PruneJuice: Pruning Trillion-edge Graphs to a Precise Pattern-Matching Solution

Tahsin Reza\textsuperscript{1,2}, Matei Ripeanu\textsuperscript{2}, Nicolas Tripoul\textsuperscript{2}, Geoffrey Sanders\textsuperscript{1}, Roger Pearce\textsuperscript{1}
\textsuperscript{1}Center for Applied Scientific Computing, Lawrence Livermore National Laboratory
\textsuperscript{2}Department of Electrical and Computer Engineering, University of British Columbia
\{treza, matei, mtripoul\}@ece.ubc.ca, \{sanders29, rpearce\}@llnl.gov

Abstract—Pattern matching is a powerful graph analysis tool. Unfortunately, existing solutions have limited scalability, support only a limited set of search patterns, and/or focus on only a subset of the real-world problems associated with pattern matching. This paper presents a new algorithmic pipeline that: (i) enables highly scalable pattern matching on labeled graphs, (ii) supports arbitrary patterns, (iii) enables trade-offs between precision and time-to-solution (while always selecting all vertices and edges that participate in matches, thus offering 100\% recall), and (iv) supports a set of popular data analytics scenarios. We implement our approach on top of HavoqGT and demonstrate its advantages through strong and weak scaling experiments on massive-scale real-world (up to 257 billion edges) and synthetic (up to 4.4 trillion edges) graphs, respectively, and at scales (1,024 nodes / 36,864 cores) orders of magnitude larger than used in the past for similar problems.

I. INTRODUCTION

Pattern matching in graphs, that is, finding subgraphs that match a small template graph within a large background graph, is fundamental to graph analysis and has applications in multiple areas such as social network analysis \cite{1}, bioinformatics \cite{2}, and information mining \cite{3}. A match can be broadly categorized as either exact - i.e., there is a bijective map between the vertices/edges in the template and those in the matching subgraph, or approximate - the template and the match are just similar by some defined similarity metric \cite{4}. If the template size is not limited, exact matching is not known to have a polynomial time solution in the general case \cite{5}. Berry et al. \cite{6} introduced the problem of type-isomorphism: metadata graphs where vertices and edges are labeled and, in addition to topological constraints, a match identifies nodes and edges with the same labels in the template and the background graph. While the labeled version does not reduce the worst-case complexity of the original problem, past experience \cite{11,12,13} has demonstrated that label-based matching can be a powerful tool with potential for practical, real-world applications such as social network analysis.

The Challenge. Applications that mine graphs consist of tens of billions of edges are common \cite{9,10}. However, existing pattern matching solutions (we survey related work in \cite{C}) have limited capabilities: most importantly, they do not scale to massive graphs and/or support only a restricted set of search templates. Additionally, the algorithms at the core of the existing techniques are not suitable for today’s infrastructures relying on horizontal scalability and share-
two types of constraints: local and non-local, to possibly be part of a match. Local constraints involve only the vertex and its neighborhood: a vertex in an exact match needs to (i) match the label of a corresponding vertex in the template, and (ii) have edges to vertices labeled as prescribed in the adjacency structure of this corresponding vertex in the template. Non-local constraints are topological requirements beyond the immediate neighborhood of a vertex (e.g., that the vertex must be part of a cycle). We describe how these constraints are generated, and our algorithmic solution to verify them in III.

Third, we observe that, full match enumeration is not the most efficient avenue to support many high-level graph analysis scenarios. Depending on the final goal of the user, pattern matching problems fall into a number of categories which include: (a) determining if a match exists (or not) in the background graph (yes/no answer), (b) selecting all the vertices and edges that participate in matches, (c) ranking these vertices or edges based on their centrality with respect to the search template, i.e., the frequency of their participation matches, (d) counting/estimating the total number of matches (comparable to the well-known triangle counting problem), or (e) enumerating all distinct matches in the background graph. The traditional approach is to perform (e) and to use the result of the enumeration to answer (a) – (d). However, this approach is limited to small background graphs or is dependent on a low number of near and exact matches within the background graph (due to exponential growth of the search state-space). We argue that a pruning-based pipeline is not only a practical solution to (a) – (d) (and to other pattern-matching-related analytics), when full match enumeration is not the main interest, but also an efficient path towards full match enumeration. We demonstrate that a solution starting from the techniques we develop for pruning, efficiently supports match enumeration for two reasons: First, the pruned graph can be multiple orders of magnitude smaller than the background graph, and existing high-complexity enumeration routines (which otherwise would be intractable following the conventional approach) are now applicable. Second, our pruning techniques collect additional key information to accelerate match enumeration: for each vertex in the pruned graph, our algorithms build a list of its potential matches in the template [V-C].

Contributions. We capitalize on our preliminary study [15] that highlighted the effectiveness of pruning (yet in the context of a restricted set of templates) and design a pattern matching solution that is: generic - no restrictions on the set of patterns supported, precise - no false positives, offers 100% recall - retrieving all matches, efficient - low generated network traffic, and scalable - able to process graphs with up to trillions of edges on tens of thousands of cores (as demonstrated). In particular, we make the following contributions:

(i) Novel Asynchronous Algorithms. We have developed asynchronous vertex-centric algorithms able to prune the background graph to a precise, enumeration of all vertices and edges that participate in a match for arbitrary templates. The key gap we bridge is the ability support templates with repeated vertex labels, cycles, and arbitrary edge density. Importantly, the algorithms map well on a distributed asynchronous graph processing platform, thus enabling scalability and high-performance. For a given template, we propose heuristics to generate the constraints that are later used for pruning. We have also developed correctness proofs to show that these constraints eliminate all non-matching vertices and offer full recall (not included here due to space constraints).

(ii) Optimized Distributed Implementation. We offer an efficient implementation of these algorithms on the top of HavoqGT [21], an open-source asynchronous graph processing framework. The prototype includes two key optimizations that dramatically reduce the generated traffic: aggressive edge elimination, and what we call work aggregation, a technique that skips duplicate checks in non-local constraint checking, thus preventing possible combinatorial explosion. Additionally, our pruning implementation collects potential matching information: not only it prunes away all vertices and edges that do not participate in any match, but, for each of the vertices that remain, it collects their mappings to the search template. We use this information to accelerate match enumeration.

(iii) Proof of Feasibility. We demonstrate the applicability of this solution by experimenting on real-world and synthetic datasets orders of magnitude larger than prior work [V]. We evaluate scalability through two experiments: first, a strong scaling experiment using real-world datasets, including the largest openly available webgraph whose undirected version has over 257 billion edges; second, a weak scaling experiment using synthetic, R-MAT [22] graphs of up to 4.4 trillion edges, on up to 1,024 compute nodes (36,864 cores). We demonstrate support for search patterns representative of practical queries in both relatively low selectivity and needle in the haystack scenarios, and, to stress our system, consider patterns containing only the highest-frequency vertex labels (up to 14B instances). We show that our technique prunes the graph by orders of magnitude, which, combined with using the intermediary state generated by pruning, makes match enumeration feasible on graphs with trillions of edges.

(iv) Application Demonstration. We demonstrate the ability of our solution to support practical graph analytics queries. To this end, we use two real-world metadata graphs which we have curated from publicly available datasets, Reddit (3.9B vertices, 14B edges) and the smaller International Movie Database (IMDB), and demonstrate practical use cases of our technique to support rich pattern mining. (V-D).
(v) Exploring Trade-offs, and the Impact of Strategic Design Choices and Optimizations. Our approach has the added flexibility that search can be stopped early, leading to the ability to trade faster time to an approximate solution (or even precise solution, yet without 100% precision guarantees) for precision (rate of false positives in the pruned graph) and precision guarantees. We also explore the impact of each optimization used: the cumulative impact of these optimizations is a multiple orders of magnitude of runtime reduction, bringing pattern matching on massive metadata graphs in the realm of possible graph analytics.

II. PRELIMINARIES

We aim to identify all structures within a large background graph, \(G\), identical to a small connected template graph, \(G_0\). We describe general graph properties for \(G\), and use the same notation (summarized in Table I) for other graph objects.

A graph \(G(V, E)\) is a collection of \(n\) vertices \(V = \{0, 1, ..., n-1\}\) and \(m\) edges \(i, j \in E\), where \(i, j \in V\) (i is the edge’s source and \(j\) is the target). Here, we only discuss simple (i.e., no self-edges), undirected, vertex-labeled graphs, although the techniques are applicable to directed, non-simple graphs, with labels on both edges and vertices. An undirected \(G\) satisfies \((i, j) \in E\) if and only if \((j, i) \in E\). Vertex \(v\)’s adjacency list, \(adj(v)\), is the set of all \(j\) such that \((v, j) \in E\). A vertex-labeled graph also has a set of \(n_t\) labels \(L\) of which each vertex \(v \in V\) has an assignment \(\ell(i) \in L\).

A walk in \(G\) is an ordered subsequence of \(V\) where each consecutive pair is an edge in \(E\). A walk with no repeated vertices is a path. A path with equal first and last vertex is a cycle. An acyclic graph has no cycles.

We discuss several graph objects simultaneously: the template graph \(G_0(V_0, E_0)\) the background graph \(G(V, E)\), and the current solution subgraph \(G^\ast(V^\ast, E^\ast)\), with \(V^\ast \subset V\) and \(E^\ast \subset E\). Our techniques iteratively refine \(V^\ast\) and \(E^\ast\) until they converge to the union of all subgraphs of \(G\) that exactly match the template, \(G_0\).

For clarity, when referring to vertices and edges from the template graph, \(G_0\), we will use the notation \(q_i \in V_0\) and \((q_i, q_j) \in E_0\). Conversely, we will use \(v_i \in V\) and \((v_i, v_j) \in E\) for vertices and edges from the background graph \(G\) or the solution subgraph \(G^\ast\).

Definition 1. A subgraph \(H(V_H, E_H)\), where \(V_H \subset V\) and \(E_H \subset E\), is an exact match of template graph \(G_0(V_0, E_0)\) (in notation, \(H \sim G_0\)) if there exists a bijective function \(\phi: V_0 \leftrightarrow V_H\) with the properties (Note that \(\phi\) may not be unique for a given \(H\)):

(i) \(\ell(\phi(q)) = \ell(q)\), for all \(q \in V_0\) and

(ii) \(\forall (q_1, q_2) \in E_0\), we have \((\phi(q_1), \phi(q_2)) \in E_H\)

(iii) \(\forall (v_1, v_2) \in E_H\), we have \((\phi^{-1}(v_1), \phi^{-1}(v_2)) \in E_0\)

Intuition for our Solution. The algorithms we develop here iteratively refines vertex-match functions \(\omega(v) \subset V_0\) such that, for every \(v \in V\), \(\omega(v)\) stores a super-set of all template vertices \(v\) can possibly match. Set \(\omega(v)\) converges to contain all possible values of \(\phi^{-1}(v)\), were \(v\) involved in one or more matching subgraphs. When a single constraint involving

<table>
<thead>
<tr>
<th>Table I: Symbolic notation used.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Object(s)</strong></td>
</tr>
<tr>
<td>template graph, vertices, edges</td>
</tr>
<tr>
<td>template graph sizes</td>
</tr>
<tr>
<td>template vertices</td>
</tr>
<tr>
<td>template edges</td>
</tr>
<tr>
<td>set of vertices adjacent to (q_i) in (G_0)</td>
</tr>
<tr>
<td>background graph, vertices, edges</td>
</tr>
<tr>
<td>background graph sizes</td>
</tr>
<tr>
<td>background vertices</td>
</tr>
<tr>
<td>background edges</td>
</tr>
<tr>
<td>set of vertices adjacent to (v_i) in (G)</td>
</tr>
<tr>
<td>maximum vertex degree in (G)</td>
</tr>
<tr>
<td>label set</td>
</tr>
<tr>
<td>vertex label and label degree of (q_i)</td>
</tr>
<tr>
<td>matching subgraph, vertices, edges</td>
</tr>
<tr>
<td>solution subgraph, vertices, edges</td>
</tr>
<tr>
<td>vertex match function</td>
</tr>
<tr>
<td>set of non-local constraints for (G_0)</td>
</tr>
<tr>
<td>(q \in V_0) violated/unmet. (q) no longer a possibility for (v) in a match and (q) is removed: (\omega(v) \leftarrow \omega(v) \setminus {q})</td>
</tr>
</tbody>
</table>

Our previously developed pruning algorithms \([13]\) required that all vertex labels in the template \(G_0\) were unique, and that \(G_0\) be acyclic or edge-monocyclic to ensure 100% precision. Our new techniques achieve 100% precision for arbitrary \(G_0\).

Remark 1. Given an ordered sequence of all \(n_t\) vertices \((q_1, q_2, ..., q_{n_t}) \subset V_0\), a simple (although potentially expensive) search from \(v_1 \in V^\ast\) verifies if \(v_1\) is in a match, with \(\phi(q_1) = v_1\), or not. The search lists an ordered sequence \((v_1, v_2, ..., v_{n_0}) \subset V^\ast\), with \(\phi\) defined as \(\phi(q_k) = v_k\). Search step \(k\) proposes a new \(v_k\), checking Def\([13]\)(i) and (ii). If all checks are passed, the search accepts \(v_k\) and moves on to step \((k + 1)\), but terminates if no such \(v_k\) exists in \(V^\ast\). If the full list is generated with all label and edge checks passed then there exists a \(H \sim G\) with \(V_H = \{v_1, v_2, ..., v_{n_0}\}\).

We call this Template-Driven Search (TDS) and develop an efficient distributed version in \([15]\) to apply to the solution \(G^\ast(V^\ast, E^\ast)\). If TDS has been applied successfully then there are no false positives remaining.

III. GRAPH PRUNING FOR SCALABLE MATCHING

Our goal is to realize a technique which systematically eliminates all the vertices and edges that do not participate in any match \(H \sim G_0\). This approach is motivated by viewing the template \(G_0\) as specifying a set of constraints the vertices and edges that participate in a match must meet. As a trivial example, any vertex \(v\) whose label \(\ell(v)\) is not present in \(G_0\), cannot be present in an exact match. A vertex in an exact match also needs to have non-eliminated edges to non-eliminated vertices labeled as prescribed in the adjacency structure of the corresponding template vertex. Local constraints that involve a vertex and its neighborhood can be checked by having vertices communicate their (tentative) template match(s) with their one-hop neighbors in the solution subgraph \(G^\ast(V^\ast, E^\ast)\) (i.e., the currently pruned background graph). We call this process
**Local Constraint Checking (LCC).** Our experiments show that LCC is responsible for removing the bulk of non-matching vertices and edges.

Some classes of templates (with cycles or repeated vertex labels), require additional routines to check non-local properties and to guarantee that all non-matching vertices are eliminated. (Fig. [2] highlights the need for these additional checks). To support arbitrary templates, we have developed a process which we dub Non-local Constraint Checking (NLCC): first, based on the search template $G_0$, we generate the set of constraints $K_0$ that are to be verified, then prune the graph using each of them.

Alg. [1] presents an overview of our solution. This section provides high-level descriptions of the local and non-local constraint checking routines, while [14] provides the detailed distributed algorithms. For better understanding, Fig. [4] illustrates the complete workflow for the graph and pattern in Fig. [1] for which constraint generation is detailed in Fig. [3].

**Algorithm 1 Main Pruning Loop**

1: Input: background graph $G(V, E)$, template $G_0(V_0, E_0)$
2: Output: the solution subgraph $G^*(V^*, E^*)$
3: generate non-local constraint set $K_0$ from $G_0(V_0, E_0)$
4: $G^* ←$ LOCAL_CONSTRAINT_CHECKING ($G, G_0$)
5: while $K_0$ is not empty do
6: pick and remove next constraint $C_0$ from $K_0$
7: $G^* ←$ NON_LOCAL_CONSTRAINT_CHECKING ($G^*, G_0, C_0$)
8: if any vertex has been eliminated or
9: has one of its potential matches remove then
10: $G^* ←$ LOCAL_CONSTRAINT_CHECKING ($G^*, G_0$)

**Non-local Constraint Checking (NLCC)** aims to exclude vertices that fail to meet topological and label constraints beyond the one-hop neighborhood that is covered by LCC (Fig. [2]). We have identified three types of non-local constraints which can be verified independently: (i) Cycle Constraints (CC), (ii) Path Constraints (PC), and (iii) constraints that require Template-Driven Search (TDS) (see Remark [1]). For arbitrary templates, TDS constraints based on aggregating multiple paths/cycles enable further pruning and, when based on full template enumeration, ensure that pruning yields no false positives. Checking TDS constraints, however, can be expensive. To reduce the overall cost, we first generate single cycle- and path-based constraints which are usually less costly to verify and prune the graph before using TDS (the effectiveness of this ordering is evaluated in Fig. [14]c)).

**High-level Algorithmic Approach.** Regardless of constraint type, NLCC leverages a token passing approach: tokens are issued by background graph vertices whose corresponding template vertices are identified to have non-local constraints. After a fixed number of steps, we check if a token has arrived with multiple neighbors with the same label, the algorithm verifies if a matching vertex in the background graph has a minimum number of distinct neighbors with the same label as prescribed in the template. (ii) Edge elimination: this excludes edges to eliminated neighbors and edges to neighbors whose labels do not match the labels prescribed in the adjacency structure of its corresponding template vertex (e.g., Fig. [4] Iteration #1). Edge elimination is crucial for scalability since, in a distributed setting, no messages are sent over eliminated edges, significantly improving the overall efficiency of the system (evaluated in Fig. [5] Fig. [8]).

---

1 In a number of corner cases not all TDS constraints need to be aggregated for full precision, we skip this discussion for lack of space.
where expected (e.g., back to the originating vertex for checking the existence of a cycle). If not, then the issuing vertex does not satisfy the required constraint and is eliminated. Along the token path, the system verifies that all expected labels are encountered and, where necessary, uses the path information accumulated with the token to verify that different/repeated node identity constraint expectations are met. Next, we discuss how each of the constraints is verified.

Cycle Constraints (CC). Higher-order structures within \( G \) that survive LCC, but do not contain \( G_0 \), are possible if \( G \) contains a cycle (this happens if \( G \) contains one or more unrolled cycles as in Fig. 2 Template (a)). To address this, we directly check for cycles of the correct length.

Path Constraints (PC). If the template \( G_0 \) has two or more vertices with the same label three or more hops away from each other, then structures in \( G \) that survive LCC, yet contain no match, are possible (Fig. 2 Template (b)). Thus, for every vertex pair with the same label in \( G_0 \), we directly check the existence of a path of correct length and label sequence for prospective matching vertices in \( G^* \). Opposite to cycle checking, after a fixed number of steps, a token must be received by a vertex different from the initiating vertex but with an identical label.

TDS Constraints. These are partial or complete (i.e., including all edges of the template) walks on the template. The token walks the constraint in the background graph and verifies that each node visited meets its neighborhood constraints (Remark 1) - in our distributed memory setting, this is done by maintaining a history of the walk and checking that previously visited vertices are revisited as expected. TDS constraints are crucial to guarantee no-false positives for templates that are non-edge-monocyclic or have repeated labels (Fig. 2).

Token Generation. For CC and TDS constraints, a token must be initiated from each vertex that may participate in the substructure, whereas for PC, tokens are only initiated from terminal vertices.

Further Optimization - Work Aggregation. All NLCC constraints attempt to identify if a walk exists from a specific vertex and though vertices with specific labels. Since the goal is to identify the existence of any such path, and multiple intermediate paths in the background graph often exist, to prevent combinatorial explosion, our duplicate work detection mechanism prevents an intermediated vertex (in the token path) from forwarding a duplicate token (evaluated in [AA-A]).

Non-local Constraint Generation. For a relatively large class of templates (i.e., those with unique labels and mono-cyclic edges), LCC and CC are sufficient to generate a precise solution. For the general case, we generate non-local constraints using the following heuristic. (Fig. 3 shows, for a given template, the non-local constraints to be verified and Fig. 4 shows how pruning progresses.) First, all the leaf vertices with unique labels are identified and ignored from this process (as LCC guarantees pruning if there is no match). Next, if the template has cycles, then individual cycles are identified (e.g., Fig. 3 Step 3) and a cycle constraint is generated for each cycle. Next, vertices with identical label are identified and all path constraints are generated for all such pairs (e.g., Fig. 3 Step 4, pentagonal vertices).

Finally, we identify TDS constraints in three steps. First, for templates with multiple cycles sharing more than one edge (i.e., non-edge-monocyclic), a TDS cyclic constraint is generated through union of previously identified cycle constraints. This results in a higher-order cyclic structure with a maximal set of edges that cover all the non-edge-mono cycles (e.g., Fig. 3 Step 5(1)). Second, for templates with repeated labels, a new TDS constraint is generated through a union of all previously identified path constraints. This procedure generates higher order structure that covers all the template vertices with repeated labels (e.g., Fig. 3 Step 5(2)). The final step generates a TDS constraint as the union of the previously identified two constraints (e.g., Fig. 3 Step 5(3)). (Note that the above is a heuristic, more constraints could be generated by creating various possible combinations of cycles and paths. Only this third step is mandatory to eliminate all false positives.)

IV. Asynchronous Distributed Algorithms

This section presents the constraint checking algorithms on top of HavoqGT [23], a MPI-based framework that supports asynchronous graph algorithms in distributed environments. Our choice for HavoqGT’s is driven by multiple considerations: first, unlike most graph processing frameworks that only support the Bulk Synchronous Parallel (BSP) model, HavoqGT has been designed to support asynchronous algorithms, essential to achieve high-performance; second, the framework has excellent scaling properties [24] [21]; and, finally, it enables load balancing: HavoqGT’s delegate partitioned graph distributes the edges of each high-degree vertices across multiple compute nodes, which is crucial for achieving scalability for scale-free graphs with skewed degree distribution.

In HavoqGT, graph algorithms are implemented as vertex-callbacks: the user-defined visit() callback can only access and update the state of a vertex. The framework offers the ability to generate events (a.k.a. ‘visitors’ in HavoqGT lingo) that trigger this callback - either at the entire graph level using the do traversal() method, or for a neighboring vertex using the push(visitor) call (this enables asynchronous vertex-to-vertex communication). The asynchronous graph computation completes when all events have been processed, which is determined by a distributed quiescence detection algorithm [25].

Alg. 1 outlines the key steps of the graph pruning procedure. Below, we describe the distributed implementation of the local and non-local constraint checking routines. Alg. 2 lists the state maintained at each vertex and its initialization.

Local Constraint Checking is implemented as an iterative process (Alg. 3 and the corresponding callback, Alg. 4). Each iteration initiates an asynchronous traversal by invoking the do traversal() method and, as a result, each active vertex receives a visitor with msg type = init. In the triggered visit() callback, if the label of a vertex \( v \) in the graph is a match for the label of any vertex in the template and the vertex is still
Algorithm 2 Vertex State and Initialization

1: status of vertex $v_j$: $\alpha(v_j) \leftarrow \text{true}$ (active) if $\exists \ell \in \mathcal{V}_0$ s.t. $\ell(v_j) = \ell(q_j)$, otherwise $\text{false}$ (i.e., $v_j$ has been pruned)
2: set of possible matches in template for vertex $v_j$: $\omega(v_j) \leftarrow$ initially all $q_k \in \mathcal{V}_0$ s.t. $\ell(q_k) = \ell(v_j)$
3: map of active edges of vertex $v_j$: $\varepsilon(v_j)$ keys are initialized to $\text{adj}(v_j)$. The value field, which is initially $\emptyset$, is set to $\omega(v_j)$, for each $v_i \in \varepsilon(v_j)$ that has communicated its state to $v_j$
4: set of already forwarded tokens by vertex $v_j$: $\tau(v_j) \leftarrow$ initially empty, used for work aggregation in NLCC

Algorithm 3 Local Constraint Checking

1: $\eta(v_j, v_j)$ - tests if $v_j$ satisfies a local constraint of $v_j$; returns $\omega(s)$
2: if constraints are satisfied, $\emptyset$ otherwise
3: procedure LOCALCONSTRAINTCHECKING ($G, G_0$)
4: do
5:     doTraversal($\text{msgtype} \leftarrow \text{init}$)
6:   barrier
7: for all $v_j \in V$ do
8:     $\omega' \leftarrow \emptyset$ // set of matches in template for neighbors of $v_j$
9:   for all $v_i \in \varepsilon(v_j)$ do
10:      if $\eta(v_i, v_j) = \emptyset$ then
11:         $\varepsilon(v_j).remove(v_i)$ ▷ edge eliminated
12:     continue
13:   else
14:      $\omega' \leftarrow \omega' \cup \eta(v_j, v_j)$ ▷ accum. matches of the nbrs.
15:      reset the value field of $v_i \in \varepsilon(v_j)$ for the next iteration
16: for all $q_k \in \omega(v_j)$ do
17:   if $\text{adj}(q_k) \not\subseteq \omega'$ then
18:      $\omega(v_j).remove(q_k)$ ▷ remove from potential matches
19:      continue
20: if $\varepsilon(v_j) = \emptyset$ or $\omega(v_j) = \emptyset$ then
21:    $\alpha(v_j) \leftarrow \text{false}$ ▷ vertex eliminated
22: while vertices or edges are eliminated ▷ global detection

Algorithm 4 Local Constraint Checking Visitor

1: visitor state: $v_j$ - vertex that is visited
2: visitor state: $v_s$ - vertex that originated the visitor
3: visitor state: $\omega(v_s)$ - set of possible matches in template for vertex $v_s$
4: visitor state: $\text{msgtype} \leftarrow \text{init}$ or $\text{alive}$
5: procedure VISIT ($G, v_q$) // $v_q$ - visitor queue
6: if $\alpha(v_j) = \text{false}$ then return
7: if $\text{msgtype} = \text{init}$ then
8: for all $v_i \in \varepsilon(v_j)$ do
9:     $\text{LCC_VISITOR}(v_i, v_j, \omega(v_j), \text{alive})$
10: $v_q.push(vis)$
11: else if $\text{msgtype} = \text{alive}$ then
12: $\tau(v_j).get(v_q) \leftarrow \omega(v_j)$

Non-local Constraint Checking routine iterates over $\mathcal{C}_0$, the set of non-local constraints to be checked, and validates each $C_0 \in \mathcal{C}_0$ one at a time. Alg. describes the solution to verify a single constraint: tokens are initiated through an asynchronous traversal by invoking the doTraversal() method. Each active vertex $v_j \in \mathcal{G}$ that is a potential match for the vertex $q_0$ at the head of a path $C_0$, broadcasts a token to all its active neighbors in $\varepsilon(v_j)$. A map $\gamma$ is used to track these token issuers. A token is a tuple $(t, r)$ where $t$ is an ordered list of vertices that have forwarded the token and $r$ is the hop-counter; $t_0 \in t$ is the token-issuing vertex in $\mathcal{G}$. The ordered list $t$ is essential for TDS since it enables detection of distinct vertices with the same label in the token path.

Algorithm 5 Non-local Constraint Checking

1: procedure NONLOCALCONSTRAINTCHECKING ($G, G_0, C_0$)
2: $\gamma \leftarrow$ map of token source vertices (in $G$) for $C_0$; the value field
3: (initialized to false) is set to true if the token source vertex meets the requirements of $C_0$
4: doTraversal($\text{msgtype} \leftarrow \text{init}$)
5: barrier
6: for all $v_j \in \gamma$ do
7:   if $\gamma.get(v_j) \neq \text{true}$ then
8:      $\omega(v_j).remove(q_i)$ where $q_i$ is the first vertex in $C_0$ ▷ violates $C_0$, eliminate potential match
9:      $\omega(v_j).remove(q_j)$ where $q_j$ is the first vertex in $C_0$ ▷ no potential match left
10:     $\alpha(v_j) \leftarrow \text{false}$ ▷ vertex eliminated
11: $\forall v_j \in V$, reset $\tau(v_j)$

When an active vertex $v_j$ receives a token with $\text{msgtype} = \text{forward}$, it verifies that if $\omega(v_j)$ is a match for the next entry in $C_0$, if it has received the token from a valid neighbor (with respect to entries in $C_0$), and that the current hop count is $< \lvert C_0 \rvert$. If these requirements are satisfied (i.e., $\mu$ returns true), $v_j$ sets itself as the forwarding vertex (added to $t$), increments the hop count, and broadcasts the token to all its active neighbors in $\varepsilon(v_j)$. If any of the constraints are not met, $v_j$ drops the token. If the hop count $r$ is equal to $\lvert C_0 \rvert$ and $v_j$ is the same as the source vertex in the token, for a cyclic template, a path has been found and $v_j$ is marked true in $\gamma$. For path constraints, an acknowledgement is sent to the token issuer to update its status in $\gamma$ (Alg. [6] lines 28 – 31). Once verification of a constraint $C_0$ has been completed, the vertices that are not marked true in $\gamma$, are invalidated, i.e., $\alpha(v_j) \leftarrow \text{false}$ (Alg. [5] line #12). NLCC uses an unordered set $\tau(v_j)$ (Alg. [4] line #4) for work aggregation (see Alg. [6] line #14): at each vertex, this is used to detect if another copy of token has already visited the vertex $v_j$ taking a different path.
Algorithm 6 Non-local Constraint Checking Visitor

1: visitor state: \( v_j \) - vertex that is visited
2: visitor state: tokens - the token is a tuple \((t, r)\) where \(t\) is an ordered list of vertices that have forwarded the token and \(r\) is the hop-counter; \(t_0 \in t\) is the vertex that originated the token.
3: visitor state: msgtype - init, forward or ack
4: \( \mu(v_j, C_0, \text{token}) \) - tests if \( v_j \) satisfies requirements of \( C_0 \) for the current state of \text{token}; returns true if constraints are met, false otherwise
5: procedure VISIT(G, \text{token})

6: if \( \alpha(v_j) = \text{false} \) then return
7: if msgtype = init and \( \exists q_k \in \omega(v_j) \) where \( q_k = q_0 \in C_0 \) then
8: \( \gamma.\text{init}(v_j); r \leftarrow 1; \text{token} \leftarrow (t, r); \gamma.\text{insert}(v_j, \text{false}) \)
9: for all \( v_i \in e(v_j) \) do
10: \( \text{vis} \leftarrow \text{NLCC_VISITOR}(v_i, \text{token}, \text{forward}) \)
11: \( \text{msg}.\text{push}(\text{vis}) \)
12: else if msgtype = forward then \( \triangleright v_j \) received a token
13: if \( \mu(v_j, C_0, \text{token}) \) then \( \triangleright \) work aggregation optimization
14: \( \tau(v_j).\text{insert}(\text{token}) \)
15: else return \( \triangleright \) ignore token if it was previously forwarded by \( v_j \)
16: if \( \mu(v_j, C_0, \text{token}) \) then \( \triangleright \) walk can be extended with \( v_j \) and it has not yet reached \( |C_0| \) length
17: \( \text{token}.t.\text{add}(v_j); \text{token}.r \leftarrow \text{token}.r + 1; \)
18: for all \( v_i \in e(v_j) \) do \( \triangleright \) forward the token
19: \( \text{vis} \leftarrow \text{NLCC_VISITOR}(v_i, \text{token}, \text{forward}) \)
20: \( \text{msg}.\text{push}(\text{vis}) \)
21: else if \( \mu(v_j, C_0, \text{token}) \) then \( \triangleright \) walk can be extended with \( v_j \) and it has reached \( |C_0| \) length
22: \( \gamma.\text{get}(v_j) \leftarrow \text{true} \) return \( \triangleright v_j \) meets requirements of \( C_0 \)
23: if \( C_0 \) is cyclic and \( t_0 = v_j \) then
24: \( \gamma.\text{get}(v_j) \leftarrow \text{true} \) return \( \triangleright v_j \) meets requirements of \( C_0 \)
25: if \( C_0 \) is acyclic and \( t_0 \neq v_j \) then
26: \( \text{vis} \leftarrow \text{NLCC_VISITOR}(v_0, \text{token}, \text{ack}) \)
27: \( \text{msg}.\text{push}(\text{vis}) \) \( \triangleright \) send \text{ack} to the token originator \( t_0 \in t \)
28: else msgtype = ack then
29: \( \gamma.\text{get}(v_j) \leftarrow \text{true} \) return \( \triangleright v_j \) meets requirements of \( C_0 \)
30:

V. Evaluation

We present strong \((V-B)\) and weak \((V-A)\) scaling experiments of pruning on massive real-world and synthetic graphs; additionally we demonstrate full match enumeration starting from the pruned graph \((V-C)\); we evaluate the effectiveness of the optimizations our system incorporates \((A-A)\), we highlight the use of our system in the context of realistic data analytics scenarios \((V-D)\); we explore time-to-solution vs. precision/guarantees trade-offs \((V-E)\); and finally, we compare our solution with a recent work, QFrag \([28]\) \((A-C)\).

Testbed. The testbed is the 2.6Pflop Quartz cluster at the Lawrence Livermore National Lab., comprised of 2.634 nodes and the Intel Omni-Path interconnect. Each node has two 12-core Intel Xeon E5-2695v4 @ 2.10GHz processors and 128GB of memory [29]. We run one MPI process per core.

Datasets. We summarize the main characteristics of the datasets used for evaluation and explain how we have generated vertex labels where necessary. For all graphs, we created undirected versions of the graphs; two directed edges are used to represent each undirected edge.

The datasets used for evaluation.

| Dataset                  | Type    | \(|V|\)   | \(|E|\)   | \(d_{max}\) |
|--------------------------|---------|----------|----------|------------|
| Web Data Commons [27]    | Real    | 3.5B     | 257B     | 95M        |
| Reddit [28]              | Real    | 3.9B     | 14B      | 789K       |
| Internet Movie Database [31] | Real   | 5.5M     | 298M     | 552K       |
| Patent [28]              | Real    | 2.7M     | 288M     | 789        |
| Youtube [28]             | Real    | 4.6M     | 88M      | 2.5K       |
| R-MAT up to Scale 37 [24]| Synthetic | 137B    | 4.4T     | 612K       |

Web Data Commons (WDC) is a web-graph whose vertices are webpages and edges are hyperlinks. To create vertex labels, we extract the top-level domain names from the webpage URL, e.g., .org or .edu. If the URL contains a common second-level domain name, it is chosen over the top-level domain name. For example, for .ox.ac.uk, .ac is selected as the vertex label. A total of 2,903 unique labels are distributed among the 3.5B graph vertices. We curated the Reddit social-media graph from an open archive of billions of public posts and comments from Reddit.com. Reddit allows its users to rate (upvote or downvote) others’ posts and comments. The graph has four types of vertices: Author, Post, Comment and Subreddit (a category for posts). For the Post and Comment there are three possible labels: Positive, Negative, and Neutral (indicating the overall balance of positive and negative votes) or No rating. An edge is possible between an Author and a Post, an Author and a Comment, a Subreddit and a Post, a Post and a Comment (to that post), and between two Comments that have a parent-child relationship. The International Movie Database (IMDB) graph was curated from the publicly available data. The graph has five types of vertices: Movie, Genre, Actress, Actor and Director. An edge is only possible between a Movie type vertex and a non-Movie type vertex. We use the smaller Patent and YouTube graphs to compare published results by Serafini et al. [28]. The synthetic R-MAT graphs exhibit approximate power-law degree distribution. These graphs were created following the Graph 500 [32] standards; 2\(^{Scale}\) vertices and an
undirected edge factor of 16. For example, a Scale 30 graph has $|V| = 2^{30}$ and $|E| \approx 32 \times 2^{30}$ (as we create a directed version). We leverage degree information to create vertex labels, computed using the formula, $\ell(v_i) = \lceil \log_2(d(v_i)+1) \rceil$.

Search Templates. To stress our system, we use templates based on patterns naturally occurring, and relatively frequent, in the template graphs. The R-MAT (Fig. 5) and WDC (Fig. 7) patterns include vertex labels that are among the most frequent in the respective graphs. The Reddit and IMDB patterns (Fig. 11) include most of the vertex labels in these two graphs. We chose templates to exercise different constraint checking scenarios: the search templates have repeated vertex metadata and non-edge-monicypirical properties.

Experimental Methodology. All runtime numbers provided are averages over 10 runs. For weak scaling experiments, we do not present scaling numbers for a single node as this experiment does not involve network communication and benefits from data locality. For strong scaling experiments, the smallest experiment uses 64 nodes, as this is the lowest number of nodes that can load the graph in memory.

A. Weak Scaling Experiments

To evaluate the ability to process massive graphs, we use weak scaling experiments and the synthetic R-MAT graphs up to Scale 37 ($\sim 4.4T$ edges) and up to 1024 nodes (36,864 cores). Fig. 5 shows the two search pattern used and Fig. 6 presents the runtimes. Since there are multiple vertices in the pattern with identical labels (at more than two-hop distance), the patterns require NLCC - path checking - to ensure no false positives in the pruned graph. We see steady scaling all the way to the Scale 37 graph, which has $\sim 4.4$ trillion edges, on 1024 nodes (36,864 cores). Runtime is broken down to the individual iteration to evaluate scaling and the individual contribution of each intermediate step. As a graph gets pruned, the subsequent iterations require less time. Fig. 6 includes (at the top of each bar) the final number of vertices and edges in the pruned graph, respectively. For better visibility, for WDC-1 (left plots), runtime for different iterations are split into two scales on the Y-axis: LCC and NLCC-path constraints are at the bottom, and LCC and NLCC-TDS constraints are at the top. To highlight the impact of edge elimination, for WDC-3 (right plots), we present results with edge elimination disabled (top) and enabled (bottom). Speedup over the 64 node configuration is also shown on the top of each stacked bar plot.

B. Strong Scaling Experiments

Fig. 8 shows the runtimes for strong scaling experiments when using the real-world WDC graph on up to 1024 nodes (36,864 cores). Intuitively, pattern matching on the WDC graph is harder than on the R-MAT graphs, as the WDC graph is both denser and has a more skewed degree distribution. We use the patterns presented in Fig. 7. WDC-1 is acyclic, yet has multiple vertices with the same label and thus, requires non-local constraint checking (PC and TDS). For better visibility, the plot splits checking initial LCC and NLCC-path constraints (bottom left) from NLCC-TDS constraints (top left). We notice near perfect scaling for the LCC phases, however, some of
the NLCC phases do not show linear scaling (explained in {A-B}). WDC-2 is an example of a pattern with multiple cycles sharing edges, and relies on CC and TDS constraint checking to guarantee no false positive matches. WDC-2 shows near-linear scaling with ~1/3 of the total time spent in the first LCC phase and little time spent in the NLCC phases. WDC-3 is a monocyclic template and, when edge elimination is used (bottom right) shows near linear scaling for both LCC and NLCC phases. The top right plot highlights the key performance impact of edge elimination: without it, the NLCC phases take almost one order of magnitude longer and the entire pruning 2–9× longer.

Table II: Match enumeration statistics: Number of matches for the chain and tree patterns (Fig. 5 top table), and WDC (Fig. 7), Reddit and IMDB (Fig. 11) patterns (bottom table) and the enumeration times, starting from the pruned graphs. Note that for WDC-1 and WDC-3, due to the extremely large number of matches (over half billion in each case) we stop enumeration early.

<table>
<thead>
<tr>
<th>R-MAT Scale</th>
<th>#Compute Nodes</th>
<th>Chain Count</th>
<th>Chain Time (s)</th>
<th>Tree Count</th>
<th>Tree Time (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>28</td>
<td>2</td>
<td>2,716</td>
<td>10.36</td>
<td>1,186</td>
<td>10.38</td>
</tr>
<tr>
<td>31</td>
<td>16</td>
<td>3,747</td>
<td>10.54</td>
<td>1,488</td>
<td>10.40</td>
</tr>
<tr>
<td>34</td>
<td>128</td>
<td>7,529</td>
<td>11.28</td>
<td>3,766</td>
<td>11.28</td>
</tr>
<tr>
<td>37</td>
<td>1024</td>
<td>55,710</td>
<td>10.10</td>
<td>32,532</td>
<td>5.53</td>
</tr>
</tbody>
</table>

Table II: Match enumeration statistics: Number of matches for the chain and tree patterns (Fig. 5 top table), and WDC (Fig. 7), Reddit and IMDB (Fig. 11) patterns (bottom table) and the enumeration times, starting from the pruned graphs. Note that for WDC-1 and WDC-3, due to the extremely large number of matches (over half billion in each case) we stop enumeration early.

C. Match Enumeration

As our technique prunes the graph by orders of magnitude (see Fig. 9 and 12), match counting and full match enumeration are now feasible. Table II (top) lists the number of distinct matches and the time to enumerate the chain and tree patterns on some of the R-MAT graphs we used. While these results prove that our match enumeration routines scale well, the match counts and enumeration time for WDC, Reddit and IMDB patterns listed on Table II (bottom) are more revealing.

There are three important takeaways: First, while our match enumeration technique is able to enumerate an immense number of matches (see, for example results for WDC-1 and WDC-3 with 500+ million matches, or even IMDB-1 with 1.5+ million matches), presenting results as pruned vertex/edge sets (with less than 100,000 vertices) avoids the combinatorial explosion and makes is feasible to imagine further analytics. Second, as Fig. 10 clearly shows, presenting the results as the union of all matches (rather than explicit match enumeration) is not only more space efficient, but also, in some cases even easy to directly understand by a human analyst. Finally, while we omit details due to lack of space, we note that key to supporting match enumeration, is edge pruning: this reduces the edge density in the pruned WDC graph by a factor of 10–15× (compared to using vertex pruning alone).

D. Example Use Case: Social Network Analysis

We demonstrate the ability of our scalable pattern matching technique to support complex data analytics scenarios in the context of social networks. Today’s user experience on social media platforms is tainted by the existence of malicious actors such as bots, trolls, and spammers. This highlights the importance of detecting unusual activity patterns that may indicate potential malicious attacks. We present three use cases: one for the IMDB graph and two queries that attempt to uncover suspicious activity in the Reddit dataset.

Fig. 11 summarizes the scenarios we target and presents the corresponding search patterns. Fig. 12 shows runtime for these scenarios, broken down to individual LCC and NLCC iteration levels. Although RDT-1 is much less frequent than RDT-2, on the same 64 nodes, pruning for RDT-1 takes more than 3× longer to complete as it spends more time verifying the non-local constraints. Although both patterns have a 6-cycle, RDT-2 allows verification of the two smaller cycles in isolation. (For NLCC, a longer path typically results in larger generated message traffic.) IMDB-1, on 8 nodes, spends the majority of the time verifying non-local, specifically TDS, constraints.

E. Pruning Precision/Guarantees vs. Time-to-Solution

Our approach gradually refines \( G^*(V^*, E^*) \) down to the complete set of vertices and edges that participate in at least one match and guarantees no false positives. Given that this is an iterative process, it is natural to investigate at what rate \( G^* \) is refined, and whether there are opportunities to trade between the precision (or existence of precision guarantees) of an intermediary solution, and compute time.

Fig. 13 shows the evolution of the precision of the intermediate solution over time for the various patterns. (We define precision as the ratio between the number of distinct vertices that participate in at least one match, and the size of refined.
We believe that the rate at which precision is achieved, is partly influenced by the order in which constraints are verified and our heuristics for constraint verification ordering leave room for improvement.

Please continue to Appendix for the remaining of the paper: evaluation (§A), a brief discussion on the potential limitations of this work (§B), related work (§C), and conclusion (§D).

ACKNOWLEDGEMENT

This work was performed in part under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under contract DEAC52-07NA27344 (LLNL-CONF-735657). Experiments were performed at the Livermore Computing facility.

REFERENCES


APPENDICES

A Evaluation (Continued)

A. Impact of Design Decisions and Strategic Optimizations

Edge Elimination. Fig. 8 (right plots) show scalability and performance gains as a result of enabling edge elimination. Without edge elimination, the WDC-3 pattern results in 3,180,678 edges selected (some are false positives). Edge elimination identifies the true positive matches and reduces the number of active edges to 255,022. In other words, the graph is $12.5\times$ sparser which in turn improves overall message efficiency of the system.

Asynchronous Communication. Our system is designed to harness the advantages of an asynchronous graph processing framework, yet a synchronous one could easily support the same algorithms. Fig. 14(a) shows runtime for two patterns that benefit the most from asynchronicity: for WDC-1 and RDT-1 ($2.65\times$ and $3.5\times$ gains, respectively) compared with a synchronous version that adds a barrier after each NLCC token propagation step. Asynchronous NLCC makes it possible for all walks to progress independently without synchronization overheads. Synchronous NLCC is implemented within HavoqGT as well.

Work Aggregation. Fig. 14(b) shows the performance gains enabled by the work aggregation strategy employed by NLCC (presented in [H] and Alg. 6). The magnitude of the gain: 10-50%, is data dependent and more pronounced when the pattern is abundant, e.g., WDC-1 has 600M+ instances (Table II).

Constraint Selection. For patterns for which full TDS is required for precision guarantees, the path and cycle constraints are there for performance optimizations (their goal is to prune away the non-matching part of the graph early) and we are interested to evaluate the impact of these optimizations. To this end, we compare time-to-solution for a configuration that generates and uses all NLCC constraints, and one that uses only the TDS constraints required to guarantee 100% precision. Our experiments show that, although it increases the number of iterations, verifying simpler and smaller substructures first is extremely effective: for some patterns (e.g., RDT-1) the system is not able to complete in reasonable time without these constraints, for others (e.g., WDC-2) these constraints enable a $2.39\times$ speedup (Fig. 14(c)).

B. Load Balancing

Load imbalance issues are inherent to problems involving irregular data structures, such as graphs, when these need to be partitioned for processing over multiple nodes. For our pattern matching solution, load imbalance can be further caused by two artifacts: first, over the course of execution, our solution causes the workload to mutate (i.e., we prune away vertices and edges), and, second, by nonuniform distribution of matches in the background graph: the vertices and edges that participate in the matches may reside on a small, potentially concentrated, portion of the graph. Load imbalance can indeed occur: for example, for the relatively rare WDC-2 pattern when using 128 nodes, for example, the vertices that participate in the final selection are distributed over as few as 114 partitions out of 4,608. The distribution is concentrated: 90% of the matching edges are on 85 partitions while more than half of the matching edges are located on only 15 partitions. For the more frequent WDC-1 pattern, 99% of the matching vertices are part of a single connected component. 50% of the matching edges are on less than 5% of the total partitions on a 64 node deployment, which becomes less than 3% on a 128 node deployment.

We employ a pseudo-dynamic, load balancing strategy. First, we checkpoint the current state of execution: the pruned graph, i.e., the set of active vertices and edges and the per-vertex state indicating template matches, $\omega(v_j)$ in Alg. 2. Next, using HavoqGT’s graph partitioning module, we reschedule vertex-to-processor assignment to evenly distribute vertices and edges across processing cores. Processing is then resumed on the rebalanced workload. (Note that, depending on the size of the pruned graph, it is possible to resumed processing on a smaller deployment.) Over the course of the execution, checkpointing and rebalancing can be repeated as needed.

As a proof of feasibility, to examine the impact of this technique, we analyze the runs for WDC-1 and RDT-1 patterns. (We chose some of the real-world workloads as they are more likely to lead to imbalance than synthetically generated load.) Fig. 14(d) compares performance of the pruning algorithms with and without load balancing. For these examples, we perform workload rebalancing only once: for WDC-1, before verifying the TDS constraints, and for RDT-1, when the pruned graph is four orders of magnitude smaller. The extent of load imbalance is more severe for WDC-1 on the smaller 64 node deployment compared to using 128 nodes; workload
rebalncing improves runtime by 3.1× and 1.3× on 64 and 128 nodes, respectively. In the case of RDT-1, the gain in runtime as a result of load balancing is 1.7×. Given the pruned graphs are orders magnitude smaller then the original graph, check pointing, rebalncing, and relaunching the computation takes less than two minutes, which is negligible compared to the gain achieved in time-to-solution. In Table I we run enumeration for WDC-1, WDC-2 and RDT-1 on the rebalnced graphs.

TABLE III: Performance comparison between QFrag and our pattern matching system. The table shows runtime in seconds for full enumeration for QFrag; pruning and full enumeration for our distributed system (labeled PruneJuice-distributed), and for a single node implementation of our graph pruning-based approach tailored for a shared memory system (labeled PruneJuice-shared). For PruneJuice, we split time-to-solution into pruning (top row) and enumeration (bottom row) times. We use the same graphs (Patent and YoutTube) and some of the query templates (Q4 – Q7) used for evaluation of QFrag in [28].

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Q4</td>
<td>4.19</td>
<td>8.08</td>
<td>0.238</td>
<td>0.704</td>
<td>0.100</td>
<td>0.400</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.223</td>
<td>1.143</td>
<td>0.010</td>
<td>0.010</td>
</tr>
<tr>
<td>Q6</td>
<td>5.99</td>
<td>10.26</td>
<td>0.065</td>
<td>0.301</td>
<td>0.005</td>
<td>0.010</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.874</td>
<td>2.340</td>
<td>0.070</td>
<td>1.730</td>
</tr>
<tr>
<td>Q7</td>
<td>6.36</td>
<td>11.89</td>
<td>0.596</td>
<td>1.613</td>
<td>0.130</td>
<td>0.820</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.039</td>
<td>0.180</td>
<td>0.005</td>
<td>0.010</td>
</tr>
<tr>
<td>Q8</td>
<td>10.05</td>
<td>14.48</td>
<td>0.959</td>
<td>2.633</td>
<td>0.100</td>
<td>1.370</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.049</td>
<td>0.738</td>
<td>0.001</td>
<td>0.010</td>
</tr>
</tbody>
</table>

C. Comparison with Existing Work: QFrag

We empirically compare our work with a recent work (2017) on pattern matching from the database community: QFrag [28].

Similar to our solution, QFrag targets exact pattern matching, yet there are two main differences: QFrag assumes that the entire graph fits in the memory of each compute node and uses data replication to enable search parallelism. More importantly, QFrag employs a sophisticated load balancing strategy between parallel instances of a search to achieve good scalability. QFrag is implemented on top of Apache Spark and Giraph [19]. In QFrag, each replica runs an instance of a pattern enumeration algorithms called TurboISO [41]. Through evaluation, the authors demonstrated QFrag’s performance advantages over two other distributed pattern matching systems: (i) TriAD [42], an MPI-based distributed RDF [43] engine based on an asynchronous distributed join algorithm, and (ii) GraphFrames [44], [45], a graph processing library for Apache Spark, also based on distributed join operations.

Given that we have demonstrated the good scalability of our solution (Serafini et al. [28] demonstrate equally good scalability properties for QFrag yet on much smaller graphs), we are interested to establish a comparison baseline at single node scale. To this end, we run experiments on a modern shared memory machine with 60 CPU-cores, and use four template queries and two real-world graphs (Patent and Youtube) that were used for evaluation of QFrag [28]. We run QFrag with 60 threads and HavoqGT with 60 MPI processes. The results are summarized in Table I. QFrag runtimes for match enumeration (first pair of columns) are comparable with the results presented in [28], so we have reasonable confidence that we replicate their experiments well. With respect to combined pruning and enumeration time, our system (second pair of columns, presenting pruning and enumeration time separately) is consistently faster than QFrag on all the graphs, for all the queries. We note that our solution does not take advantage of shared memory of the machine at the algorithmic or implementation level (we use different processes, one MPI process per core), and has the system overhead of MPI-communication between processes. Additionally, unlike QFrag, our system is not handicapped by the memory limit of a single machine as it supports graph partitioning. To highlight the effectiveness of our technique and get some intuition on the magnitude of the MPI overheads in this context, we implemented our technique for shared memory and present runtimes (when using 60 threads) for the same set of experiments in Table III (right two columns). We notice up to an order of magnitude better performance, the main cost for our technique, compared to the distributed implementation running on a single node.

In summary, our system works about 4–10× faster than QFrag, and, if excluding distributed system memory overheads and considering the pruning time for the shared memory solution and conservatively reusing enumeration runtime for the distributed solution, it is about 6–100× faster than QFrag.

B LIMITATIONS / DISCUSSION

We categorize the limitations of our proposed system based on their respective sources.

Limitations stemming from major design decisions. Our system inherits limitations of a system that performs exact matching (compared to a system that focuses on approximate matching, e.g., based on graph simulation [7]). Similarly, our system inherits all limitations of its communication and middleware infrastructure, MPI and HavoqGT. One example is the lack of flow control in these infrastructures which sometimes leads to message buildup and system collapse.

Limitations stemming from the targeted uses cases. In the same vein, we note that our system targets a graph analytics scenario (queries that need to cover the entire graph), rather than the traditional graph database queries that attempt to find a specific pattern around a vertex indicated by the user (where other systems may perform better).

Limitations stemming from attempting to design a generic system. Systems optimized for specific patterns may perform better (e.g., systems optimized to enumerate triangles or treetlets [44] or systems relying on multi-join indices to support patterns with limited diameter).

Limitations stemming from incomplete understanding and work in progress. While we propose heuristics that appear to work well for our experiments, one of the key challenges is identifying an optimal set of constraints and their execution order. We believe graph statistics at different stages in execution can be used to dynamically select effective constraints.

C RELATED WORK

The volume of related work on graph processing in general [46], [19], [20], [47], [48], [49] and on pattern match-
TABLE IV: Comparison of past work on distributed pattern matching. The table highlights the characteristics of the solution presented (exact vs. approximate matching), its implementation infrastructure, and summarizes the details of the largest-scale experiment performed. We highlight the fact that our solution is unique in terms of demonstrated scale, ability to perform exact matching, and ability to retrieve all matches.

<table>
<thead>
<tr>
<th>Contribution</th>
<th>Model</th>
<th>Framework/ Language</th>
<th>Match Type</th>
<th>Max-Query Size</th>
<th>Metadata</th>
<th>#Compute Nodes</th>
<th>Max-Real Graph</th>
<th>Max-Synthetic Graph</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plantenga [33]</td>
<td>Graph Walk</td>
<td>Hadoop</td>
<td>Approx.</td>
<td>4-cliques Real</td>
<td>64</td>
<td>107B edges</td>
<td>R-MAT Scale 20</td>
<td>N/A</td>
</tr>
<tr>
<td>QFrag [28]</td>
<td>Tree-based</td>
<td>Spark</td>
<td>Exact</td>
<td>7 edges Real</td>
<td>10</td>
<td>117M edges</td>
<td>N/A</td>
<td>269M edges</td>
</tr>
<tr>
<td>SAHAD [34]</td>
<td>Explore-Join</td>
<td>Hadoop</td>
<td>Approx.</td>
<td>12 vertices Synthetic</td>
<td>40</td>
<td>N/A</td>
<td>15</td>
<td>117M edges Erdős-Renyi 1M edges</td>
</tr>
<tr>
<td>FASICA [35]</td>
<td>Explore-Join</td>
<td>MPI</td>
<td>Approx.</td>
<td>12 vertices N/A</td>
<td>15</td>
<td>117M edges</td>
<td>N/A</td>
<td>2B edges (Unif. rand.)</td>
</tr>
<tr>
<td>Chakaravarthy et al. [36]</td>
<td>Explore-Join</td>
<td>MPI</td>
<td>Approx.</td>
<td>10 vertices N/A</td>
<td>15</td>
<td>117M edges</td>
<td>N/A</td>
<td>2B edges (Unif. rand.)</td>
</tr>
<tr>
<td>PGX.D/Async [14]</td>
<td>Async. DFS</td>
<td>Java/C++</td>
<td>Exact</td>
<td>4 edges Synthetic</td>
<td>32</td>
<td>N/A</td>
<td>15</td>
<td>16.5M edges 4B vertices</td>
</tr>
<tr>
<td>Gao et al. [37]</td>
<td>Explore-Join</td>
<td>Giraph</td>
<td>Approx.</td>
<td>50 vertices Synthetic</td>
<td>28</td>
<td>3.7B edges</td>
<td>N/A</td>
<td>100M vertices</td>
</tr>
<tr>
<td>Sun et al. [38]</td>
<td>Explore-Join</td>
<td>C#.Net4</td>
<td>Exact</td>
<td>15 vertices Type only</td>
<td>16</td>
<td>5.1M edges</td>
<td>8</td>
<td>300M edges N/A</td>
</tr>
<tr>
<td>Ma et al. [39]</td>
<td>Graph Simulation</td>
<td>Python</td>
<td>Approx.</td>
<td>15 vertices Type only</td>
<td>8</td>
<td>300M edges</td>
<td>N/A</td>
<td>100M vertices</td>
</tr>
<tr>
<td>Fard et al. [40]</td>
<td>Graph Simulation</td>
<td>GPS</td>
<td>Approx.</td>
<td>N/A</td>
<td>8</td>
<td>300M edges</td>
<td>N/A</td>
<td>100M vertices</td>
</tr>
</tbody>
</table>

**Approximate Matching.** Recently, a new family of approximate matching algorithms based on graph simulation [53] has been proposed [11]. As opposed to exact matching, graph simulation algorithms relax matching constraints, e.g., matching based on vertex attributes and their connectivity constraints in the query. Simulation based algorithms have quadratic/cubic time-complexity and have been proposed as a possible solution for emerging matching problems when large-scale graphs are involved [53]. Two inexact matching algorithms based on graph simulation are introduced in [40]. Although results from both are only presented on relatively small real-world graphs. Gao et al. introduce another approximate matching algorithm based on explore-join and evaluate it on even larger query patterns than in [53]. Here, a query pattern is converted into a single-sink Directed Acyclic Graph (DAG) and message transition follows its topology.

**D Conclusion**

This paper presents a new algorithmic pipeline to support pattern matching on large-scale metadata graphs on large...
distributed-memory machines. We capitalize on the idea of graph pruning and develop asynchronous algorithms that use both vertex and edge elimination to iteratively prune the original graph and reduce it to a subgraph which represents the union of all matches. We have developed pruning techniques that guarantee a solution with 100% precision (i.e., no false positives in the final pruned graph) and 100% recall (i.e., all vertices and edges participating in matches are included) for arbitrary search patterns, including patterns with repeated vertex metadata, and patterns that have arbitrary cycles. Our algorithms are vertex-centric and asynchronous, thus, they map well onto existing high-performance graph frameworks. Our evaluation using up to 257 billion edge real-world web-graphs and up to 4.4 trillion edge synthetic R-MAT graphs, on up to 1,024 nodes (36,864 cores), confirms the scalability of our solution. We demonstrate that, depending on the search template, our approach prunes the graph by orders of magnitude which enables full pattern enumeration and counting on graphs with trillions of edges. Our success stems from a number of key design ingredients: asynchronicity, aggressive vertex and edge elimination while harnessing massive parallelism, intelligent work aggregation to ensure low message overhead, effective pruning constraints, and lightweight per-vertex state.
Here, we present an example of searching a pattern in a \textsc{R-MAT} generated graph using our program. The code is developed on top of \textsc{HavoqGT}. Detailed instructions on how to use our tool are available at \url{https://github.com/LLNL/havoqgt}.

Clone the code from \url{https://github.com/LLNL/havoqgt.git}.

You will require the latest releases of OpenMPI or MAVPICH2 and the Boost library to run \textsc{HavoqGT}. The code has only been tested on latest generation of Linux distributions. Once you have checked out the code, make sure you are on the master branch.

Go to the directory, build/quartz/ and build the \textsc{R-MAT} generator:

```
make generate
```

Create a directory, e.g., /usr/graph/, to store the generated graph.

Assuming you are in a Slurm environment, run the following command to generate a \textsc{R-MAT} graph:

```
srun -N1 –ntasks-per-node=4 –distribution=block ./src/generate_rmat -s 18 -p 1 -f 1 -o /dev/shm/rmat -b /usr/graph/rmat
```

This will create a graph with four partitions, to be run using four MPI processes. Note that this is a Scale 18 graph. (Notice the parameter for the -s flag.) The mmap/binary graph file will be store in /usr/graph/

Next, we build the pattern matching executable:

```
make run_pattern_matching_beta
```

We will search the following Tree pattern on the graph we just created. The numeric values on each vertex is the label of the respective vertex.

```
2
 / \
4 5
 / \
3 7 3 7
```

Tree

We use degree information to create numeric vertex labels, computed using the formula. \( \lceil \log_2(d(v_i) + 1) \rceil \). Here, \( d(v_i) \) is the degree of a vertex \( v_i \).

The input pattern is available at \url{https://github.com/LLNL/havoqgt} (See the instructions on the readme page regarding how to download the sample input pattern.)