G-thinker: Big Graph Mining Made Easier and Faster

Da Yan§1, Hongzhi Chen§2, James Cheng§3, M. Tamer Özsu‡4, Qizhen Zhang§5, John C. S. Lui§6

§Department of Computer Science and Engineering, The Chinese University of Hong Kong

{yanda, hzchen, jcheng, qzzhang, cslui}@cse.cuhk.edu.hk

†Department of Computer and Information Sciences, The University of Alabama at Birmingham

1yanda@uab.edu

‡David R. Cheriton School of Computer Science, University of Waterloo

tozsu@uwaterloo.ca

Abstract

This paper proposes a general system for computation-intensive graph mining tasks that find from a big graph all subgraphs that satisfy certain requirements (e.g., graph matching and community detection). Due to the broad range of applications of such tasks, many single-threaded algorithms have been proposed. However, graphs such as online social networks and knowledge graphs often have billions of vertices and edges, which requires distributed processing in order to scale. Unfortunately, existing distributed graph processing systems such as Pregel and GraphLab are designed for data-intensive analytics, and are inefficient for computation-intensive graph mining tasks since computation over any data is coupled with the data’s access that involves network transmission. We propose a distributed graph mining framework, called G-thinker, which is designed for computation-intensive graph mining workloads. G-thinker provides an intuitive graph-exploration API for the convenient implementation of various graph mining algorithms, and the runtime engine provides efficient execution with bounded memory consumption, light network communication, and parallelism between computation and communication. Extensive experiments were conducted, which demonstrate that G-thinker is orders of magnitude faster than existing solution, and can scale to graphs that are two orders of magnitude larger given the same hardware resources.

1 Introduction

We focus on a class of graph mining problems, namely subgraph finding problems, which aim to find all subgraphs in a graph that satisfy certain requirements. It may enumerate (or count) all of the subgraphs, or find only those subgraphs with top-k highest scores, or simply output the largest subgraph. Examples of subgraph finding problems include graph matching [12], maximum clique finding [26], maximal clique enumeration [1], quasi-clique enumeration [13], triangle listing and counting [9], and densest subgraph finding [11]. These problems have a wide range of applications including social network analysis [17, 20], searching knowledge bases [10, 27] and biological network investigation [8, 33]. Although many serial algorithms have been proposed to solve these problems, they cannot scale to big graphs such as social networks and knowledge graphs. Moreover, it is non-trivial to extend these algorithms for parallel processing, because (1) the input graph itself may be too big to fit in the memory of one machine, (2) a serial algorithm checks subgraphs using backtracking, where only one candidate subgraph is constructed (incrementally from the previous candidate subgraph) and examined at a time; a parallel algorithm that checks many subgraphs simultaneously in memory may cause memory overflow. The issue is further complicated in the distributed setting due to the high cost of accessing large amounts of remote data.

A plethora of distributed systems have been developed recently for processing big graphs [29], but most of them adopt a think-like-a-vertex style (or, vertex-centric) computation model [3, 7, 22, 31] which is pioneered by Google’s Pregel [15]. These systems only require a programmer to specify the behavior of one generic vertex (e.g., sending messages to other vertices) when developing distributed graph algorithms, but the resulting programs are usually data-intensive. Specifically, the processing of each vertex is triggered by incoming messages sent (mostly) from other machines, and the CPU cost of vertex processing is negligible compared with the communication cost of message transmission. Moreover, subgraph finding problems operate on subgraphs rather than individual vertices, and it is unnatural to translate a subgraph finding problem into a vertex-centric program.

As a result, big graph analytics research mostly focuses on problems that naturally have a vertex-program implementation, such as PageRank and shortest path, and few work has been devoted to large-scale subgraph
finding. To our knowledge, only two existing systems, NScale [19] and Arabesque [24], attempted to attack large-scale subgraph finding with a subgraph-centric programming model. Unfortunately, their execution engines still mine subgraphs in a data-intensive manner. Specifically, they construct all candidate subgraphs in a synchronous manner (i.e., large subgraphs are constructed from small ones), before the actual computation that examines these subgraphs. Materializing candidate subgraphs incurs large network and storage overhead (while CPU is under-utilized), and the actual computation-intensive mining process is delayed until the costly subgraph materialization is completed.

Another critical drawback of existing distributed frameworks is that they process each individual subgraph as an independent task, which loses many optimization opportunities. For example, if multiple subgraphs on a machine contain a common vertex \( v \), their tasks may share \( v \)'s information (e.g., adjacent edges). But in existing systems, each subgraph will maintain its own copy of \( v \)'s information (likely received from another machine), leading to redundant communication and storage. The synchronous execution model of existing systems is also prone to the straggler problem, due to imbalanced workload distribution among different machines.

In this paper, we identify the following five requirements that a distributed system for subgraph finding should satisfy in order to be efficient and user-friendly:

- The programming interface should be subgraph-centric.
- Computation-intensive processing should be native to the programming model. For example, a programmer should be able to backtrack a portion of graph to examine candidate subgraphs like in a serial algorithm, without materializing and transmitting any subgraph.
- There should exist no global synchronization among the machines, i.e., the processing of different portions of a graph should not block each other.
- Since a subgraph finding algorithm checks many (possibly overlapping) subgraphs whose cumulative volume can be much larger than the input graph itself, it is important to schedule the subgraph-tasks properly to keep the memory usage bounded at any point of time.
- Obviously, each machine should stream and process its subgraphs on its local disk (if memory is not sufficient), to minimize network and disk IO overhead.
- Subgraphs on a machine that contain a common vertex \( v \) should be able to share \( v \)'s information (e.g., adjacent edges), to avoid redundant data transmission and storage.

Based on these criteria, we designed a novel subgraph-centric system, called G-thinker, as a unified framework for developing scalable algorithms for various subgraph finding problems. To write a G-thinker program, a user only needs to specify how to grow a portion of the input graph \( g \) by pulling \( g \)'s surrounding vertices, and how to process \( g \) (e.g., by backtracking). Communication and execution details in G-thinker are transparent to end users. In G-thinker, each machine only keeps and processes a small batch of tasks in memory at any time (to achieve high throughput through task batching while keeping memory usage bounded). Subgraphs that are waiting to be processed (e.g., to grow its frontier by pulling remote vertices) are buffered in a disk-based priority queue. The priority queue is organized by a min-hashing based task scheduling strategy, in order to maximize the opportunity that the processing of different subgraphs share common vertices (including their adjacent edges) that are cached in the local machine.

We have used G-thinker to develop significantly more efficient and scalable solutions for a number of subgraph finding problems, including triangle counting, maximum clique finding, and graph matching. Compared with existing systems, G-thinker is up to hundreds of times faster and scales to graphs that are two orders of magnitude larger given the same hardware resources.

The rest of this paper is organized as follows. We motivate the need of a computation-intensive subgraph finding framework with the problem of maximal clique enumeration in Section 2, and then explain why existing systems are inefficient in Section 3. Section 4 provides an overview of the design of G-thinker, Section 5 introduces the programming interface of G-thinker, and Section 6 illustrates how to write application programs in G-thinker. We present the implementation of G-thinker in Section 7 and report experimental results in Section 8. Finally, we conclude the paper in Section 9.

2 A Motivating Example

A common feature of subgraph finding problems is that, the computation over a graph \( G \) can be decomposed into that over subgraphs of \( G \) that are often much smaller (called decomposed subgraphs), such that each result subgraph is found in exactly one decomposed subgraph. In other words, the decomposed subgraphs partition the search space and there is no redundant computation. We illustrate by considering maximal clique enumeration, which serves as our running example. Table 1 summarizes the notations used throughout this paper.
Example: Maximal Clique Enumeration. We decompose a graph \( G = (V, E) \) into a set of \( G_i \)’s subgraphs \( \{G_1, G_2, \ldots, G_n\} \), where \( G_i \) is constructed by expanding from a vertex \( v_i \in V \). Let us denote the neighbors of a vertex \( v \) by \( \Gamma(v) \). If we construct \( G_i \) as the subgraph induced by \( \{v_i\} \cup \Gamma(v_i) \) (\( G_i \) is called \( v_i \)’s 1-ego network), then we can find all cliques from these 1-ego networks since any two vertices in a clique must be neighbors of each other. However, a clique could be double-counted.

Let us define \( \Gamma_{g1}(v) = \{u \in \Gamma(v) | u \succ v\} \) where vertices are compared according to their IDs. To avoid redundant computation, we redefine \( G_i \) as induced by \( \{v_i\} \cup \Gamma_{g1}(v_i) \), i.e., \( G_i \) does not contain any neighbor \( v_j < v_i \). This is because any clique containing both \( v_i \) and \( v_j \) has already been computed when processing \( G_j \). Obviously, any clique \( C \) (let the smallest vertex in \( C \) be \( v_i \)) is only computed once, i.e., when \( G_i \) is processed.

We can distribute these decomposed subgraphs to different machines, so that each decomposed subgraph is processed using a serial backtracking algorithm to find cliques without network communication. Since the computation complexity of maximal clique enumeration is exponential to graph size, the computation cost of processing \( G_i \) is super-linear (to \( G_i \’s \) size) with a small constant (i.e., computation-intensive), while the transmission cost of creating \( G_i \) is linear with a large constant (due to limited network transmission rate). Thus, the computation cost and communication cost strike a balance when \( G_i \) is sufficiently large, and overlapping computation and communication over decomposed subgraphs significantly improves the overall performance.

However, since real graphs often follow power-law degree distribution, there may exist some vertex \( v_i \) with a very high degree, thus generating a large \( G_i \). Due to high computational complexity, the machine processing \( G_i \) may become the straggler that keeps processing \( G_i \) while other machines finish their tasks and become idle. To tackle this problem, a system should allow \( G_i \) to be further decomposed, so that the resulting decomposed subgraphs can be distributed to different machines for processing. In maximal clique enumeration, we can decompose \( G_i \) exactly as how we decompose \( G \), conditioned on that \( v_i \) is already in any clique found therein. The decomposition may recurse by looking at more vertices until the resulting subgraphs are small enough for balanced workload distribution.

As we shall see in Section 6, the above ideas generalize to numerous subgraph finding problems.

### 3 Limitations of Related Work

In this section, we review existing distributed solutions to subgraph finding, and explain their weaknesses.

**Vertex-Centric In-Memory Solutions.** Most vertex-centric systems are in-memory systems, where vertices (along with their adjacency lists) are partitioned among different machines in a cluster and kept in memory \([3, 14, 6, 22, 31]\). Vertices communicate with each other by message passing, and messages are also buffered in memory to avoid slow disk access.

However, the vertex-centric API is not suitable for subgraph finding, and each vertex \( v_i \) needs to communicate with its surrounding vertices in a breadth-first manner (one more hop per iteration) to get their information for constructing \( G_i \). The solution cannot scale to large graphs since the total volume of (possibly overlapping) decomposed subgraphs may easily exceed the memory capacity of a cluster. Vertex-centric systems also do not provide any mechanism for decomposed subgraphs to share common vertex’s information\

The key problem is, nevertheless, that vertex-centric computation is mainly for data-intensive computation, and generates a large number of messages to transmit for subgraph finding (e.g., for pulling vertices to construct decomposed subgraphs). We call the problem as communication-in-the-chain. In fact, [32] indicates that a vertex-centric program is most scalable if each iteration requires linear computation and communication cost, and it runs for a small number of iterations. This essentially implies that vertex-centric systems are for graph problems with a low computational complexity.

**Vertex-Centric Disk-Based Solutions.** The prohibitive memory requirement can be eliminated using a disk-based system. For example, MapReduce [4] can be used to simulate vertex-centric graph computation (e.g., message sending & receiving) [19], and Pregelix [2] translates a vertex-centric program into a dataflow execution plan for out-of-memory processing. However, the large amount of intermediate data (including messages and subgraphs) need to be dumped to disk and then

---

1 Pregel’s message combiner [15] does not help, since it is to aggregate messages towards the same target vertex, while we consider getting information from the same source vertex.
loaded for each iteration of synchronous computation, making the running time prohibitive. We call the problem as disk-in-the-chain, which adds upon the communication-in-the-chain problem already suffered by a vertex-centric model. In fact, MapReduce even writes intermediate data to Hadoop Distributed File System (HDFS), which is much slower than local disk writes since HDFS replicates each data block on three machines for fault tolerance (termed the remote write problem).

These synchronous frameworks also prevent the computation-intensive processing of decomposed subgraphs from beginning until all decomposed subgraphs are synchronously constructed, leading to CPU underutilization. Data sharing among subgraphs is also not possible since a subgraph is processed by a reducer.

**Systems with Subgraph-Based API.** Recently, NScale [19] and Arabesque [24] attempted to attack subgraph finding problems through a subgraph-based API rather than a vertex-centric one. Albeit becoming more user-friendly, the execution engines of these systems still perform data-intensive processing like vertex-centric solutions mentioned before, and they actually introduce new performance issues.

NScale [19] uses the MapReduce solution we mentioned above, and it brings additional overheads. NScale only supports the top-level decomposed subgraphs, and there is no mechanism to balance workload through recursive decomposition. Assuming that each $G_i$ spans the $k$-hop neighborhood around $v_i$, then NScale first constructs all decomposed subgraphs using $k$ rounds of MapReduce. The large number of decomposed subgraphs are then packed into larger compact subgraphs, each of which can fit in the memory of a reducer. Vertices common to multiple decomposed subgraphs are stored only once in their packed subgraph. Finally, each compact subgraph is distributed to a reducer, which processes all decomposed subgraphs packed in the compact subgraph in memory. Obviously, NScale suffers from all the performance issues of a MapReduce-based vertex-centric solution; moreover, NScale further packs decomposed subgraphs through expensive disk-based computation, and it is very likely that the cost of packing $G_i$ already surpasses that of processing $G_i$ right after it is constructed in memory.

Arabesque [24] proposed an embedding-centric model where an embedding is a subgraph of the input graph $G$. Arabesque requires the entire $G$ to reside in the memory of every machine, and constructs and processes subgraphs iteratively. In the $i$-th iteration, it grows the set of embeddings with $i$ edges/vertices by one adjacent edge/vertex, to construct embeddings with $(i + 1)$ edges/vertices for processing. New embeddings that pass a filtering condition are further processed and then passed to the next iteration. For example, to find cliques, the filtering condition checks whether an embedding $e$ is a clique; if so, $e$ is reported and passed to the next iteration to grow larger clique candidates.

Unfortunately, Arabesque suffers from new performance and scalability issues. Firstly, while previous solutions still permit efficient backtracking within each decomposed subgraph, Arabesque materializes and transmits every single candidate subgraph it examines. Arabesque also compresses/decompresses the large number of materialized embeddings using a data structure called ODAG to save space, which consumes additional CPU cycles. To additionally support frequent subgraph pattern mining, automorphism checking is performed for every newly-expanded embedding to avoid generating duplicate embeddings, which adds unnecessary overhead for subgraph finding. Finally, since $G$ resides in the memory of every machine, scalability is limited by the memory space of a single machine.

**Other Systems.** Blogel [30] and Giraph++ [25] proposed a block-centric model which partitions a graph into disjoint subgraphs called blocks to be distributed among machines for iterative processing, eliminating the need of communication inside each block. However, these systems do not target subgraph finding problems, but rather the acceleration of vertex-centric models.

## 4 System Overview

We now overview the design of G-thinker, including its programming model and system components.

**Programming Model.** G-thinker performs computation on subgraphs. Each subgraph $g$ is associated with a *task*, which performs computation on $g$ and grows $g$ when needed. G-thinker grows subgraphs starting from a set of seed vertices in $V$. For example, in clique enumeration, one may create a task from each vertex $v_i \in V$, which forms the initial subgraph $g$ containing only $v_i$; the task grows $g$ into $G_i$ by pulling vertices in $\Gamma_{gr}(v_i)$ along with their adjacent edges, and then enumerates cliques in $G_i$. In case $G_i$ is too big, users may instead further decompose $G_i$ and create new tasks associated with the newly decomposed subgraphs, which can then be distributed to different machines to improve load balancing.

**G-thinker Components.** A G-thinker program runs on a cluster of workers, $W = \{W_1, W_2, \ldots\}$, where each worker is a basic computing unit that processes its assigned tasks in serial, and a machine may run multiple workers. Each worker alternates between subgraph-centric task computation and vertex pulling (into subgraphs), both are processed in batches. To be memory efficient, we keep the memory requirement of each worker at approximately $O(\frac{d_{avg} \cdot |V|}{|W|})$, where $d_{avg}$ is the average vertex degree.
**G-thinker** partitions the vertices in $V$ (along with their adjacency lists) among different workers, and each worker maintains its assigned vertices in a local table $T_{local}$. Let us denote the $i$-th vertex maintained in $T_{local}$ of worker $W_j$ by $v^j_i$. Figure 1 shows the components of $W_1$, where we can see that $T_{local}$ maintains vertices $v_1^1, v_1^2, v_1^3, \ldots$; each vertex also keeps its neighbors in its adjacency list, so that it can pull its neighbors (along with their adjacency lists) from $T_{local}$ of other workers, by providing the neighbor IDs. The local tables of all workers collectively constitute a distributed key-value store where key is the ID of a vertex $v$ and value is $v$’s adjacency list $\Gamma(v)$.

When a vertex is pulled (along with its adjacency list) from another worker, it is not directly added to the subgraph of the requesting task; instead, it is put in an LRU cache $T_{cache}$. The cache keeps the non-local vertices (i.e., not in $T_{local}$) that are previously received, so that a non-local vertex can be shared by all the tasks that pull it. It is up to the user to decide whether the task’s subgraph should be updated, and if so, what information of that vertex should be added to the subgraph.

As Figure 1 shows, $W_1$’s $T_{cache}$ keeps non-local vertices $v_2^1, v_2^2, v_3^1, \ldots$ (along with their adjacency lists), which are pulled by local tasks previously executed at $W_1$. Note that the adjacency list of a non-local vertex may contain a local vertex, such as $v_1^3$ in the adjacency list of vertex $v_2^3$ in $T_{cache}$ in Figure 1.

During subgraph-centric computation, when a task requires the adjacency list of a vertex $u$, if $u$ is in $T_{local}$ or $T_{cache}$, the task can directly obtain $\Gamma(u)$ by table lookup. Otherwise, the task needs to first pull $u$ from $T_{local}$ of $u$’s worker into the cache table $T_{cache}$ before accessing it.

As Figure 1 shows, each worker also maintains an in-memory task buffer for keeping tasks that are currently being processed, and a disk-based task queue for keeping tasks that are waiting to be processed. This design allows tasks to be processed with high throughput and less redundant communication, as we shall discuss next.

**Batch Processing & Communication Reduction.** A task usually only generates a small number of pull-requests at a time, and sending small messages wastes network bandwidth. Therefore, in **G-thinker**, a worker fetches a batch of tasks from the disk-based task queue at each time, sends their pull-requests together, receives all the requested vertices, and then processes these tasks. Tasks that need to pull more vertices are then added to the task queue for further processing.

Batch processing hides the round-trip delay of each task’s pull-requests, since if tasks are processed one at a time, each task needs to wait for its requested vertices to arrive, which wastes CPU cycles. Batch processing also reduces redundancy in communication. Specifically, if many tasks in a batch pull a remote vertex $u$, only one pull-request needs to be sent, and $\langle u, \Gamma(u) \rangle$ will be received only once and cached in $T_{cache}$ for access by all these tasks. This is in contrast to existing solutions like NScale, where $\langle u, \Gamma(u) \rangle$ needs to be transmitted to every subgraph $g$ that requires it. We organize the task queue $Q$ using locality sensitive hashing (detailed in Section 7), to increase the probability that tasks fetched from $Q$ share common vertices to pull.

To further reduce communication, **G-thinker** allows a user to prune useless items in $\Gamma(v)$ before responding to $\langle v, \Gamma(v) \rangle$ to a worker that pulls $v$. For example, in clique enumeration, a vertex $v$ only needs to respond to a pull-request with $\Gamma_{gt}(v)$ instead of the entire $\Gamma(v)$.

**Memory Cost Analysis.** Since **G-thinker** partitions the vertices evenly among the workers, $T_{local}$ of each worker contains around $O(|V|/|W|)$ vertices. To keep the memory consumption bounded by $O(d_{avg} \cdot |V|/|W|)$, we also set the capacity of $T_{cache}$ to cache at most $O(|V|/|W|)$ vertices at any time, and vertex eviction is based on the LRU (Least Recently Used) policy. As for the memory requirement of tasks, since each task keeps a subgraph $g$, one cannot afford to keep all tasks in memory (e.g., consider all maximal cliques of a graph). Our solution is to keep only a small number (e.g., 1000) of active tasks in memory for batch processing, so that their small subgraphs consume $O(d_{avg} \cdot |V|/|W|)$ memory space.

**Computation & Communication Cost Analysis.** Each task in **G-thinker** (1) pulls required vertices to its subgraph (linear communication cost, and pull requests can further be shared with other tasks) and (2) then performs higher-complexity computation on the subgraph in local machine. Step (2) is computation-intensive and avoids any communication when exploring the large search space by backtracking.

To overlap communication (i.e., Step (1)) with computation (i.e., Step (2)), **G-thinker** treats tasks independently. Different tasks can have different progress, and no synchronization among all machines is required. **G-thinker** proceeds the computation of a task as long as all its requested vertices are locally accessible, and the only communication type in **G-thinker** is point-to-point communication between two workers for vertex pulling, and dynamic task (or decomposed subgraph) reassignment if
Tasks: Subgraph \( g \) 
Context
Functions:
- pull(neighbor)
- add_task(task)
/UDF:
- bool compute(\text{vertex}_\text{frontier})

| Worker \text{<}\text{I, C, V, E}>
| --- |
| Functions: |
| - run(config_info) |
| - add_task(task) |
| /UDF: |
| - seedTask\_gene(vertex) |
| Vertex respond(vertex) |

Figure 2: Programming Interface of G-thinker

```
\begin{figure}
\centering
\includegraphics[width=0.5\textwidth]{figure2.png}
\caption{Programming Interface of G-thinker}
\end{figure}
```

```
```

The Task class. The Task class has another template argument \( <C> \) that specifies the type of context information for a task \( t \), which can be, for example, \( t \)'s iteration number (a task in G-thinker proceeds its computation in iterations). Each Task object \( t \) maintains a subgraph \( g \) and the user-specified context object (of type \( <C> \)). The Task class has only one UDF, \( t\.compute(frontier) \), where the input \( \text{frontier} \) keeps the set of vertices requested by \( t \) in its previous iteration. Each element of \( \text{frontier} \) is actually a pointer to a vertex object in \( T_\text{local} \) or \( T_\text{cache} \). Of course, users may also access \( t \)'s subgraph and context object in \( \text{compute(frontier)} \).

UDF \( \text{compute(frontier)} \) specifies how a task computes for one iteration. If \( t\.compute() \) returns \( \text{true} \), \( t \) needs to be processed by more iterations; otherwise, \( t \)'s computation is finished after the current iteration. In G-thinker, when \( t \) is fetched from the task queue for processing, \( t\.compute() \) is executed repeatedly until either \( t \) is complete, or \( t \) needs a non-local vertex \( v \) that is not cached in \( T_\text{cache} \), in which case \( t \) is added to the task queue waiting for all \( t \)'s requested vertices to be pulled. When a task is completed or queued to disk, G-thinker automatically garbage collects the memory space of the task to make room for the processing of other tasks.

Inside \( t\.compute() \), a user may access and update \( g \) and \( \text{context} \), and call \( \text{pull(u)} \) to request vertex \( u \) for use in \( t \)'s next iteration. Here, \( u \) is usually in the adjacency list of a previously pulled vertex, and \( \text{pull(u)} \) expands the frontier of \( g \). To improve network utilization, \( g \) is usually expanded in a breadth-first manner, so that each call of \( \text{compute()} \) generates pull-requests for all relevant vertices adjacent to \( g \)'s growing frontier. A user may also call \( \text{add_task(task)} \) in \( t\.compute() \) to add a newly-created task to the task queue.

The Worker Class. Each object of the Worker class corresponds to a worker that processes its assigned tasks in serial. Figure 2 shows the key functions of the Worker class, including two important UDFs.

UDF \( \text{seedTask\_gene(v)} \) specifies how to create tasks according to a seed vertex \( v \in T_\text{local} \). A worker of G-thinker starts by calling \( \text{seedTask\_gene(v)} \) on every \( v \in T_\text{local} \), to generate seed tasks and to add them to the disk-based task queue. Inside \( \text{seedTask\_gene(v)} \), users may examine the adjacency list of \( v \), create tasks accordingly (and may let each task pull neighbors of \( v \)), and add these tasks to the task queue by calling \( \text{add_task()} \).

UDF \( \text{respond(v)} \) is used to prune \( \Gamma(v) \) before sending it back to requesting workers. By default, \( \text{respond(v)} \) returns \( \text{NULL} \) and G-thinker directly uses the vertex object of \( v \) in \( T_\text{local} \) to respond. Users may overload \( \text{respond(v)} \) to return a newly created copy of \( v \), with items in \( \Gamma(v) \) properly pruned to save communication (e.g., \( \Gamma_{\text{eq}}(v) \) for clique enumeration). In this case, G-thinker will respond by sending the new object and then garbage-collect it.

Figure 3: Data Types in G-thinker

```
```

5 Programming Interface

G-thinker is written in C++, and it defines two important base classes, Task and Worker, as sketched in Figure 2. To write a G-thinker program, a user needs to subclass Task and Worker with their template arguments properly specified, and implement their abstract functions according to the application logic; these functions are called user-defined functions (UDFs). We remark that although we use C++ terminology here such as “template”, the API is general enough to be implemented in any object-oriented language (e.g., “generic types” in Java).

Data Types. As Figure 2 shows, the Task class takes four template arguments \( <I> \), \( <C> \), \( <V> \) and \( <E> \). Among them, \( <I> \), \( <V> \) and \( <E> \) specify the data types of vertices and edges: (1) \( <I> \): the type of vertex ID; (2) \( <V> \): the type of vertex attribute; (3) \( <E> \): the type of the attribute of an adjacency list item. Other system-defined types (e.g., those for subgraph, vertex, and adjacency list) are automatically derived by G-thinker from them, and can be directly used in the UDFs once a user specifies these three template arguments.

Figure 3 illustrates the inferred system-defined types. Specifically, a subgraph is shown on the left, which consists of a table of vertices (stored with their adjacency lists). The structure of Vertex 3 is shown on the right, where the vertex is stored with its ID (of type \( <I> \)) and a vertex label “c” (of type \( <V> \)), and an adjacency list. Each item in the adjacency list is stored with a neighbor ID (of type \( <I> \)) and an attribute (of type \( <E> \)) indicating the label of the neighbor and the edge label. For example, the first item corresponds to Vertex 1 with label “a”, and the edge label of (3, 1) is “B”. Attributes (i.e., \( <V> \) and \( <E> \)) are optional and are not needed for finding subgraphs with only topology constraints (e.g., triangles, cliques, and quasi-cliques).
The worker class also contains formatting UDFs, e.g., for users to define how to parse a line in the input file on HDFS into a vertex object in $T_{local}$, which will be used during graph loading.

To run a G-thinker program, one may subclass Worker with all UDFs properly implemented, and then call run(config_info) to start the job, where config_info contains job configuration parameters such as the HDFS file path of the input graph.

The Aggregator Class. The Worker class optionally admits a second template argument <aggT>, which needs to be specified if aggregator is used to collect some statistics such as triangle count or maximum clique size. Each task can aggregate a value to its worker’s local aggregator when it finishes. These locally aggregated values can either be globally aggregated at last when all workers finish computing their tasks (which is the default setting), or be periodically synchronized (e.g., every 10 seconds) to make the globally aggregated value available to all workers (and thus all tasks) timely for use (e.g., in compute() to prune search space). In the latter case, users need to provide a frequency parameter.

6 Applications

We consider two categories of applications: (1) finding dense subgraph structures such as triangles, cliques and quasi-cliques, which is useful in social network analysis and community detection; (2) graph matching, which is useful in applications such as querying semantic networks and pattern recognition.

For simplicity, we only consider top-level task decomposition, i.e., we grow each vertex $v_i \in V$ into exactly one decomposed subgraph $G_i$ and every qualified subgraph will be found in exactly one decomposed subgraph.

Triangle Counting. Assume that for any vertex $v$, neighbors in $\Gamma(v)$ are already sorted in increasing order of vertex ID (e.g., during graph loading). We also denote the largest (i.e., last) vertex in $\Gamma(v)$ by $\text{max}_\Gamma(v)$.

We want each triangle $\Delta v_1 v_2 v_3$ (w.l.o.g., $v_1 < v_2 < v_3$) to be counted exactly once, i.e., in $v_i$’s decomposed subgraph. We let $v_1$ count $\Delta v_1 v_2 v_3$ by checking whether $v_3 \in \Gamma(v_2)$. Since $v_1$ only examines $\Gamma(v_2)$ for every neighbor $v_2$ with $v_1 < v_2 < v_3$, $v_1$ only needs to pull the adjacency list of every neighbor in $\Gamma_G(v_1) - \{\text{max}_\Gamma(v_1)\}$. Also, since $\Gamma_G(v_2)$ is only checked against $v_3 > v_2$, $v_2$ only needs to respond $\Gamma_G(v_2)$ to $v_1$.

According to the above discussion, among the UDFs of Worker, respond(v2) creates a copy of $v_2$ with adjacency list $\Gamma_G(v_2)$ for responding; if $|\Gamma_G(v_1)| \geq 2$, seedTask_gene(v1) creates a task $t$ for $v_1$ and let $t$ pull every vertex in $\Gamma_G(v_1) - \{\text{max}_\Gamma(v_1)\}$. The context of $t$ keeps $\text{max}_\Gamma(v_1)$ and a triangle counter count (initialized as 0).

Figure 4: An Example of Graph Matching

In $t$.compute(frontier), frontier contains all the pulled vertices (i.e., $v_2$) in increasing order of their IDs as they were requested in seedTask_gene(v1). We check every $v_2 \in \text{frontier}$ as follows. For each $v_2$, we loop through all vertices $v_3 > v_2$ in $\Gamma_G(v_1)$ ($\Gamma_G(v_1)$ is obtained by appending $\text{max}_\Gamma(v_1)$ in $t$’s context to frontier), and increment $t$’s counter if $v_3 \in \Gamma_G(v_2)$. Finally, compute() returns false since we have checked all $(v_2,v_3)$ pairs and the task is finished.

Whenever a task $t$ is finished, its counter (in $t$’s context) is added to the locally aggregated value, and when all workers finish computation, these local counts are sent to the master to get the total triangle count.

Maximum Clique. We adapt the serial backtracking algorithm of [26] to G-thinker. The original algorithm maintains the size of the maximum clique currently found, denoted by $|Q_{max}|$, to prune the search space.

To allow timely pruning, each worker in our G-thinker program maintains $|Q_{max}|$ and keeps it relatively up to date by periodic aggregator synchronization, so that if a worker discovers a larger clique and updates $|Q_{max}|$, the value can be synchronized to other workers timely to improve their pruning effectiveness. In seedTask_gene(v1), we create a task $t$ whose graph $g$ contains $v_1$, and let $t$ pull all vertices in $\Gamma_g(v_1)$. Then in $t$.compute(frontier), we collect vertices in frontier (i.e., $\Gamma_g(v_1)$), add them to $g$ but filter those adjacency list items that are not in $\{v_i\} \cup \Gamma_g(v_1)$, to form the decomposed subgraph $G_i$, and then run the algorithm of [26] on $G_i$.

This solution can be easily extended to find quasi-cliques, where in a quasi-clique, every vertex is adjacent to at least $\gamma \geq 0.5$ fraction of other vertices. In such a quasi-clique, two vertices are at most 2 hops away $[13]$. The G-thinker algorithm is similar to that for finding maximum clique, except that (1) for each local seeding vertex $v_i$ compute(frontier) runs for up to $n$ iterations to pull vertices (larger than $v_i$) within 2 hops of $v_i$; (2) compute(frontier) then constructs $G_i$ as the 2-hop ego-network of $v_i$ and runs the quasi-clique algorithm of [13] on $G_i$ to compute the quasi-cliques.

Graph Matching. Graph matching finds all subgraph instances in a data graph that match the query graph.
Consider the problem of finding all occurrences of the query graph pattern given by Figure 4(a) in the data graph shown in Figure 4(b). In this example, each vertex in the query graph (and the data graph) has a unique integer ID and a label. We define \( k_1k_2k_3k_4k_5 \) as a mapping where vertex with ID \( k_i \) in the data graph is mapped to vertex \( 1 \) in the query graph. A mapping is a matching if vertex \( k_i \) and vertex \( 1 \) have the same label (for any \( i \)), and for any edge \((1,1)\) in the query graph, the corresponding edge \((k_i,k_j)\) exists in the data graph. For example, 25478 is a matching, while 25178 is not since the data graph does not have edge (1,5) that corresponds to \((3,2)\).

Existing works on distributed graph matching combine vertex-centric graph exploration with subgraph join. Note that when a query graph contains cycles, vertex-centric graph exploration alone is not sufficient. For example, in Figure 4(b), suppose that we perform vertex-centric exploration on the data graph along query graph path \( 3 - 1 - 2 \), we will explore from Vertex 1 (or 4) to 2 and then to 5 simply according to neighbors’ labels. Then, we need to check all \( b \)-labeled neighbors of Vertex 5 to find Vertex 1 (or 4), which is essentially an equi-join on the ID of \( k_3 \). We now present our algorithm, which can be safely skipped if you are not interested in reading the details.

**The Algorithm:** in \( \text{seedTask}_{\text{gene}}(v) \), we only create a task \( t \) for \( v \) if \( v \)'s label is “\( a \)”, and \( \Gamma(v) \) contains neighbors with both labels “\( b \)” and “\( c \)”. If this is the case, we add vertex \( v \) to \( g \), and pull all vertices in \( \Gamma(v) \) with labels “\( b \)” and “\( c \)”.

Then, in iteration 1 of \( t.\text{compute}(\text{frontier}) \), we split \( \text{frontier} \) into two vertex sets: \( V_b \) (resp. \( V_c \)) consists of vertices with label “\( b \)” (resp. “\( c \)”). However, while a vertex in \( V_c \) definitely matches vertex \( 2 \) in Figure 4(a), a vertex in \( V_b \) may match either Vertex \( 3 \) or Vertex \( 4 \). For each vertex \( v_c \in V_c \), we split all vertices in \( \Gamma(v_c) \) with label “\( b \)” into two sets: \( U_1 \) consisting of those vertices that are also in \( V_b \) (i.e., they can match Vertex \( 3 \) or Vertex \( 4 \)), and \( U_2 \) consisting of the rest (i.e., they can only match Vertex \( 4 \) since they are not neighbors of \( v_c \)). We prune \( v_c \), (i) if \( |U_1| = 0 \) since \( v_c \) does not have a neighbor matching Vertex \( 3 \), or (ii) if \( |U_1| = 1 \) and \( U_2 = \emptyset \), since \( v_c \) does not have two neighbors with label “\( b \)”.

Then in iteration 2 of \( t.\text{compute}(\text{frontier}) \), \( \text{frontier} \) contains all pulled vertices with label “\( b \)” that can match Vertex \( 4 \). Let the set of all vertices with label “\( c \)” in \( \text{g} \) (i.e., matching Vertex \( 2 \)) be \( V_c \). Then, for each vertex \( v_b \in \text{frontier} \), we denote the set of all vertices of \( \Gamma(v_b) \) with label “\( d \)” (i.e., matching Vertex \( 5 \)) be \( V_d \); if \( V_d \neq \emptyset \), (1) we add \( v_b \) to \( g \), (2) for every vertex \( v_c \in V_c \cap \Gamma(v_b) \), we add edge \((v_c,v_b)\) (i.e., matching \((2,4)\)) to \( g \), (3) for every vertex \( v_d \in V_d \), we add \( v_d \) and edge \((v_b,v_d)\) to \( g \). Finally, we run a backtracking algorithm on \( g \) to enumerate all subgraphs that match the query graph.

Lastly, we can let \( \text{UDF}_{\text{respond}}(v) \) return a copy of \( v \) by pruning items in \( \Gamma(v) \) whose labels do not fall into \{a, b, c, d\} to save communication.

We remark that only top-level subgraphs decomposed by Vertex \( 1 \) in the query vertex have been considered. If a resulting decomposed subgraph \( G_i \) is still too big, one may continue to decompose \( G_i \) by looking at one more vertex in the query graph (given that Vertex \( 1 \) is already matched to \( v_1 \)).

### 7 System Implementation

**Task Queue & Task Buffers.** Since tasks contain subgraphs that may overlap with each other, it is impractical to keep all tasks in memory. Thus, each worker of \( G\text{-thinker} \) maintains a disk-based queue \( Q \) to keep those tasks waiting to be processed. Figure 5 shows the procedure of task computation on a worker of \( G\text{-thinker} \). Specifically, tasks are fetched from the task queue \( Q \) one at a time and added to a task buffer \( B^T \) for batch process-
ing, while the processed tasks (and those newly-created by `add_task(.)`) are appended to buffer $B^\ell_{\text{in}}$ and then merged to $Q$ in batches.

One baseline approach to organize $Q$ is to treat it as a local-disk stream of tasks, which allows tasks to be sequentially read from (and appended to) $Q$. To support two-sided streaming, we organize tasks in $Q$ with files each containing $C$ tasks, where $C$ is a user-defined parameter (100 by default) to amortize the random IO cost of reading (and writing) a task file. We call this queue organization as stream-queue, which does not consider whether tasks fetched into $Q$ share common vertices to pull. To increase the probability that tasks in $B^\ell_{\text{in}}$ share common vertices to pull, we designed another queue organization called LSH-queue based on min-hashing.

Specifically, assume that a task $t$ has called `compute(.)`, and let us denote the set of vertices that $t$ needs to pull from remote machines by $P(t)$. Before adding $t$ to $Q$, we append a key $k(t)$ to $t$, which consists of a sequence of $\ell$ ($= 4$ by default) MinHash signatures [21] of $P(t)$. Due to the locality sensitivity of min-hashing, for two tasks $t_1$ and $t_2$, the more similar $k(t_1)$ and $k(t_2)$ are, the more likely $P(t_1)$ and $P(t_2)$ overlap [21]. Therefore, we keep tasks in $Q$ ordered by their keys (in alphabetic order), so that tasks fetched tasks from the head of $Q$ have similar keys (e.g., see $B^\ell_{\text{out}}$ in Figure 5) and are likely to share more common vertices to pull. Reducing redundancy by ordering data according to min-hashing keys has been shown to be effective by previous works [19][16].

To avoid random disk IOs, we organize $Q$ as depicted in Figure 5. We maintain an in-memory buffer $B^\ell_{\text{in}}$ to receive incoming tasks (from $B^\ell_{\text{out}}$), and an in-memory buffer $B^\ell_{\text{out}}$ to buffer ordered tasks to be fetched. The waiting tasks on local disk are grouped into files, where each file contains $\lfloor C/2, C \rfloor$ tasks ordered by their keys and $C$ is a user-defined parameter. The ranges of keys in different files are disjoint (except at boundaries where keys may be equal), and all task files are linked in the order of key ranges by an in-memory doubly-linked list $L^\ell$. Each element in $L^\ell$ points to a task file and records its key range. Here, $L^\ell$ is like the leaf level of a $B^+$-tree, but it is small enough to be memory-resident since only metadata of files are kept. When $B^\ell_{\text{out}}$ overflows, we merge all its tasks to the list of task files efficiently by utilizing $L^\ell$ while guaranteeing that each file still contains $\lfloor C/2, C \rfloor$ tasks after merging, using a $B^+$-tree style algorithm.

A computing thread fetches tasks one by one from $B^\ell_{\text{out}}$ for computation, and when $B^\ell_{\text{out}}$ becomes empty, we load to $B^\ell_{\text{out}}$, those tasks in the first file of $L^\ell$ if $L^\ell$ is not empty; otherwise, we fill $B^\ell_{\text{out}}$ with tasks obtained from the head of $B^\ell_{\text{in}}$. Since tasks in a file are clustered by their keys, tasks in $B^\ell_{\text{out}}$ tend to share common vertices to pull.

**Task Computation.** A worker of G-thinker processes its tasks in rounds, where each round consists of three steps, (1) task fetching, (2) vertex pulling, and (3) task computing. Figure 5 illustrates these three steps.

The first step fetches tasks from $Q$ into $B^\ell_{\text{in}}$, until either (i) $B^\ell_{\text{in}}$ becomes full, or (ii) there is no more room in $T_{\text{cache}}$ to accommodate more vertices to pull. The second step then pulls all requested vertices that are not already hit in $T_{\text{cache}}$. Note that the pulling frequency is influenced by the capacity of $B^\ell_{\text{in}}$ and $T_{\text{cache}}$. Now, for every task $t \in B^\ell_{\text{in}}$, its requested vertices in $\text{frontier}$ are either in $T_{\text{local}}$ or $T_{\text{cache}}$, and thus we start the third step to process the tasks in $B^\ell_{\text{in}}$. We compute each task $t$ iteratively until either $t$ is complete, or there exists a newly-requested vertex that is neither in $T_{\text{local}}$ nor in $T_{\text{cache}}$. In the latter case, we compute $t$’s key using $P(t)$, and then add $t$ to $B^\ell_{\text{out}}$. If new tasks are created by $t$, they are also added to $B^\ell_{\text{out}}$. Whenever $B^\ell_{\text{out}}$ is full, it merges its tasks to $Q$.

Since a machine runs multiple workers, and the independence of their execution allows computation and communication to overlap.

**Other Issues.** Real graphs may contain some high-degree vertex $v$, and the task seeded from $v$ may have $|P(t)|$ larger than the capacity of $T_{\text{cache}}$. To allow such a task to proceed, we treat $t$ as a singleton task batch to perform vertex pulling, by temporarily increase the capacity of $T_{\text{cache}}$. After $t$.compute(.) returns, we recover the original capacity of $T_{\text{cache}}$ by evicting overflowed vertices, before starting the next round.

A worker initially seeds the tasks from all vertices in $T_{\text{local}}$ into $Q$, to maximize the opportunity of finding tasks that share common vertices to pull. The seeded tasks are merge-sorted by their min-hashing key to efficiently create the file list of $Q$.

There exists some work that uses heuristics to estimate the computation cost of a task $t$ from its decomposed subgraph [28], and an online regressor may also be trained to improve the cost estimation after each task is finished. Since estimating the cost of $t$ from a partially grown subgraph is difficult, we only estimate $t$’s cost when its decomposed subgraph is fully constructed.

Figure 5: Computation Framework of a Worker

- **Task Fetching**: $B^\ell_{\text{in}} \rightarrow T_{\text{cache}} \rightarrow B^\ell_{\text{out}}$
- **Vertex Pulling**: $\Gamma(v_1) \rightarrow \Gamma(v_2) \rightarrow \cdots$
- **Task Computing**: $t_1 \rightarrow t_2 \rightarrow \cdots$

![Figure 5: Computation Framework of a Worker](image_url)
To allow task prefetching, each worker buffers a small set of tasks with estimated costs, and if all other tasks are exhausted, the work requests tasks from a coordinating master while continuing to process the buffered tasks. The master collects task summary from all workers to decide which tasks to be redistributed when some worker requests more tasks. We remark that task stealing strategies are still under development and are thus not reported in this paper.

8 Experiments

We evaluate the performance of G-thinker, which is implemented in C++ and communicates with HDFS (Hadoop 2.6.0) using libhdfs. All our experiments were run on a cluster of 15 machines, each with 12 cores (two Intel Xeon E5-2620 CPUs) and 48GB RAM. The connectivity between any pair of nodes in the cluster is 1Gbps. All system and application codes are released in G-thinker’s website.

Table 2 shows the four real-world graph datasets used in our experiments. We chose these graphs to be undirected since our applications described in Section 6 are for undirected graphs, while G-thinker can also handle directed graphs. These graphs are also chosen to have different sizes: Youtube, Skitter, Orkut, and Friendster have 2.99 M, 11.1 M, 117 M and 1,806 M undirected edges, respectively.

We ran the algorithms described in Section 4 and list the triangle count, maximum clique size (denoted by \( |Q_{\text{max}}| \)), and number of matched subgraph instances in Table 2. For graph matching, we used the query graph of Figure 4(a) and randomly generated a label for each vertex in the data graph among \{a, b, c, d, e, f, g\} (following a uniform distribution). We can see that the job is highly computation-intensive; e.g., the number of matched subgraphs is in the order of \( 10^{11} \) for Orkut and Friendster.

Recall from Table 5 that each worker in G-thinker maintains four task buffers, \( B_{\text{in}}^0, B_{\text{out}}^0, B_{\text{in}}^1, \) and \( B_{\text{out}}^1 \). We set the capacity of \( B_{\text{in}}^0, B_{\text{out}}^0, B_{\text{in}}^1, \) and \( B_{\text{out}}^1 \) to be the same, which is the maximum number of tasks that are processed in each round. We call this capacity as buffer capacity. We also set the capacity of \( B_{\text{out}}^0 \) to be the file capacity \( C \), since it loads a file of tasks from disk each time.

Unless otherwise stated, the default setting is as follows. Each machine runs 8 workers (i.e., 120 workers in total). The buffer capacity is set to 1000 tasks, and the file capacity \( C \) is set to 100 tasks. Moreover, we set \( T_{\text{cache}} \) to accommodate up to 1 M non-local vertices.

Comparison with Serial Algorithms. We first compare the performance of the serial algorithms for subgraph finding with their distributed G-thinker counterparts. Since serial algorithms need to hold an entire input graph in memory, we run them in a high-end machine with 1TB DDR3 RAM and 2.2GHz CPUs. Table 3 reports the comparison results for triangle counting (with relatively low computation intensity) and graph matching (which is highly computation-intensive). We see that G-thinker is orders of magnitude faster than the serial algorithm for graph matching, but only several times faster for triangle counting. This is because the light computation workload of triangle counting cannot offset the communication cost of vertex-pulling.

Comparison with Other Systems. We also compare G-thinker with Arabesque [24] and Pregelix (version 0.2.12) [2]. Both Arabesque and Pregelix have already implemented triangle counting and maximal clique enumeration, and we used these programs directly for comparison. Unfortunately, neither Arabesque nor Pregelix could successfully finish maximal clique enumeration on even the smallest graph Youtube in our cluster. Arabesque failed after running for 1.5 hours due to memory overflow, while Pregelix reported a frame size error since the 1-ego network of some vertex cannot fit in a frame as required by Pregelix. In contrast, G-thinker found the maximum clique (of size 17) in Youtube in just around 20 seconds.

Table 4 reports the results for triangle counting. We
can see that G-thinker is orders of magnitude faster than both Arabesque and Pregelix, while memory-based Arabesque is a few times faster than disk-based Pregelix. However, Arabesque failed to process Orkut and Friendster due to insufficient memory space. Pregelix failed to process Friendster because it used up the disk space, probably due to its space-consuming B-tree structure for storing vertex data.

Finally, although NScale [19] is not public, [19] reported that it takes 1986 seconds to count the triangles of Orkut (not including the expensive subgraph construction & packing) on a 16-node cluster, while G-thinker takes only 78 seconds on our 15-node cluster.

**Scalability.** We tested the vertical scalability of G-thinker by running various applications with all 15 machines, where each machine runs 1, 2, 4 and 8 workers, respectively. We also tested the horizontal scalability of G-thinker by running various applications with 5, 10 and 15 machines, where each machine runs 4 workers. The results are reported in Table 5 where we can see that G-thinker scales well with the number of workers per machine, and the total number of machines, especially for highly computation-intensive problems like graph matching.

**Effect of System Parameters.** We conducted extensive experiments to study the impact of various parameters on system performance, and find that the performance is mainly sensitive to buffer capacity and the capacity of $T_{cache}$, the impact of other parameters such as the file capacity $C$ is minor. Due to space limit, we only report the results for triangle counting over Friendster. Table 6(a) shows the results when we change the capacity of $T_{cache}$ while keeping all other parameters as default. We can see an obvious reduction in runtime as the capacity of vertex cache increases. Overall, the performance difference is not significantly influenced by the system parameters (e.g., less than doubled), and thus G-thinker is expected to perform well even when the memory space is limited.

**Stream-Queue v.s. LSH-Queue.** Due to space limit, we only report results for triangle counting; the results for maximum clique and graph matching are similar. Table 7 reports the total number of random disk reads and writes incurred by LSH-queue during the whole period of job execution, for the vertical scalability experiments we reported in Table 8, we can see that the number of random IO is small, which demonstrates that LSH-Queue exhibits near-sequential disk IO. We also repeated the experiments using stream-queue instead of LSH-queue, and Table 8 reports the results. Comparing Table 8(a) to Table 8(b) we can see that LSH-Queue improves the performance of most jobs, especially those with heavy computation workload. For some jobs with relatively light workload, the overhead incurred by LSH-Queue (e.g., key computation and random IO) stands out and stream-queue is more efficient.

**Stream-Queue v.s. LSH-Queue.** Due to space limit, we only report results for triangle counting; the results for maximum clique and graph matching are similar. Table 7 reports the total number of random disk reads and writes incurred by LSH-queue during the whole period of job execution, for the vertical scalability experiments we reported in Table 8, we can see that the number of random IO is small, which demonstrates that LSH-Queue exhibits near-sequential disk IO. We also repeated the experiments using stream-queue instead of LSH-queue, and Table 8 reports the results. Comparing Table 8(a) to Table 8(b) we can see that LSH-Queue improves the performance of most jobs, especially those with heavy computation workload. For some jobs with relatively light workload, the overhead incurred by LSH-Queue (e.g., key computation and random IO) stands out and stream-queue is more efficient.

**9 Conclusions and Future Work**

We presented a new framework called G-thinker for scalable subgraph finding, whose computation-intensive execution engine beats existing data-intensive systems by orders of magnitude, and scales to graphs with size two orders of magnitude larger given the same hardware resources. Future work of G-thinker include designing effective task stealing strategy for load balancing, and acceleration through new hardware (e.g., SSD and GPU).
References

[1] C. Bron and J. Kerbosch. Finding all cliques of an undirected graph (algorithm 457). Commun. ACM, 16(9):575–576, 1973.

[2] Y. Bu, V. R. Borkar, J. Jia, M. J. Carey, and T. Condie. Pregelix: Big(ger) graph analytics on a dataflow engine. PVLDB, 8(2):161–172, 2014.

[3] A. Ching, S. Edunov, M. Kabiljo, D. Logothetis, and S. Muthukrishnan. One trillion edges: Graph processing at facebook-scale. PVLDB, 8(12):1804–1815, 2015.

[4] J. Dean and S. Ghemawat. Mapreduce: Simplified data processing on large clusters. In OSDI, pages 137–150, 2004.

[5] J. Gao, C. Zhou, J. Zhou, and J. X. Yu. Continuous pattern detection over billion-edge graph using distributed framework. In ICDE, pages 556–567, 2014.

[6] J. E. Gonzalez, Y. Low, H. Gu, D. Bickson, and C. Guestrin. Powergraph: Distributed graph-parallel computation on natural graphs. In OSDI, pages 17–30, 2012.

[7] J. E. Gonzalez, R. S. Xin, A. Dave, D. Crankshaw, M. J. Franklin, and I. Stoica. Graphx: Graph processing in a distributed dataflow framework. In OSDI, pages 599–613, 2014.

[8] H. He and A. K. Singh. Graphs-at-a-time: query language and access methods for graph databases. In SIGMOD, pages 405–418, 2008.

[9] X. Hu, Y. Tao, and C. Chung. I/o-efficient algorithms on triangle listing and counting. ACM Trans. Database Syst., 39(4):27:1–27:30, 2014.

[10] G. Kasneci, F. M. Suchanek, G. Ifrim, M. Ramanath, and G. Weikum. NAGA: searching and ranking knowledge. In ICDE, pages 953–962, 2008.

[11] S. Khuller and B. Saha. On finding dense subgraphs. In ICALP, pages 597–608, 2009.

[12] J. Lee, W. Han, R. Kasperovics, and J. Lee. An in-depth comparison of subgraph isomorphism algorithms in graph databases. PVLDB, 6(2):133–144, 2012.

[13] G. Liu and L. Wong. Effective pruning techniques for mining quasi-cliques. In PKDD, pages 33–49, 2008.

[14] Y. Low, J. Gonzalez, A. Kyrola, D. Bickson, C. Guestrin, and J. M. Hellerstein. Distributed GraphLab: A framework for machine learning in the cloud. PVLDB, 5(8):716–727, 2012.

[15] G. Malewicz, M. H. Austern, A. J. C. Bik, J. C. Dehnert, I. Horn, N. Leiser, and G. Czajkowski. Pregel: a system for large-scale graph processing. In SIGMOD Conference, pages 135–146, 2010.

[16] J. Mondal and A. Deshpande. Eagr: supporting continuous ego-centric aggregate queries over large dynamic graphs. In SIGMOD, pages 1335–1346, 2014.

[17] J. Pattillo, N. Youssef, and S. Butenko. On clique relaxation models in network analysis. European Journal of Operational Research, 226(1):9–18, 2013.

[18] L. Qin, J. X. Yu, L. Chang, H. Cheng, C. Zhang, and X. Lin. Scalable big graph processing in mapreduce. In International Conference on Management of Data, SIGMOD 2014, Snowbird, UT, USA, June 22-27, 2014, pages 827–838, 2014.

[19] A. Quamar, A. Deshpande, and J. Lin. Nscale: neighborhood-centric large-scale graph analytics in the cloud. The VLDB Journal, pages 1–26, 2014.

[20] L. Quick, P. Wilkinson, and D. Hardcastle. Using pregel-like large scale graph processing frameworks for social network analysis. In ASONAM, pages 457–463, 2012.

[21] A. Rajaraman, J. D. Ullman, J. D. Ullman. Mining of massive datasets, volume 1. Cambridge University Press Cambridge, 2012.

[22] S. Salihoglu and J. Widom. GPS: a graph processing system. In SSDBM, page 22, 2013.

[23] Y. Shao, B. Cui, L. Chen, L. Ma, J. Yao, and N. Xu. Parallel subgraph listing in a large-scale graph. In SIGMOD, pages 625–636, 2014.

[24] C. H. C. Teixeira, A. J. Fonseca, M. Serafini, G. Siganos, M. J. Zaki, and A. Aboulnaga. Arabesque: a system for distributed graph mining. In SOSP, pages 425–440, 2015.

[25] Y. Tian, A. Balmin, S. A. Corsten, S. Tatikonda, and J. McPherson. From “think like a vertex” to “think like a graph”. PVLDB, 7(3):193–204, 2013.

[26] E. Tomita and T. Seki. An efficient branch-and-bound algorithm for finding a maximum clique. In DMTCS, pages 278–289, 2003.
[27] W. Wu, H. Li, H. Wang, and K. Q. Zhu. Probase: a probabilistic taxonomy for text understanding. In SIGMOD, pages 481–492, 2012.

[28] J. Xiang, C. Guo, and A. Aboulnaga. Scalable maximum clique computation using mapreduce. In ICDE, pages 74–85, 2013.

[29] D. Yan, Y. Bu, Y. Tian, and A. Deshpande. Big graph analytics platforms. Foundations and Trends in Databases, 7(1-2):1–195, 2017.

[30] D. Yan, J. Cheng, Y. Lu, and W. Ng. Blogel: A block-centric framework for distributed computation on real-world graphs. PVLDB, 7(14):1981–1992, 2014.

[31] D. Yan, J. Cheng, Y. Lu, and W. Ng. Effective techniques for message reduction and load balancing in distributed graph computation. In WWW, pages 1307–1317, 2015.

[32] D. Yan, J. Cheng, K. Xing, Y. Lu, W. Ng, and Y. Bu. Pregel algorithms for graph connectivity problems with performance guarantees. PVLDB, 7(14):1821–1832, 2014.

[33] L. Zou, L. Chen, and M. T. Özsu. Distancejoin: Pattern match query in a large graph database. PVLDB, 2(1):886–897, 2009.