a <http://dbpedia.org/resource/placeOfResidence> ?a. ?a <http://dbpedia.org/resource/placeOfResidence> ?a. ?a <http://dbpedia.org/resource/placeOfResidence> ?a.}

Q5: Select ?a, ?b where { ?a <http://dbpedia.org/resource/placeOfResidence> ?a. ?a <http://dbpedia.org/resource/placeOfResidence> ?a. ?a <http://dbpedia.org/resource/placeOfResidence> ?a. ?a <http://dbpedia.org/resource/placeOfResidence> ?a.}

Q6: Select ?a, ?b, ?l, ?o where { ?a <http://dbpedia.org/resource/placeOfResidence> ?a. ?a <http://dbpedia.org/resource/placeOfResidence> ?a. ?a <http://dbpedia.org/resource/placeOfResidence> ?a. ?a <http://dbpedia.org/resource/placeOfResidence> ?a.}
Q_6: Select distinct \( ?n1, ?n2 \) where
\[
\{ ?f \text{<http://www.w3.org/1999/02/22-rdf-syntax-ns#type> <http://dbpedia.org/ontology/Film>. ?f <http://dbpedia.org/ontology/starring> ?person2. ?person1 <http://www.mpii.de/yago/resource/isMarriedTo> ?person2. ?person2 <http://www.mpii.de/yago/resource/hasPreferredName> ?n2. ?f <http://data.linkedmdb.org/resource/movie/producer> ?person1. ?person1 <http://data.linkedmdb.org/resource/movie/producer_name> ?n1. \}
\]

### B.4 Queries with Different Diameters in Experiment 1

| Diameter | Queries |
|----------|---------|
| 1 | Select \( ?x, ?y \) where \( ?x \text{ ub:undergraduateDegreeFrom http://www.University0.edu.} \) |
| 2 | Select \( ?x, ?y \) where \( ?x \text{ ub:advisor ?a.} \) \( ?x \text{ ub:undergraduateDegreeFrom http://www.University0.edu.} \) |
| 3 | Select \( ?x, ?y \) where \( ?x \text{ ub:advisor ?a.} \) \( ?x \text{ ub:undergraduateDegreeFrom ?y.} \) \( ?y \text{ ub:name University0.} \) |
| 4 | Select \( ?x, ?y \) where \( ?a \text{ ub:teacherOf ?c.} \) \( ?x \text{ ub:advisor ?a.} \) \( ?x \text{ ub:undergraduateDegreeFrom ?y.} \) \( ?y \text{ ub:name University0.} \) |

Table 6: Queries with Different Diameters