**DENSEST SUBGRAPH DISCOVERY ON THE GPU**

by

Hunter G. Gareau

A Thesis

Submitted to the

Department of Computer Science

College of Science and Mathematics

In partial fulfillment of the requirement

For the degree of

Master of Science in Computer Science

at

Rowan University

August 19, 2025

Thesis Chair: Guimu Guo, Ph.D., Assistant Professor, Department of Computer Science

Committee Members:

Sihan Yu, Ph.D., Assistant Professor, Department of Computer Science

Richard Rabbitz, MSc., Assistant Teaching Professor, Department of Computer Science

© 2025 Hunter G. Gareau

**Acknowledgements**

I would like to deeply thank Professor Guimu Guo for his guidance throughout our research. We have worked together for a long time on this project, and I wouldn’t have been able to do any of this without his skills, direction, and help.

I would like to thank Professor Sihan Yu and Professor Richard Rabbitz for being a part of my committee, being available to assist me in this endeavor, and for reading over and reviewing my paper.

I would also like to thank Wajid Manzoor, a Ph.D. student and fellow student of Professor Guo, who was a massive help on developing the CUDA code.

# **Abstract**

Hunter Gareau

DENSEST SUBGRAPH DISCOVERY ON THE GPU

2024 – 2025

Guimu Guo, Ph.D.

Master of Science in Computer Science

The Densest Subgraph Discovery (DSD) problem is a prevalent problem in the field of graph mining, aiming to find the subgraph of highest edge to vertex ratio. There have been many solutions to this problem. However, one avenue that other graph mining problems have gone down is parallel programming, which executes many commands simultaneously for more efficient programs. What we have done is applied parallel programming on the GPU to a DSD solution called CoreExact in a two-fold effort. This is done in hopes of both creating a more efficient solution to the DSD while also serving as further research and precedent for the efficacy of parallel programming using the GPU.

**Table of Contents**

[Abstract iv](#_Toc203129859)

[List of Figures vi](#_Toc203129860)

[List of Tables vii](#_Toc203129861)

Chapter 1: [Introduction 1](#_Toc203129862)

[The Densest Subgraph Discovery Problem 1](#_Toc203129863)

[The GPU And Parallel Programming With CUDA 3](#_Toc203129864)

Chapter 2: [Existing Solutions and Related Works 7](#_Toc203129865)

[Goldberg’s Maximum Flow-Based Algorithm 7](#_Toc203129866)

[Greedy Peeling Algorithm 8](#_Toc203129867)

[The Greedy++ Algorithm 9](#_Toc203129868)

[Densest K-Subgraph Approximation 9](#_Toc203129869)

[DSD for Directed Graphs 11](#_Toc203129870)

[Optimal Quasi-Clique Problem 12](#_Toc203129871)

Chapter 3: [The CoreExact Algorithm 13](#_Toc203129872)

[Adjacency Lists and Adjacency Matrices 13](#_Toc203129873)

[K-Cores and Decomposition 15](#_Toc203129874)

[Connected Components 19](#_Toc203129875)

[Flow Networks and Minimum st-Cuts 20](#_Toc203129876)

**Table of Contents (Continued)**

[An in Depth Look at Goldberg’s Algorithm 24](#_Toc203129877)

[The Full Algorithm 28](#_Toc203129878)

Chapter 4: [Our Parallelized Implementation 31](#_Toc203129879)

[DAG Generation 32](#_Toc203129880)

[Finding all Cliques 33](#_Toc203129881)

[Clique-Based Core Decomposition 35](#_Toc203129882)

[Construction and Pruning of the Densest K-Core 37](#_Toc203129883)

[Component Decomposition 40](#_Toc203129884)

[Bound Generation 41](#_Toc203129885)

[Flow Network Creation 43](#_Toc203129886)

[In Summary 44](#_Toc203129887)

Chapter 5: [Experiment and Evaluation 46](#_Toc203129888)

Chapter 6: [Conclusion 51](#_Toc203129889)

[References 52](#_Toc203129890)

# **List of Figures**

Figure Page

[Figure 1. DSD Example. 2](#_Toc199072638)

[Figure 2. GPU Architecture [5]. 5](#_Toc199072639)

[Figure 3. Adjacency List Representation [9]. 14](#_Toc199072640)

[Figure 4. Adjacency Matrix Representation. 15](#_Toc199072641)

[Figure 5. Graph with outline of 1-core, 2-core, and 3-core [11]. 16](#_Toc199072642)

[Figure 6. Connected components example [12]. 20](#_Toc199072643)

[Figure 7. Flow Network example [13]. 22](#_Toc199072644)

[Figure 8. st-cut example. 23](#_Toc199072645)

[Figure 9. Minimum st-cut example. 24](#_Toc199072646)

[Figure 10. CoreExact flow network. 26](#_Toc199072647)

[Figure 11. CoreExact algorithm [2]. 29](#_Toc199072648)

[Figure 12. k-clique examples [17]. 31](#_Toc199072649)

[Figure 13. Processing vertex neighbors by thread. 33](#_Toc199072650)

[Figure 14. Processing vertices by thread. 36](#_Toc199072651)

[Figure 15. Handling cliques by warp and its edges by thread. 39](#_Toc199072652)

[Figure 16. Handling components by warp. 42](#_Toc199072653)

[Figure 17. Handling components by warp and their vertices by thread. 44](#_Toc199072654)

# **List of Tables**

Table Page

[Table 1. Graphs by Number of Edges and Vertices 47](#_Toc203481083)

[Table 2. Highest Density by Graph (Rounded to 2 Decimal Places) 48](#_Toc203481084)

[Table 3. Graphs by DSD Runtime Over Edge Density 48](#_Toc203481085)

[Table 4. Graphs by DSD Runtime Over 3-Clique Density 49](#_Toc203481086)

**Chapter 1**

# **Introduction**

When it comes to analyzing large or complex groups of data, it is often useful to examine the connections and relationships shared between its subjects, and graphs can be used to model such relations. Graphs are composed of two components: vertices (which represent individual members of a data set) and edges (which represent the connections between these members). A graph is usually represented as *G = (V, E)*, where *V* is the set of vertices in the graph, and *E* is the set of edges in the graph. A couple of the major types of data that graphs can be used to model are social networks (e.g. Facebook, Twitter, etc.) and biological data (DNA, neural networks, etc.) [1]. It’s also worth noting that edges can be directed, as in a connection between two vertices specifically goes from one to another. A directed graph can be used to model other types of data sets, or specific types of relations, such as people following others in an online social network. The analyzation of these graphs using various tools or techniques to find additional data and patterns is known as *graph mining*.

## **The Densest Subgraph Discovery Problem**

While there are many studies and problems in graph mining, a fundamental one is known as the *densest subgraph discovery problem* (DSD). The aim of the DSD is that given an undirected graph *G*, you must find a subgraph *S* such that it has the highest density of all subgraphs of *G*. The definition of edge density is as follows:

Definition 1 (Edge Density [2]): Given a graph G (V, E), edge density is

Additionally, density can also be applied to network motifs, which are small structures of vertices and edges such as shapes or cliques:

Definition 2 (Motif Density [2]): Given a graph G (V, E) and its set of motifs of a chosen type M, its motif density is .

Motif Density can be applied as edge density or clique density, so it is the definition we will use. We can define the DSD as such:

Definition 3 (Densest Subgraph Discovery Problem): Given a graph G (V, E) and its set of motifs of a chosen type, the densest subgraph *S* is such that and the motif density of all other subgraphs of G.

A diagram of a network

AI-generated content may be incorrect.

Figure 1. DSD Example.

The denser a graph is, the more connected the members of that graph are. In simple terms, the DSD aims to find the most connected group of vertices within a graph. An example of a densest subgraph is found in Figure 1, where the subgraph of the pink vertices is denser than all other subgraphs in the graph. The densest subgraph (and therefore a solution to the DSD) is a notable piece of information to have for a dataset and has plenty of notable applications in real data sets including finding and filtering out fake users or identifying echo chambers in social networks, or identification of regulatory motifs in DNA or gene annotation graphs in biological data [1].

## **The GPU And Parallel Programming With CUDA**

Being such a notable problem, there are many solutions to the DSD. However, all are serialized programs that run on the CPU. By the nature of graph mining, it is almost always required to process every vertex and edge in these large graph datasets, which can take a notable amount of time in a serialized program processing these one by one. But by programming in parallel, a great number of vertices can be processed in concurrence and thus be more efficient. This is where the GPU comes in. The CPU and GPU differ in their processing cores. The CPU runs on a handful of powerful processing cores that can take complex orders, while the GPU runs on many weaker processing cores that take simpler orders. So, while a powerful CPU can run a taxing program efficiently, utilizing the full power of a GPU by running commands in parallel across its many cores is much more computationally efficient. Although this comes with the limitation that the GPU’s commands are much more restrictive than that of the CPU [3].

These processes and commands are all run across threads, which can be thought of as a unit of execution within an overall process with its own data register. For example, if you had an array of integers that you needed to add 10 to individually, each thread of the GPU would take an integer from the array and add 10 to it at the same time. Threads are grouped hierarchically. 32 threads make up a warp, with every thread in a warp following the same instructions. Data is easily shared between thread registers in a warp. A step up from this, some number of warps make up a block (also referred to as a thread block). The number of warps per block varies between GPUs, and the blocks are mapped to the processors of the GPU. There is a cache of shared memory for the warps and threads in a block, but the collective register memory of every thread is often greater than the amount of available memory in the block’s cache. Meaning that data can be shared between warps in a block, but it is more computationally taxing. The final level of the hierarchy is grids, which are made up of blocks. When a kernel (a process to be run in parallel) is launched, that kernel gets mapped to a grid using the processors mapped to the grids blocks to complete the task. The entire GPU shares an L2 cache and a global memory, but sharing memory across blocks, warps, and threads using this memory is taxing due to the synchronous nature of the threads and memory updates being dependent. This architecture is visualized in Figure 2. With this architecture, work and different processes can be divided up between blocks, warps, and threads, allowing the user notable control and options over what can be processed and how it can be worked with in parallel [4].

A diagram of a computer code

AI-generated content may be incorrect.

Figure 2. GPU Architecture [5].

So, the computational power and structure of parallel programming on the GPU is well suited for graph mining, and thus the DSD. With the complexities of graph mining and the DSD however, one would have a very tough time writing a program to analyze a graph only using the simple commands available to the GPU. But thanks to NVIDIA, there’s a tool to work around that. CUDA (which stands for Compute Unified Device Architecture) is a parallel computing platform and application programming interface model. CUDA allows for a serialized C++ program run on the CPU to execute kernels on the GPU and transfer data between the CPU and GPU, allowing the user to leverage the massive computational power of the GPU as needed [3]. This interaction between serialized code and running kernels on the GPU can be seen in Figure 2.

This is the goal of the project, creating a solution to the DSD which utilizes the GPU through CUDA. This goal is two-fold, both being a more efficient solution to the DSD, as well as providing further research on the computational power of parallel programming on the GPU.

**Chapter 2**

# **Existing Solutions and Related Works**

As previously stated, there are many existing solutions to the DSD. Algorithms that have been created and developed to be increasingly more efficient. Some algorithms will get the exact solution to the DSD, meaning these algorithms will return the densest subgraph without fail. Other algorithms will approximate the densest subgraph, finding one of the densest subgraphs, if not the densest. So, it’s worth covering some of these algorithms to see how others have solved it before, and to introduce some important topics and ideas.

### **Goldberg’s Maximum Flow-Based Algorithm**

This algorithm was first developed by Andrew Goldberg in 1984. The basic outline of this algorithm follows the construction of a flow network based on the given graph, where every vertex is connected to two newly added vertices *s* and *t* (the source and sink). A binary search is run on this network, maintaining an upper and lower bound on the greatest density, and tightening these bounds with every iteration until the lower bound is within the margin of being exact. The flow network and upper and lower bounds are updated using a maximum flow (or minimum st-cut) approach which will try to return a subgraph of density equal to the average of the upper and lower bounds [6]. The time complexity of this algorithm depends on the implementation of the mininmum st-cut, to the point where the complexity given to Algorithm 1 (Goldberg’s algorithm) of [6] is such that the time complexity is given by where *T* is the complexity of the minimum st-cut algorithm used and *v* is the number of vertices, being log(*v*) iterations of the minimum st-cut algorithm since it’s a binary search on the size of the graph. There is a lot to unpack and understand with this algorithm, but these topics will be explored more in depth in the next chapter.

### **Greedy Peeling Algorithm**

This is a straightforward approximation algorithm developed by Moses Charikar in 2000 [7]. Given graph *G*, the algorithm will remove the vertex of lowest degree from *G* every iteration (where degree is the number of vertices a given vertex is connected to). The current subgraph of highest density is stored, and every iteration it is compared to the new subgraph. If the new subgraph is denser than the currently stored one, it replaces the old one. This goes on until all vertices have been removed, and thus the subgraph that had the highest density as these vertices were peeled away is the result. This is not a very reliable algorithm to find the densest subgraph as it leaves a lot of possible subgraphs unchecked. However, it was found and proven that through this method, the resulting subgraph would always be *at least* half as dense as the actual densest subgraph [6]. Meaning this is a decent algorithm for finding dense subgraphs, and there is notable premise in the idea of peeling off vertices by lowest degree. Additionally, its time complexity is straightforward, being where *v* and *e* are the number of vertices and edges for a non-weighted graph. This complexity is due to the fact that after the initial scan for degrees, each vertex and edge is only operated on once, being when they are removed, thus the number of steps is reliant on the number of vertices plus the number of edges [6].

### **The Greedy++ Algorithm**

Building on Charikar’s Greedy Peeling Algorithm is an algorithm proposed by D. Boob, Y. Gao, R. Peng, and J. Wang in 2019 [6]. This algorithm iterates through the Greedy algorithm *T* times, with each iteration updating the priority of each vertex so that in subsequent iterations, vertices of higher priority are kept in for longer. By running the Greedy algorithm multiple times and utilizing the results of previous iterations, denser subgraphs can be found. The time complexity for this is similar to the original Greedy algorithm with , but with two notable changes. First, since priority is being used,each selection of the lowest degree vertex takes a bit of additional time to determine and update priority, coming out to time in a binary search for each vertex, getting us to . And since we are doing a chosen number *T* iterations, we find that the time complexity of Greedy++ is . Additionally, while this algorithm is still an approximation, it was found that with enough iterations, the densest subgraph found would be arbitrarily close to the exact optimal densest subgraph. Meaning this relatively simple solution can find a subgraph of negligible difference from the optimal densest subgraph at worst, making for an efficient and relatively easy to implement solution to the DSD [6].

### **Densest k-Subgraph Approximation**

This is a notable variant of the DSD where the subgraphs being searched for are specifically of size *k* (subgraphs with *k* vertices). This can be useful information in various contexts and is also an interesting problem to dissect on its own since it essentially reduces the DSD from looking for to just In other words, you’re looking for the subgraph of size *k* with the most edges.

There are several existing solutions to this problem, but one of the major ones was developed by U. Feige, G. Kortsarz, and D. Peleg in a combinatorial approximation algorithm [8]. The basic algorithm (referred to as Algorithm A in the paper) combines three procedures for finding a dense subgraph of size *k* and returns the densest of the three. The first procedure simply takes *k/2* random edges and returns the set of vertices connected by these edges, adding additional arbitrary vertices to get the set to size *k* if needed. This procedure provides a baseline, always returning a subgraph of density ≥ 1. The following 2 procedures act as Greedy approximations to try and find an even denser subgraph. The first of the two sorts the vertices by degree and takes the *k*/2 vertices of highest degree into the subset, then re-sorts the remaining vertices by how many neighbors they have in the initial subset, and adding the last *k*/2 vertices from the top of that ranking. The second of these procedures constructs a subgraph for every vertex *a* in the graph, being constructed by ranking how many 2-step paths each vertex has to *a* and then ranking the neighbors of *a* by how many of those vertices they’re connected to, followed by taking the *k*/2 vertices of the highest degree connected to those vertices. The union of this set and the 2-step neighbors serves as the resulting subgraph, adding in arbitrary vertices if the result does not reach *k*. As stated before, the result of these three procedures with the greatest density is taken as the result of the algorithm. This algorithm is of accuracy . In the same paper, they dive into ways to approximate even closer, and other algorithms have been made to approximate the densest k-subgraph which have a higher accuracy, but this algorithm served as the first notable solution [6].

### **DSD for Directed Graphs**

As was mentioned earlier, the definition of density we utilize only applies to undirected graphs, as that definition does not take directionality into account. A definition for density in directed graphs was proposed by Kannan and Vinay in 1999 [2], and is still in use today. It is defined as follows:

Definition 4 (Directed Graph Density [9]): Given a directed graph G (V, E), let and E (S, T) is all the edges from S to T. We define the density of these sets as

In layman’s terms, directed graph density represents how connected one set of vertices is to another. And as follows, the Directed Densest Subgraph Problem (DDSP) aims to find sets such that *d(S, T)* is maximized.

There are multiple solutions to this problem as one may expect, and one such algorithm developed by Charikar will find the exact solution. This solution is a Relaxed Linear Programming problem based on the value of *|S| / |T|*. As proven in [7], the optimal value of the linear programming problem on *|S| / |T|* is equivalent to the optimal directed density, where the optimal sets *S* and *T* can be computed from the results of the LP [6].

### **Optimal Quasi-Clique Problem**

The definition of density most widely used is the one covered in Definition 1. However, some have put forward different definitions to gather different results. One such definition is called Edge-Surplus:

Definition 5 (Edge – Surplus [10]): Given graph G (V, E), let , where S (X, Y). Given some functions g and h, and some > 0, we define edge surplus as:

. In the case of .

This definition is flexible, as the definition of *g*, *h*, and are determined for the specific case. But this framework sets up the problem such that it favors more edges and penalizes more vertices, hence why is positive, and why is negative.

This definition can be used to evaluate the DSD as normal, but in the case of quasi-clique: , , and [6]. These definitions favor subgraphs that are tighter knit and have a small diameter (where diameter is the longest path in a graph). This differs from the normal density definition which does not distinguish graph size. And as follows, the Optimal Quasi-clique Problem (OQP) aims to find the subgraph that maximizes the value of this function. One such solution to OQP is based on the Greedy peeling algorithm for the DSD. The main differences lie in that this algorithm checks for quasi-clique value rather than standard edge density, and that they increase efficiency by keeping lists of all possible degree values, updating them as vertices are removed and using them to decide which vertex to remove next [10].

**Chapter 3**

# **The CoreExact Algorithm**

While these various solutions are all very interesting to discuss and investigate, let’s cover the most important algorithm of our project. Our program is based on the CoreExact algorithm [9]. This is the algorithm that will be rewritten into a C++/CUDA program, but to understand the implementation we should first explain CoreExact itself. There are a lot of concepts and smaller algorithms that make up CoreExact though, so each will be covered before going over the whole algorithm.

## **Adjacency Lists and Adjacency Matrices**

Because algorithms must be applied to actual code, graphs must be implemented in a more numerical way. There are multiple ways to do so, but the method used for graphs in our program (and in the CoreExact program) is the Adjacency List. This is a straightforward way of storing how vertices are connected using a 2-dimensional list (a list of lists). The first dimensional list represents every vertex in the graph, where the index matches the vertex number. The second dimension is a list of all the vertices that the represented vertex is connected to. A representation of this is shown in Figure 3. Of note, because our graph is undirected the edges will be stored bidirectionally, so if an edge connects vertices 1 and 2, 2 and 1 show up in their lists respectively.



Figure 3. Adjacency List Representation [11].

Additionally, since CoreExact is allowing for the use of motifs other than regular edges the motif structures also need to be stored. In theory, adjacency lists could be used here as well. However, CoreExact uses adjacency matrices to do so. Adjacency matrices, in comparison to adjacency lists, work well for small structures like motifs as they make construction and representation easier in the code (which holds true for large graphs as well, but they become much more costly in memory and iteration efficiency).

As the name suggests, the adjacency matrix represents graphs through a full 2-dimensional matrix. There is a row and column for every vertex, with each intersection being filled with a 0 or 1. If the intersection has a 0, the two vertices are not connected by an edge, whereas a 1 indicates they are connected. A representation of this is shown in Figure 4, where the graph on the left has its edges represented by 1’s in the table matrix.

A diagram of a network

AI-generated content may be incorrect.

Figure 4. Adjacency Matrix Representation.

In the actual code, both adjacency lists and matrices are handled as 2-dimensional vectors, but their structures still reflect what has been outlined above.

## **K-Cores and Decomposition**

A major aspect of the CoreExact methodology is that we can prune down the input graph to a subgraph and break that down into parts to save time searching. The first of these steps is finding the densest k-core.

A k-core is a graph in which every vertex is connected to at least *k* other vertices. In a 1-core, every vertex is connected to at least one other vertex. In a 2-core, every vertex is connected to at least 2 other vertices, and so on. But within a large graph, there are likely to be subgraphs of a much higher *k* value. To find these graphs, we can use a peeling technique like the Greedy peeling algorithm covered earlier called core decomposition. Like with the Greedy algorithm, core decomposition finds the vertex of lowest degree one at a time and removes it from the graph, where degree is the number of vertices a vertex is connected to. Using a 1-core as an example, if you remove all the vertices of degree 1 (and all subsequent vertices that have degree 1 after the removal of the other vertices), you will be left with only vertices that have degree 2 or higher, going from a 1-core to a 2-core. Going until all vertices have been removed will mean every k-core up to the highest possible *k* value will have been found. So, using this method you can record the information on all these k-core subgraphs, and as stated earlier this can be done in time (and as a reminder, *e* is the number of edges in the graph and *v* is the number of vertices).

Figure 5 outlines an example of a graph which can be broken down into a 1-core, 2-core, and 3-core. The orange vertices are in the 1-core since they all have a degree of 1 (they’re only connected to 1 vertex) and are removed at the first stage. The blue vertices are in the 2-core since they only have a degree of 2 after the orange vertices’ removal. This leaves the red vertices, which are all connected to 3 vertices each, leaving a 3-core.



Figure 5. Graph with outline of 1-core, 2-core, and 3-core [12].

And this process and terminology applies to any motif, not just edges. A 1-core of triangles means every vertex is part of at least one triangle, a 2-core means every vertex is part of at least 2 triangles, and so on. The decomposition process is the same as well, since you remove vertices one at a time based on motif degree. And motif core decomposition is the first step of the CoreExact algorithm.

However, to begin motif core decomposition in the code, every motif instance in the graph must be found (since individual motif data is not stored for each graph, only edges). To do so, you first need to generate a DAG of the graph (directed acyclic graph). Finding motifs requires branching down all possible paths, meaning any loops in the graph would cause potential repeats. And DAGs are designed such that they have no loops. In this case, it is set up so that all edges are directed from the vertex of higher degree to the vertex of lower degree. This means the first step is regular edge decomposition, which has already been covered. Then it’s a simple process of generating a “new” graph where each edge is only stored from the higher degree vertex to the one of lower degree, rather than being stored for both directions.

Once the DAG is created, you can find all the motifs in the graph. This process is done with a recursive algorithm. It is called on the DAG, and it passes the size of our chosen motif to it: *k*. By branching down paths in the DAG that only result in a list of vertices of size *k*, it is effectively making it so only possible motifs are checked. This algorithm can be divided into two cases: if *k* is equal to 2 (the motif type is edge), or not. In the former case, for every neighbor of the current vertex, a complete motif is found. So, for each neighbor the number of motifs found and the motif degree of the vertex are updated, and the path of vertices is added to the motif list. In the latter case, we check every vertex in our current cluster of neighbors (this starts as the list of all vertices in the DAG). For each vertex, add all neighbors of the vertex to a new list that are valid for the current recursive level *k* (level of connection to other vertices). Then after updating the degrees of every neighbor to reflect the next recursive level, make a recursive call, passing the current path, *k* – 1 (to reflect the next recursive level), and the list of valid neighbors. This is an in-depth algorithm, but all that is important to know is that it finds all the motifs in the graph, and stores motif information such as the number of motifs in the graph and the motif degrees of every vertex.

Once motif information is gathered, it’s just a matter of doing motif core decomposition, which as covered is basically the same as edge decomposition. In summary, we do edge decomposition, DAG generation, motif listing, and then motif core decomposition can be performed, resulting in finding every motif k-core. And the densest of these k-cores must contain the densest subgraph, as proven in [9]. This comes down to two steps. First, proving the lemma that removing any vertex from the densest subgraph will result in several motifs being removed equal to at least the ceiling value of the density of the densest subgraph (we’ll call this value ⌈*p*⌉). This was proven by contradiction. If you assume that the statement is false, and that less than ⌈*p*⌉ motifs are removed, you get a subgraph with a higher density than the previous subgraph. However, that is a contradiction, since we know that the previous subgraph was the densest subgraph. Proving the lemma. So intuitively, this means that any given vertex in the densest subgraph is a part of ⌈*p*⌉ motifs. Which is the same as the definition of a k-core, where any vertex has participated in at least *k* motifs. Thus, we can conclude that the densest subgraph must be in the k-core where *k* = ⌈*p*⌉.

## **Connected Components**

The densest k-core can be broken down into pieces known as connected components. A connected component is a subgraph where every vertex is connected to all the others in the subgraph by *some* path. In Figure 6 is an example of a graph of connected components, where the graph can be broken into 3 disjointed subgraphs {*V1, V2, V3, V4, V5, V6}, {V7, V8, V9}, and {V10, V11, V12}*. In layman’s terms, a connected component is a subgraph that has no connections to the vertices of the other connected components. Not every graph is going to have connected components, but it’s worth checking for as the connected components of the densest subgraph have the same density as each other, and by breaking down the k-core, it may be found that certain connected components are denser than others [9].



Figure 6. Connected components example [13].

To decompose into connected components in the code, multiple steps occur. First off, connected components are found using a breadth first search, which checks each vertex and edge once, resulting in time complexity. However, after this the connected component subgraphs need to be constructed and added to a list. Most of this is negligible for time complexity save for working with the motifs. Processing motifs requires processing each vertex of each motif. So, for the number of motifs *m* and the size of our motif *w* is time complexity . Since the rest of the process is negligible, the time complexity comes down to .

## **Flow Networks and Minimum st-Cuts**

One of the most important pieces of CoreExact is the usage of Goldberg’s maximum flow algorithm. A key aspect of that algorithm is a constantly updating flow network using the next best guess for greatest density, and then using minimum st-cut to try and return a subgraph of at least that density. Therefore, to understand how Goldberg’s algorithm works it’s important to understand what flow networks are and what a minimum st-cut is.

A flow network is a special type of directed graph where every edge has two values: capacity and flow. These are abstract numerical values that represent an amount of “something” passing through the edge in the direction it faces. Water through a pipe is an easy way to look at this, since water flows through a pipe in a certain direction. Capacity represents the maximum amount of this “something” that can pass through the edge, while flow is an amount greater than or equal to 0 that is less than or equal to the capacity, essentially representing how much of this “something” is passing through the edge. This is often represented as *x/y*, where *x* is the flow and *y* is the capacity. Using the water pipe example, a pipe may be able to sustain up to 5 liters of water at any moment, but the amount of water passing through may be anywhere from 0 to 5 liters. Flow networks have two special vertices called the source and the sink, usually represented as *s* and *t* respectively. The source is where all flow starts and comes from, and the sink is where all the flow ends up. Another important detail is that flow follows the rules of conservation, where the amount of flow going into any given vertex must be equal to the amount coming out of it (unless they are the source or sink).



Figure 7. Flow Network example [14].

Shown above in Figure 7 is an example of a flow network. Every edge has a flow/capacity pair, with the flow ranging from 0 to the capacity of that edge. That flow only comes out of source *s*, and flows into sink *t*. Additionally, flow conservation can be examined here. As an example, Vertex A receives 8 flow (5 from *s* and 3 from D), and outputs 8 flow (5 to B and 3 to C). This rule is held up in the other vertices as well.

An st-cut is the division of a flow network into 2 subgraphs *S* and *T*, where *S* contains the source, and *T* contains the sink. This cut is made by removing edges such that there is no remaining connection between the two subgraphs, disjointing them. Cut capacity is the sum of capacities of the removed edges, but an important note is that only the capacity of edges that flow into *T* are counted in cut capacity. This can be examined in Figure 8. A cut of the network is being depicted such that the edges connecting 0 to 2, 2 to 1, and 1 to 3 are removed from the graph. This leaves two subgraphs, *S = {s, 0, 1}* and *T = {2, 3, t}*. We can calculate the cut capacity to be 5, since we ignore the capacity of edge 2 to 1 as it flows into *S*, leaving capacities of 2 and 3.

A screenshot of a computer

AI-generated content may be incorrect.

Figure 8. st-cut example.

The minimum st-cut is the st-cut where cut capacity is minimized. Importantly, it has been proven that the value of this cut is equal to the maximum flow through the network [15], and so the problems can be used interchangeably. Using the same graph, we can find that the minimum st-cut is pictured in Figure 9, where *S = {s, 0}* and *T = {1, 2, 3, t}.* The cut capacity comes out to be 4, as the edge from 1 to 0 is excluded due to flowing into *S*, leaving capacities of 2 and 2. If you examine the example graph further, you will find no st-cut with a cut capacity less than 4, which is why Figure 9 depicts the minimum st-cut.

A screenshot of a computer

AI-generated content may be incorrect.

Figure 9. Minimum st-cut example.

In the CoreExact algorithm, the Edmonds-Karp algorithm is used to find max flow, and therefore the minimum st-cut. Edmonds-Karp uses a breadth first search to find the shortest path from *s* to *t*, pass the max flow it can push along this path, update the flow network to include this flow, then repeat until no more paths can be found. This results in a network of maximum flow, and by taking the st-cut of this graph, *S* (which will contain all vertices with flow running through them) will be the minimum st-cut subgraph. The time complexity of the Edmonds-Karp algorithm is [16]. This is because breadth first searches have time complexity of , but in this case a large number of edges is the worst case, so we have worst case of for the breadth first search. But this must run for every path, which in the worst case is roughly equal to *ve* paths. So, the worst-case time complexity comes out to be .

## **An in Depth Look at Goldberg’s Algorithm**

Now that the necessary pieces have been explained, a more in-depth explanation of how Goldberg’s maximum flow algorithm works in CoreExact can be done.

To start off, lower and upper bounds *l* and *u* are declared, and α is set as the average of *l* and *u.* Then a specifically designed flow network that is based on any input graph is utilized. In this network, every edge (*a,b*) is replaced by a pair of directed edges, one from *a* to *b* and one from *b* to *a* each with capacity 1. Two additional vertices are added to the network to be the source and sink (*s* and *t*). There is an edge added for every vertex *a* in the graph from *s* to *a* where the capacity is equal to the motif degree of *a*. There is also an edge added for every vertex *a* in the graph from *a* to *t* where the capacity is equal to α times the motif size. This flow network structure can be seen in Figure 10 (the edges connecting vertices like *0*, *u*, and *v* together are implied). As a note, in Goldberg’s original algorithm the flow network uses edge degree and just α. But CoreExact has adjusted those values to work with motifs other than edges. The time complexity of this construction comes down to the handling of the motifs. Each edge of each motif needs to be handled separately to determine the weight. This is the most complex part of the construction, resulting in a time complexity of . With this graph, the minimum st-cut is taken to see if there exists a subgraph of density α or higher.

A diagram of a graph

AI-generated content may be incorrect.

Figure 10. CoreExact flow network.

Golberg explains the logistics of this minimum st-cut in this paper [17]. The explanation states for a minimum st-cut of the constructed network if *S / {s} = ∅,* then there is no subgraph of density greater than α. But if *S* is not empty, the density of *S / {s} >* α. Goldberg does this by showing the two cases based on his cut capacity equation, where . In this equation *z* is the capacity of edges from *s* to the graph’s vertices, *V* is the set of vertices in *G*, and is *S / {s}*. In these cases, contradictions would arise if was empty or not, proving his theorem. A more through explanation of this can be found in [17], and an explanation for how it’s applied to cliques and motifs is in [13].

This process iterates numerous times. With each iteration, the bounds of *u* and *l* tighten. When α produces a new densest subgraph *l* is set as α, but otherwise *u* is set as *α*. The final key detail here is how many iterations occur. The way it’s set up is that it iterates until . The logic behind this bounding is that is the smallest possible distance between any two subgraph densities for a given graph, meaning onceis greater than this value there are no more possible densities to find through the binary search. The proof for this value is that given two densities and , the difference between them is . If then the difference (since and are interchangeable in this case). But otherwise, the smallest possible difference between and is 1, meaning you have values *v* and *v -* 1 in which case the difference would be [17].

As previously discussed with Goldberg’s algorithm, the time complexity of an implementation of the algorithm was where *v* is the number of vertices and *T* is the complexity of the minimum st-cut. After analyzing the CoreExact implementation of Goldberg’s algorithm, this holds true. It is still a binary search depending on the number of vertices *v* (hence log(*v*) iterations), and for every iteration the most complex step is the minimum st-cut, which is the Edmonds-Karp algorithm. Meaning that the time complexity for the CoreExact implementation of Goldberg’s algorithm comes out to as it’s already been explained that the time complexity of Edmonds-Karp was .

## **The Full Algorithm**

Now that the individual aspects have been covered, let’s review how this algorithm works. Before the algorithm begins, a graph and motif type are selected as input. Edge-based core decomposition is run to generate a DAG of the graph, which is then used to generate a list of all motifs in the graph. These can then be used to run motif-based core decomposition on the graph to find the k-cores and store the information on the densest of them. The densest k-core is then broken into connected components, and the highest density is found between the k-core and its connected components. The densest of these is saved as the current densest subgraph, and the upper and lower bounds (*l* and *u*) are generated from them, where *u* = the *k* value of the k-core and *l* = the greatest density. At this point, the binary search for a denser subgraph begins, being run on each of the connected components. The next best guess for highest density is used to construct the flow network. The minimum st-cut of the flow network returns *S*. Then it is checked if *S = {s}* or not. If *S = {s}* is true, a subgraph of density greater than or equal to α was not found and the upper bound is now set as α. If *S = {s}* is not true, then a subgraph of density greater than or equal to α was found, so the density of that subgraph now equals the lower bound and that subgraph is saved as the current densest. This runs until the difference between the upper and lower bound *u* and *l* are such that . Once that happens, the densest subgraph for that component has been found. The lower bound is then set as the highest density of that component. If that density is greater than the density of our current densest subgraph overall, that new subgraph is set as the densest subgraph. If there is another connected component, a new binary search for the densest subgraph of that component begins with the new lower bound and same upper bound. Otherwise, it is done and the densest subgraph has been found.

A screenshot of a computer program

AI-generated content may be incorrect.

Figure 11. CoreExact algorithm [9].

For reference in Figure 11’s terminology, *p* represents density, *p’’* is the density of the densest k-core, *k’’* is the k value of the densest k-core, ψ is motif type, (*k’’*, ψ)-core is the densest k-core, *C*’s are components, and *F*’s are flow networks.

Our worst-case time complexity for CoreExact is relatively straight forward to find. The main loop is based on the number of connected components: *c*. And for every iteration the most complex step would be Goldberg’s algorithm. Every other necessary step (including processing our motifs to find the density) is negligible by comparison. So, since the time complexity of Goldberg’s algorithm comes out to , and it occurs *c* times, we the time complexity of CoreExact comes out to. With all of this in mind, parallelization of the CoreExact algorithm can be covered.

**Chapter 4**

# **Our Parallelized Implementation**

As previously, our solution was implemented and coded with CUDA. Because of the way CUDA works, much of the solution is still serialized and written in C++. So rather than cover all the CUDA code, we will only cover the parts that were parallelized. This was mostly done with kernel functions we wrote for the CUDA program, though a few were pre-existing parallel C++ functions from the Thrust library [18].

Additionally, while our CUDA solution is mostly an implementation of the CoreExact algorithm, there is one notable change. Rather than using any type of motif, our solution is limited to cliques. Cliques are a specific type of motif, being a subset of vertices that are all connected to one another. They’re often referred to as k-cliques, where the k is how many vertices are in the clique. So, a 3-clique is a clique of 3 vertices, and a 4-clique is a clique of 4 vertices.

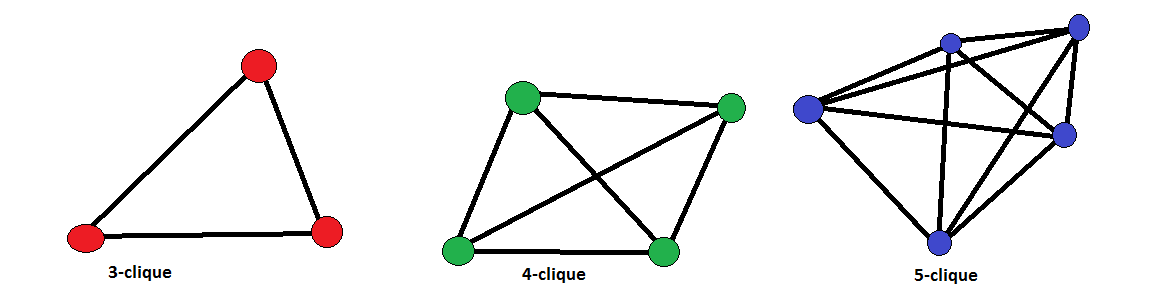


Figure 12. k-clique examples [19].

The CoreExact algorithm and the definition of motif density remain unchanged, with the only change being the process for finding all the cliques. This change was due to not being able to parallelize the listing of most motif types, which is already a process that takes a fair amount of time in the serialized program. Of note, this does not affect our solution’s ability to do edge density, as a 2-clique is an edge by definition (two vertices connected to all other vertices in the clique, which is just each other).

## **DAG Generation**

The first piece of the code to be parallelized is the generation of the DAG (of not, before this is run, edge-based core decomposition has already occurred). The first major step is generating new degrees for every vertex to reflect the changes of becoming a DAG (with edges only counting for a vertex’s degree if it has a higher core value than the vertex it’s connected to). This is done using the kernel function generateDegreeDAG, where every vertex is processed by a warp in parallel. The 32 threads of each warp work in parallel to compare the core value of the vertex to its neighbor, adding 1 to a collective degree variable for the warp (and thus the vertex) if the neighbor has a lower order. The structure of this parallelization is shown in Figure 13 visualizing how each warp handles a vertex, and each thread of the warp handles a neighbor of that vertex. This methodology stores the DAG degree since direction is now accounted for in the vertex’s degree. And once all neighbors are checked, the new degree of the vertex is stored.

A diagram of a machine

AI-generated content may be incorrect.

Figure 13. Processing vertex neighbors by thread.

The second major step is setting up the direction of every vertex to reflect the change into a DAG. This is done in a similar vein as the first step, using the kernel function generateNeighborDAG. This follows the structure seen in Figure 13, with every warp processing a vertex and each thread in the vertex checking a neighbor of the vertex (and thus the edge). In the case the neighbor has a lower order, the neighbor is stored for this vertex in a new adjacency list. This creates the DAG, storing only the edges in direction of higher order to lower order and with every thread checking a neighbor.

## **Finding all Cliques**

Just as the algorithm for finding cliques in CoreExact was in depth, so is the CUDA code for finding all cliques in the graph. The first major step is done with a kernel method called listInitialCliques. This function processes vertices by warp, using its threads to process the neighbors of the vertex. For each neighbor, the code checks if it’s a valid candidate for a clique, seeing if it has enough neighbors to be part of a clique with the current vertex. This information is stored, creating pairs of vertices that will be checked as being part of larger cliques. So, every thread is working in parallel to check for valid vertex and neighbor pairs that may be part of a larger clique. After this is the kernel method flushPartitions. This function simply updates global information based on current local information, including the current list of potential cliques and neighbors that could extend the clique.

The next major step is a loop that checks each clique level down to 2, where clique level refers to the size of the clique. Each iteration, the kernel function listMidCliques is called. This function checks all candidate vertices by warp. For each candidate, it checks whether the candidate has valid neighbors for the next clique level. If it does, all valid neighbors of the candidate are added to the list of candidates for the next clique level, and the candidate is officially added to the potential clique it may be a part of. So, each thread is checking for valid neighbors in parallel and each warp is processing candidates. After listMidCliques is finished, flushPartitions is run again to update the global clique information. Then it moves to the next iteration with a lower clique level. To summarize, this stage consists of confirming whether candidate vertices can be added to their potential cliques, adding new candidates, and then updating that information to be global each iteration through the descending clique levels until we get the clique level 2.

The final step uses the kernel method writeFinalCliques. By this point, the potential cliques are all of size *k – 1*, where *k* is the size of our k-clique. And we have a final list of candidate vertices. So, the potential cliques are all processed by warps with the threads of each warp checking the neighbors of each of its candidate vertices. For every valid neighbor (a neighbor that is connected to all other vertices in the clique), a complete clique is found. At that point the clique is stored (the vertices in the clique) and the clique degree of all the vertices in the clique are updated. Thus, every thread is checking for potential complete cliques in parallel. Afterwards, the list of cliques and the clique degrees of the vertices are stored globally.

## **Clique-Based Core Decomposition**

Next is what would be considered the first step of the CoreExact algorithm: core decomposition based on the chosen clique size. The parallelization here occurs inside a serial loop with each iteration checking a core level (a different *k* value). So, the first iteration checks the 0-core, the second iteration is the 1-core, and so on. This runs until all vertices have been peeled from the graph, and the densest k-core is found by storing data on the densest k-core each iteration, updating it when a denser k-core is found.

Each iteration, two main steps occur in parallel. The first is done with the kernel method selectNodes. Every vertex is processed on an individual thread, meaning as many vertices are being processed at the same time as there are threads. This parallel structure is visualized in Figure 14. Each thread checks whether the degree of that vertex is equal to the current level (our *k* value). If it is equal that means it is part of the current k-core, and therefore it is added to a buffer of vertices that need to be removed by the next level (we’ll call this the removal buffer).

A line with black text

AI-generated content may be incorrect.

Figure 14. Processing vertices by thread.

The second step is done with the kernel method processNodesByWarp. In this method, every vertex in the removal buffer is processed by warp and each thread of the warp processes a neighbor of the vertex. So, for every vertex in the removal buffer we check every clique that the vertex is a part of. For each neighbor in the clique, the clique degree of the neighbor is decreased by 1. In the case that this decrements that vertex to have a degree equal to our current level that vertex gets added to the removal buffer to be handled next iteration. In the case that the decrement brings that vertex’s degree below the current level, it gets incremented back up by 1 as it would already be in the buffer or it would have already been peeled. This makes it so that at the end, the array holding the degrees of every vertex will instead hold the core value of every vertex.

After these two parallelized processes have run, the core level is increased by 1 so the next k-core can be checked in the next iteration, and then density of our current k-core is checked. If the current k-core’s density is greater than the currently stored greatest k-core density, the densest k-core information is updated to be the current k-core.

## **Construction and Pruning of the Densest K-Core**

With the information about the densest k-core gathered, the next step is to construct the densest k-core subgraph so that it can be worked with rather than just knowing what vertices are in it. Therefore, a graph representative of the densest k-core must be generated and relevant information about the graph must be stored. This process is done in two major steps with two kernel functions.

First is generateDensestCore which handles the bulk of the construction. This function handles vertices by warp. For each warp, it’s checked if the vertex has a clique core value greater than or equal to that of the densest k-core, confirming whether the vertex is part of the densest k-core. If it is it is added to the new ongoing list of vertices in the subgraph. Once it’s added, all its neighbors are checked in parallel with each thread of the warp checking one neighbor at a time. For every neighbor that also has a clique core value greater than or equal to the densest k-core value, the clique degree of the vertex is increased by 1 (with it starting at 0 for this new subgraph). Meaning by the end of this parallelized function, all vertices in the densest k-core subgraph and all their clique degrees are set.

The second step of this process is done with the kernel function generateNeighborDensestCore. As with keeping the adjacency list structure for the graph, the other major piece of information that needs to be stored for the graph is what vertices each vertex is connected to in the densest k-core (this needs to exclude all neighbors outside of the densest k-core). Like the previous step, the vertices are handled by warp and each neighbor is being handled by the threads in these warps. For each thread, one of the warp vertex’s neighbors is checked to see if it is part of the densest k-core. If it is, it is added to the array of neighbors of the vertex.

And with these two steps, the list of vertices in the subgraph, the clique degree of each vertex, and all their neighbors in the subgraph are stored. Now we must prune our graph for unnecessary edges. That isn’t to say that there are edges in this subgraph that aren’t part of a clique. But as the program searches for the densest subgraph, there’s a chance certain edges aren’t part of *enough* cliques to be relevant. Therefore, it is efficient to find and eliminate all edges of low density. This will be done in two main steps, identifying the edges to remove and updating the clique degrees and vertex neighbors.

To identify edges to remove, this program will instead do the opposite and identify all edges that need to stay. So, in the kernel function pruneEdges, all the cliques that have been identified are handled by warp. For every clique, the warp checks if its clique value is greater than the densest k-core value (e.g. confirming a 5-clique is part of our 4-core but a 3-clique is not). For every clique that is part of our k-core, all of its edges are marked as necessary over the warp’s threads in parallel. Any edge that never gets marked is not part of a large enough clique is not necessary and will be removed. This parallelization is visualized in Figure 15.

A diagram of a diagram

AI-generated content may be incorrect.

Figure 15. Handling cliques by warp and its edges by thread.

Once our pruning identification is done, the pruning of unnecessary edges can be done by updating the clique degrees and neighbor lists of each vertex. This is handled by two kernel functions: generateDegreeAfterPrune and generateNeighborAfterPrune. These work in a very similar way. Both handle each vertex in a warp and check each neighbor by threads of the warp. For each neighbor, the edge’s removal status is checked. As covered, every marked edge needs to be removed. In generateDegreeAfterPrune, it keeps a count of every neighbor that isn’t marked for removal. Then once every neighbor is processed, the clique degree of the warp vertex is set to be that count as it is all the unpruned neighbors. In generateNeighborAfterPrune, a list of every neighbor that isn’t marked for removal is kept. Then once every neighbor is processed, the neighbor list of the vertex is set as this new list. Meaning that these two functions process the neighbors of every vertex in parallel, keeping track of what neighbors aren’t being removed to update the degree and neighbor list of each vertex. And once these two functions are run, the densest k-core subgraph with all of its relevant information stored.

## **Component Decomposition**

The next parallelized section is the component decomposition of the densest k-core. To start this off, component ids for every vertex are stored with every id simply starting as the vertex id in an array, so vertex numbers align with the index numbers of the array. Using these ids, the components are narrowed down by finding the minimum component id that every vertex is connected to.

This is done in a main do while loop that runs until no component id changes occur, kept track of using a boolean value named “changed”. For each iteration, the kernel function componentDecomposek is run. In this function, every vertex is handled by warp. For each warp, the current component id of the vertex is stored and set it as the minimum component id, then it starts checking the vertex’s neighbors over each thread in the warp. For each neighbor, if the neighbor’s component id is less than the current minimum id, then the neighbor’s component id is set as the minimum. After every neighbor is checked, if a minimum id was found that was less than the current vertex’s, the vertex’s new component id is set as the minimum and “changed” is updated to be true. So, in parallel the program checks to see if each vertex has a neighbor of lower component id, updating the vertices to be in the lowest component id they are connected to by any path in the graph until no more updates to component id occur.

Then, utilizing the C++ parallel functions sequence and sort\_by\_key [18], which run on the GPU, the connected components are stored. The sequence function simply creates an array of numbers from two given numbers as the range (e.g. given 3 and 7 will create an array storing the numbers 3, 4, 5, 6, and 7). In this case, it is used to create an array of all the vertices in the densest k-core. Sort\_by\_key is given 2 arrays, a list of keys (in this case the component ids) and values (the vertices). This function sorts the keys and matches them with the values. So, all the component ids are sorted then the vertices they match to are sorted to match the new order of the component ids. This way, the components are stored as a pair of linked arrays.

Additionally, two final pieces of info are stored using two more C++ parallel functions. First, by running the function unique\_copy [18] on our list of component ids, every unique component id is found and that amount is stored, effectively counting how many connected components there are. Then, it uses the function lower\_bound [18] is used to create an array of every index in the array of component ids that represents the start of a component. Of note, the parallel C++ functions utilized for creating our linked arrays and storing the connected component information are from the “thrust” library. And with that, the connected components are found and can be worked with individually.

## **Bound Generation**

For Goldberg’s algorithm an upper and lower bound for the density being checked is needed so that they can be tightened in a binary search. CoreExact used the same upper bound each time and updated the lower bound after each iteration. However, to parallelize the process our solution can’t rely on previous iterations for lower bounds and instead will generate an upper and lower bound for each connected component.

The upper and lower bounds will be stored in a pair of arrays, with their indices linking them in representation of the component. Next, like in CoreExact the maximum k-core value is stored to serve as a strict upper bound for all of the components. Then the kernel function getLbUbandSize is run. Each component is handled on a thread. This parallelization is depicted in Figure 16. For each component, the lower bound is set as the clique density of that component. Then, for upper bound, the maximum possible number of k-cliques for the component is found (a theoretical fully connected graph) and then divided by the number of vertices in the component. Whichever value is lower between the fully connected component density and the maximum k-core value is stored as the upper bound. Finally, if the lower bound is greater than the upper bound the component is discarded entirely, not storing it in an ongoing array. Otherwise, it is stored in this array. Thus, the upper and lower bounds for each connected component are calculated in parallel so each component is processed in parallel rather than sequentially.

A diagram of a line with letters and numbers

AI-generated content may be incorrect.

Figure 16. Handling components by warp.

## **Flow Network Creation**

With bounds set, the unique flow networks necessary for finding the densest subgraph can be generated for each component. As a reminder, the structure of these flow networks is a source and sink node connecting to every vertex in the graph, with edges in the graph all having capacities of 1, edges from the source to each vertex with a capacity of the vertex’s motif degree (which will now be clique degree), and edges from the vertices to the sink with a capacity of α times motif size (where α is the average of the upper and lower bound and motif size is now clique size).

This will be parallelized across the connected components. The first step is allocating memory for the edges of our flow networks. Using the kernel function createFlowNetworkOffset, the components are handled by warp. For each component, memory offsets are created according to the number of edges that need to be stored. This is done by handling the vertices in parallel, with each vertex running on a thread and setting the number of edges for that vertex as the clique degree of the vertex. This parallel structure is depicted in Figure 17. Then the source and the sink have edges set up as the number of vertices in the graph since both need to connect to every vertex.

A diagram of a diagram

AI-generated content may be incorrect.

Figure 17. Handling components by warp and their vertices by thread.

And now with offsets created and memory allocated, flow network information can be generated. Like the previous function, this is done with each component being handled by warp and each vertex being handled by thread as seen in Figure 17. The flow network is handled in a structure with two linked arrays, with one storing the edges and the other storing the capacity of the edge it’s linked to. First, every edge from a component vertex to the sink is set, with their capacity being equal to α times the clique size. Then every edge from the source to a vertex is stored, with their capacity being equal to the clique degree of that vertex. So, every edge of the flow network and its capacity is set in parallel across all the threads of their component’s warp.

## **In Summary**

Of the CoreExact algorithm, our solution has parallelized DAG generation, clique listing, clique-based k-core decomposition, densest k-core construction and pruning, component decomposition, bound generation, and flow network creation. Not every part of CoreExact was parallelized into our program. That was due to time constraints and certain parts not being able to run in parallel. For example, running a minimum st-cut algorithm or the bound tightening loop of Goldberg’s algorithm in parallel is not possible due to the consecutive dependencies of the process. While a process such as updating the flow networks based on new upper and lower bounds is potentially possible but will have to be pursued in an updated version of this code due to time constraints. However, as it stands our solution parallelizes most of the code, leaving only the main loop with sections to be parallelized. And it is this version of the code that will be tested.

In comparison to the serialized CoreExact program, this parallelized solution should in theory run larger graphs faster due to handling the many vertices and cliques in parallel. It should also handle graphs with connected components even better considering Goldberg’s exact algorithm will run in parallel across the components rather than one at a time. So, while it should in theory be quicker overall, the less connected a graph is the more efficient the parallel solution should be in comparison to the serialized code. That being said, the results should illuminate the strengths and potential weaknesses of the parallelized solution.

**Chapter 5**

# **Experiment and Evaluation**

To test the CUDA program, a series of graphs have been run through both the original CoreExact algorithm in Java and the CUDA program. This was done with 8 different graphs, 3 of which are real data and 5 of which are randomly generated.

The five real graphs include a network of the proteins in yeast (where the vertices are proteins and edges are chemical bonds [20]), a social network of the European users of the music streaming service Deezer (where vertices are the users and edges represent friends on the service [21]), and a network representation of co-authors on scientific papers in the condense matter field (where vertices are authors and edges represent whether a pair has co-authored a paper before [22]) [23] [24]. The synthetic graphs were created with a Python program which takes a number of vertices and a decimal chance for any two vertices to be connected as input. This program creates a 2-dimensional array as an empty adjacency list, then checks every possible vertex pair. For each vertex pair it randomly decides if the two are connected based on the input decimal chance, and if they are connected both are added to each other’s list of connected vertices in the adjacency list. For our randomly generated synthetic graphs 1 through 5, the chances of a connection between any vertex pair were 5%, 10%, 5%, 1%, and 0.1% respectively.

**Table 1**

*Graphs by Number of Edges and Vertices*

|  |  |  |
| --- | --- | --- |
| Graph | Number of Vertices | Number of Edges |
| Yeast Protein Network | 1,459 | 1,948 |
| European Deezer Social Network | 28,281 | 92,752 |
| Condense Matter Co-Author Network | 108,300 | 186,936 |
| Synthetic Graph 1 | 500 | 6,276 |
| Synthetic Graph 2 | 500 | 12,372 |
| Synthetic Graph 3 | 1,000 | 25,053 |
| Synthetic Graph 4 | 2,500 | 31,256 |
| Synthetic Graph 5 | 5,000 | 12,385 |

Shown above in Table 1 is the number of vertices and edges of each graph used in these experiments. These graphs give a large range of sizes and overall graph density.

The results of our experiments are focused on the results of the programs and their efficiency. Thus, Table 2 illustrates the highest density found for each graph in regard to edge density and 3-clique density. And to measure and compare the efficiency of the two programs the run times are recorded side by side. So, Table 3 illustrates the run times for both CoreExact and our parallel solution checking for the highest edge density, while Table 4 illustrates the run times for highest 3-clique density. Tables 3 and 4 also display how much faster the parallel solution processed each graph in comparison to the CoreExact program.

**Table 2**

*Highest Density by Graph (Rounded to 2 Decimal Places)*

|  |  |  |
| --- | --- | --- |
| Graph | Highest Edge Density  (edges / vertices) | Highest 3-Clique Density  (3-cliques / vertices) |
| Yeast Protein Network | 2.26 | 3.71 |
| European Deezer Social Network | 8.53 | 33.41 |
| Condense Matter Co-Author Network | 13.37 | 109.63 |
| Synthetic Graph 1 | 12.55 | 5.23 |
| Synthetic Graph 2 | 24.74 | 40.67 |
| Synthetic Graph 3 | 25.05 | 20.80 |
| Synthetic Graph 4 | 12.51 | 1.21 |
| Synthetic Graph 5 | 2.62 | 0.33 |

**Table 3**

*Graphs by DSD Runtime Over Edge Density*

|  |  |  |  |
| --- | --- | --- | --- |
| Graph | CoreExact Runtime (milliseconds) | Parallelized Solution Runtime (milliseconds) | Speed Increase  (CoreExact Runtime/ Parallel Runtime) |
| Yeast Protein Network | 403 | 153.368 | 2.63 |
| European Deezer Social Network | 23545 | 9893.91 | 2.38 |
| Condense Matter Co-Author Network | 76683 | 1558.99 | 49.19 |
| Synthetic Graph 1 | 16875 | 443.164 | 38.08 |
| Synthetic Graph 2 | 59243 | 3045.69 | 19.45 |
| Synthetic Graph 3 | 228629 | 1344.22 | 170.08 |
| Synthetic Graph 4 | 1021662 | 4312.55 | 236.9 |
| Synthetic Graph 5 | 278561 | 9916.88 | 28.09 |

**Table 4**

*Graphs by DSD Runtime Over 3-Clique Density*

|  |  |  |  |
| --- | --- | --- | --- |
| Graph | CoreExact Runtime (milliseconds) | Parallelized Solution Runtime (milliseconds) | Speed Increase  (CoreExact Runtime/ Parallel Runtime) |
| Yeast Protein Network | 114 | 131.444 | 0.87 |
| European Deezer Social Network | 6346 | 18177.1 | 0.35 |
| Condense Matter Co-Author Network | 108521 | 3347.62 | 32.42 |
| Synthetic Graph 1 | 9365 | 317.121 | 29.53 |
| Synthetic Graph 2 | 1187069 | 433087 | 2.74 |
| Synthetic Graph 3 | 1566331 | 53334.22 | 29.37 |
| Synthetic Graph 4 | 12231 | 7693.51 | 1.59 |
| Synthetic Graph 5 | 291 | 254.143 | 1.15 |

Based on these results, there is a notable range in how effectively the process was sped up. When it came to edge density, every graph saw an increase in efficiency, with only a couple graphs seeing less than at least a ten-fold speed increase. But when it came to 3-clique density, there were a couple graphs that were slower overall, a few that were only a bit faster, and a few that saw a thirty-fold speed increase. While the results are not fully consistent based on the differences between the input graphs, there is a definite increase in speed overall, and we can conclude that the parallel solution is more efficient than its serialized counterpart.

**Chapter 6**

# **Conclusion**

The Densest Subgraph Discovery Problem is a fundamental problem in graph mining with many existing solutions. In this paper we have created a parallel solution that runs using CUDA based on the CoreExact algorithm to capitalize on the processing power of the GPU. Based on our results, the parallel solution was indeed faster and overall more efficient than CoreExact, which itself is more efficient than many other DSD solutions. This reaffirms the computing power available through use of the GPU and CUDA and the belief that utilizing these tools for other graph mining problems would likely serve as a significant source for creating faster and more efficient solutions.

In the future, our solution can be improved even further. Certain sections of the CoreExact algorithm were not parallelized due to time constraints or not being able to run them in parallel. However, with more time these sections could be improved as well to create an even more efficient solution to the DSD. And with further analysis of our results, the weaknesses of our solution can be identified and updated.

Overall, our parallel CUDA program serves as a more efficient solution to the DSD is a basis that can be further improved or implemented in the future and acts as an example of the computing power available through parallel programming on the GPU.

# **References**

|  |  |
| --- | --- |
| [1] | C. M. Y. F. a. L. V. L. W. Luo, "A Survey of Densest Subgraph Discovery on Large Graphs," 14 June 2023. [Online]. Available: https://arxiv.org/abs/2306.07927. [Accessed 6 January 2025]. |
| [2] | R. K. a. V. Vinay, *Analyzing the Strucutre of Large Graphs,* August, 1999. |
| [3] | Udacity, "Intro to Parallel Programming," 5 August 2013. [Online]. Available: https://www.youtube.com/watch?v=F620ommtjqk&list=PLAwxTw4SYaPnFKojVQrmyOGFCqHTxfdv2. [Accessed 6 January 2025]. |
| [4] | A. K. E. S. M. H. A. S.-R. Lizhi Xiang, "cuTS: Scaling Subgraph Isomorphism on Distributed Multi-GPU," Boise State University, Boise, 2021. |
| [5] | U. Perego, "ResearchGate," July 2015. [Online]. Available: https://www.researchgate.net/figure/Schematic-illustration-of-GPU-architecture-and-programming-elements\_fig4\_274838864. [Accessed 15 May 2025]. |
| [6] | A. M. A. F. F. B. T. Lanciano, "A Survey of Densest Subgraph Discovery Problem and Its Variants," 18 April 2024, Apr. 18. [Online]. Available: https://arxiv.org/abs/2303.14467. [Accessed 8 January 2025]. |
| [7] | M. Charikar, "Greedy Approximation Algorithms For Finding Dense Components in a Graph," *Proceedings of the 3rd International Workshop on Approximation Algorithms for Combinatorial Optimization,* pp. 84-95, 2000. |
| [8] | G. K. a. D. P. U. Feige, "The Dense k-Subgraph Problem," in *Algorithmica*, New York, Springer-Verlag, 2001, pp. 410-421. |
| [9] | K. Y. R. C. L. V. L. X. L. Y. Fang, "Efficient Algorithms for Densest Subgraph Discovery," 17 August 2019. [Online]. Available: https://arxiv.org/abs/1906.00341. [Accessed 6 January 2025]. |
| [10] | F. B. A. G. F. G. a. C. Tsourakakis, "Denser than the densest subgraph: extracting," *KDD,* p. 104–112, 2013. |
| [11] | G. Shute, "University of Minnesota Duluth," Fall 2014. [Online]. Available: https://www.d.umn.edu/~gshute/cs2511/slides/graphs/slide017.html. [Accessed 2 February 2025]. |
| [12] | R. U. Orakzai, "What is the K-Core of a Graph," Baeldung, 5 June 2023. [Online]. Available: https://www.baeldung.com/cs/graph-k-core. [Accessed 5 February 2025]. |
| [13] | J. P. R. P. C. T. S. C. X. M. Mitzenmacher, "Scalable Large Near-Clique Detection in Large-Scale Networks via Sampling," *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining,* pp. 815 - 824, 2015. |
| [14] | J. Kogler, "Maximum Flow - Ford-Fulkerson and Edsmond-Karp," Algorithms for Competitive Programming, 13 October 2024. [Online]. Available: https://cp-algorithms.com/graph/edmonds\_karp.html. [Accessed 10 February 2025]. |
| [15] | L. Mouatadid, "Network Flows: The Max Flow/Min Cut Theorem," University of Toronto, Toronto, 2016. |
| [16] | "DSA Edmonds-Karp Algorithm," W3 Schools, [Online]. Available: https://www.w3schools.com/dsa/dsa\_algo\_graphs\_edmondskarp.php#:~:text=So%2C%20BFS%20with%20time%20complexity,(%20V%20%E2%8B%85%20E%202%20)%20.. [Accessed 11 March 2025]. |
| [17] | A. V. Goldberg, Finding a Maximum Density Subgraph, Berkley: Univerisity of California, 1984. |
| [18] | "CUDA Core Compute Libraries: Thrust," NVIDIA, 2025. [Online]. Available: https://nvidia.github.io/cccl/thrust/index.html. [Accessed 13 June 2025]. |
| [19] | "Graph kclique\_percolation," 8 January 2016. [Online]. Available: https://pythonhosted.org/trustedanalytics/python\_api/graphs/graph-/kclique\_percolation.html. [Accessed 14 April 2025]. |
| [20] | H. a. M. S. a. B. A. a. O. Z. Jeong, "Lethality and centrality in protein networks," *arXiv preprint cond-mat/0105306,* 2001. |
| [21] | B. R. a. R. Sarkar, "Characteristic Functions on Graphs: Birds of a Feather, from Statistical Descriptors to Parametric Models," in *Proceedings of the 29th ACM International Conference on Information and Knowledge Management (CIKM '20)*, 2020. |
| [22] | J. K. a. C. F. J. Leskovec, "Graph Evolution: Densification and Shrinking Diameters," in *ACM Transactions on Knowledge Discovery from Data (ACM TKDD), 1(1)*, 2007. |
| [23] | J. Leskovec, "Stanford Large Network Dataset Collection," Stanford University, [Online]. Available: https://snap.stanford.edu/data/. [Accessed 22 May 2025]. |
| [24] | R. A. R. a. N. K. Ahmed, "The Network Data Repository with Interactive Graph Analytics and Visualization," AAAI, [Online]. Available: https://networkrepository.com/index.php. [Accessed 22 May 2025]. |
| [25] | P. S. Segundo, "ResearchGate," June 2010. [Online]. Available: https://www.researchgate.net/figure/A-A-simple-undirected-graph-U-B-The-adjacency-matrix-A-of-U-C-Bit-wise-AND-operation\_fig6\_220204855. [Accessed 2 February 2025]. |
| [26] | M. Aibin, "Connected Components in a Graph," Baeldung, 18 March 2024. [Online]. Available: https://www.baeldung.com/cs/graph-connected-components. [Accessed 5 February 2025]. |
| [27] | S. Y. Cheung, "The Maximum flow and the Minimum cut," Emory University, 27 October 2019. [Online]. Available: https://www.cs.emory.edu/~cheung/Courses/253/Syllabus/NetFlow/max-flow-min-cut.html. [Accessed 10 February 2025]. |
| [28] | "Introduction to Push Relabel Algorithm," GeeksForGeeks, 6 June 2023. [Online]. Available: https://www.geeksforgeeks.org/introduction-to-push-relabel-algorithm/. [Accessed 4 May 2025]. |
| [29] | P. Sah, "Inferring social structure and its drivers from refuge use in the desert tortoise, a relatively solitary species.," *Behavioral Ecology and Sociobiology 70.8,* pp. 1277-1289, 2016. |
| [30] | D. J. a. S. S. H. Watts, "Collective dynamics of small-world networks," *nature,* vol. 393, no. 6684, pp. 440-442, 1998. |