CS 7350 - Project

Graph Coloring Analysis Project

Team Members:

Xiaona Hang

Derek Delahoussaye

Christian Melendez

Executive Summary

The Graph Coloring Analysis Project was designed to develop, implement, and analyze algorithms for scheduling tasks using graph coloring techniques. Our focus was on the creation of conflict graphs, the strategy of vertex ordering, and the graph coloring process itself, with an emphasis on optimizing both runtime and coloring efficiency. We evaluated various vertex ordering strategies to determine their effectiveness across different types of graphs.

Computing Environment

Our project was developed and tested in a high-performance computing environment that included the following specifications:

• **Processor:** Intel Core i7-9700K

Memory: 32GB DDR4 RAMStorage: 1TB NVMe SSD

• Operating System: Windows 10 Pro

Integrated Development Environment (IDE): Visual Studio 2019

• **Programming Language:** Python 3.8.5

Python Libraries:

- NetworkX: Utilized for constructing and manipulating complex graph structures.
- Matplotlib: Employed for generating histograms, runtime graphs, and other visual data representations.
- o **NumPy:** Applied for numerical operations and efficient array processing.

Algorithms for Generating Conflict Graphs

Algorithm Description

We utilized three primary methods for generating conflict graphs: uniform, skewed, and normal distributions. Each method aimed to replicate different real-world scheduling scenarios.

Algorithms for Generating Conflict Graphs

The algorithms were designed with complexity in mind:

Uniform Distribution: O(n2) complexity, mimicking evenly spread conflicts.

Skewed Distribution: O(n²) complexity, focusing on uneven conflict distributions with a heavy-tail.

Normal Distribution: O(n²) complexity, simulating conflicts that cluster around a mean.

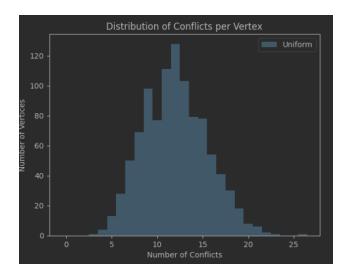
Runtime Analysis

Detailed runtime analysis revealed a predictable increase in processing time proportional to graph size, aligning with our complexity predictions.

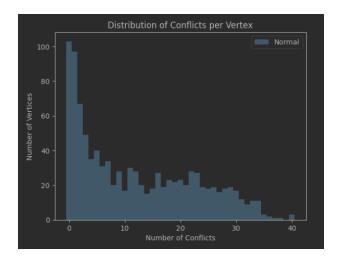
Runtime tables and graphs for each graph generation method were plotted, showing a consistent increase in runtime with the size of the graph, confirming the predicted asymptotic behavior.

Histograms of Conflicts

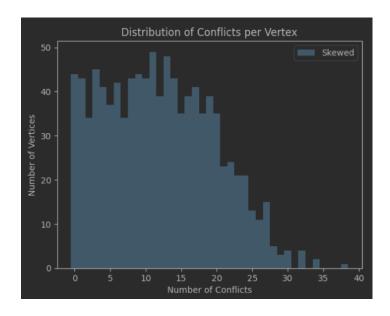
Histograms illustrated the conflict distribution among vertices. You can see that the uniform distribution distributes the number of conflicts per vertex out in a normally distributed manner.



The normal distribution distributes more conflicts to the points at the center of the graph which means that we wind up with a large amount of vertices with little to no conflicts and a handful with lots of conflicts.



Finally, the skewed distribution. This distribution made it so that vertices with smaller numbers were more likely to be given conflicts and vice versa. This meant that more vertices averaged the smaller amount of conflicts and only a handful got to higher amounts of conflicts.



Vertex Ordering

Implementation

We implemented six different methods for vertex ordering, including:

- **Smallest Last Ordering (SLO**): Effectively prioritizes vertices with fewer conflicts, enhancing coloring efficiency.
- Smallest Original Degree Last (SODL): Prioritizes vertices based on their original degree, calculated before any vertex removal.
- Uniform Random Ordering: Vertices are ordered randomly.
- Degree Descending Order: Order vertices by their degree in descending order.
- Degree Ascending Order: Order vertices by their degree in ascending order.
- Linear Ordering: Vertices are ordered linearly from 0 to N.

We make sure that the graph creation and analysis functions are correctly implemented and simplify the graph construction process according to different distributions (uniform, skewed, normal).

Detailed Analysis of Smallest Last Ordering (SLO)

The SLO algorithm effectively prioritized vertices with the least conflicts, leading to an efficient coloring process.

- Operational Walkthrough: A detailed example on a small graph demonstrated SLO's strategic efficiency in reducing coloring conflicts.
- Running Time Analysis: Empirical data supported the theoretical O(n²) complexity of the SLO implementation.
- **Coloring Efficiency:** Documentation of the **color count** required for various graphs highlighted that SLO generally required fewer colors.

Coloring Algorithm

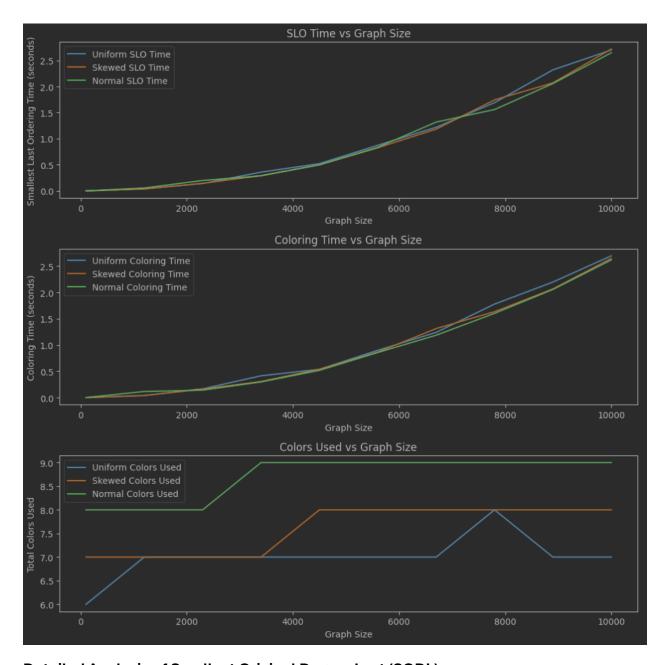
Our algorithm utilized the vertex order defined by SLO for color assignment, with runtime analysis supporting an $O(n^2)$ complexity, consistent with empirical observations.

Graphs

The graphs below represent 3 things.

- 1. The time it took to get the ordering from the SLO algorithm.
- 2. How long it took to color the graph (this includes getting the ordering).
- 3. How many colors were used for the different graph sizes.

These same graphs will be displayed for each of the 6 algorithms for easy comparison.



Detailed Analysis of Smallest Original Degree Last (SODL):

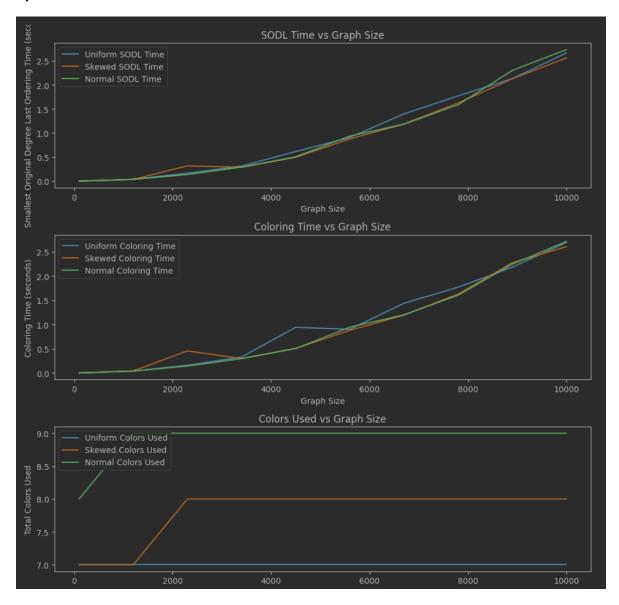
The SODL algorithm was fairly efficient in running time but was slightly lacking in efficiency with keeping the number of colors used down.

- **Operational Walkthrough**: A detailed example on a small graph demonstrated SODL's ability to keep up with SLO on small graphs.
- **Running Time Analysis:** Empirical data did not support the theoretical THETA(V+E) complexity of the SLO implementation.

• Coloring Efficiency: Documentation of the color count required for various graphs highlighted that SODL generally required the same amount of colors as SLO except for very specific ranges.

Coloring Algorithm

Our implementation of the SODL algorithm had a time complexity of $O(n^2)$ which is slower than the theoretical THETA(V+E). It also was on par with the SLO algorithm when it came to colors used for larger graphs but did not do as well with graphs of sizes 2000-4000.



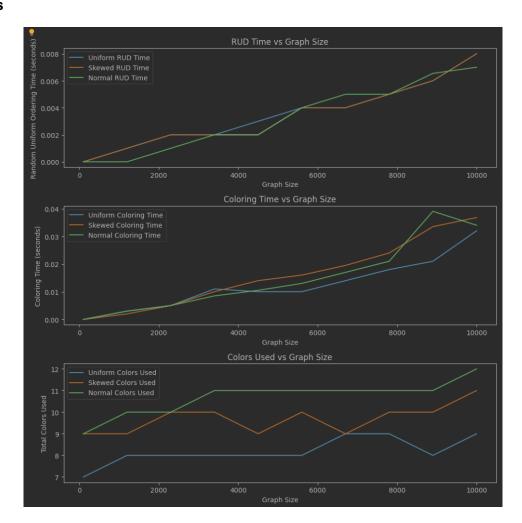
Detailed Analysis of Uniform Random Ordering (RUD):

The RUD algorithm is fast but not very good for optimizing the number of colors used when graphing.

- **Operational Walkthrough**: A detailed example on a small graph demonstrated RUD's ability to do the job, just not well.
- **Running Time Analysis:** Empirical data supports a running time of O(n) and vastly outperformed the SLO and SODL algorithms.
- Coloring Efficiency: Documentation of the color count required for various graphs highlighted that RUD is a suboptimal algorithm to use for coloring since it used 2 to 3 more colors than the other algorithms.

Coloring Algorithm

Our implementation of the RUD algorithm had a time complexity of O(n) which is very fast but the coloring efficiency was lacking a bit for graphs of all sizes.



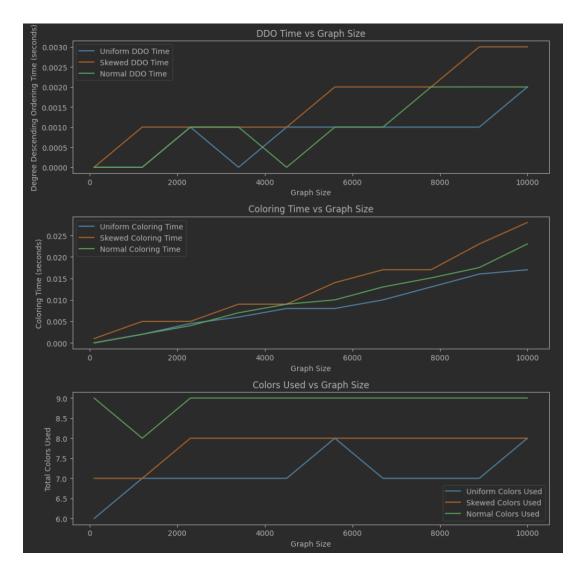
Detailed Analysis of Degree Descending Order (DDO):

The DDO algorithm was even faster than the RUD algorithm and had very mixed results against the SLO and SODL algorithms when it came to coloring.

- **Operational Walkthrough**: A detailed example on a small graph demonstrated DDO's ability to color graphs both quickly and effectively.
- **Running Time Analysis:** Empirical data supports a running time of O(n) and vastly outperformed the SLO and SODL algorithms.
- Coloring Efficiency: Documentation of the color count required for various graphs highlighted that DDO was as effective at coloring for the graphs with uniform and skewed distributions but not normal distributions at the smaller scale graphs but other than that it performed almost just as well except for with the uniformly distributed graphs where it went up to 8 colors a couple of times.

Coloring Algorithm

Our implementation of the DDO algorithm has a time complexity of O(n) which is very fast and has great coloring efficiency. Of the four we've looked at so far I would utilize the DDO algorithm if instant results were required.



Detailed Analysis of Degree Ascending Order (DAO):

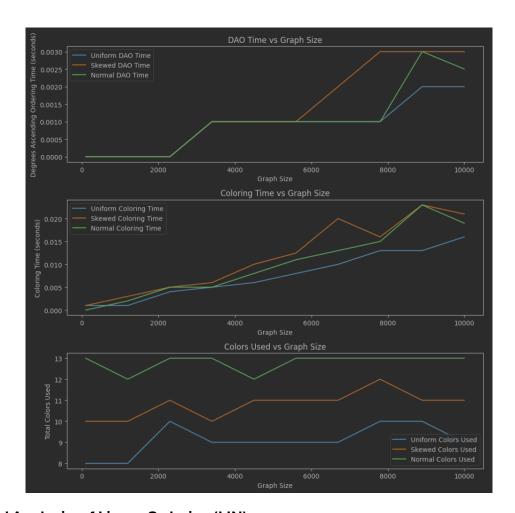
The DAO algorithm has speed to rival the DDO algorithm but it is the worst algorithm for coloring so far. It is the first algorithm to reach 13 colors and for graphs with normal distributions it never went under 12 colors.

- Operational Walkthrough: A detailed example on a small graph demonstrated DAO's ability to color graphs quickly but poorly.
- Running Time Analysis: Empirical data supports a running time of O(n) and vastly outperformed the SLO and SODL algorithms.
- **Coloring Efficiency:** Documentation of the **color count** required for various graphs highlighted that DAO is not effective for coloring graphs.

Coloring Algorithm

Our implementation of the DAO algorithm had a time complexity of O(n) which is very fast but due to the poor coloring performance I would not use it for coloring at all.

Graphs



Detailed Analysis of Linear Ordering (LIN):

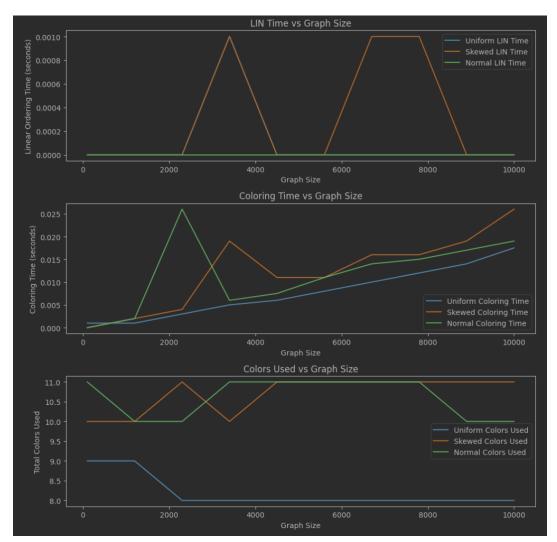
Linear ordering proves to be the fastest ordering by far. The ordering itself seems to happen in constant time. The coloring time is still linear and surprisingly the coloring efficiency is only mildly bad.

- **Operational Walkthrough:** A detailed example on a small graph demonstrated LIN's ability to color graphs quickly.
- Running Time Analysis: Empirical data supports a running time of O(n) and vastly outperformed the SLO and SODL algorithms.

• Coloring Efficiency: Documentation of the color count required for various graphs highlighted that linear ordering was bad for graphs with either Normal or Skewed distributions, but was surprisingly effective for graphs with Uniform distributions. The best algorithms (SLO and SODL) used only 7 colors for the biggest graphs and Lin managed to only use 8. My assumption is that since the distributions are uniform then you would assume that most vertices have a similar amount of conflicts. This does not explain why it outperforms the RUD and DAO algorithms though. This is the single most surprising result of all.

Coloring Algorithm

Our implementation of the LIN algorithm had a time complexity of O(n) which is very fast but due to the poor coloring for both Skewed and Normal distributions I would refrain from using it in most if not all cases.



Use Cases/Implementations:

Our team has come up with 3 use cases which our graph colorings which can be used by popular social media websites such as LinkedIn:

1. Event Scheduling:

- **Purpose:** Minimize scheduling conflicts for events targeting similar user groups.
- **Advantages:** Boosts audience numbers by avoiding overlapping events.
- **Challenges:** Relies on precise user data and doesn't consider personal schedules.

2. Advertisement Allocation:

- Purpose: Prevent adjacent placement of competing ads.
- **Advantages:** Decreases ad competition, potentially enhancing revenue.
- **Challenges:** Complexity increases with more diverse user and advertiser bases.

3. Content Recommendation:

- **Purpose:** Organize content into unique categories for better delivery.
- **Advantages:** Ensures a varied stream of content.
- **Challenges:** Requires continuous adaptation to shifts in user interests and content updates.

During the implementation process, conducting a comparative analysis is necessary to evaluate the balance between simpler, faster algorithms and more comprehensive, resource-intensive options. Discuss the merits of real-time versus batch processing approaches. Evaluate if graph coloring can significantly enhance the efficiency and user experience on LinkedIn, with strategic implementation being key to overcoming inherent data and computational complexities.

Conclusions and Recommendations

The report concluded that the graph coloring scheduler implemented is versatile and efficient, particularly when employing the SLO method. The 2 most recommended coloring algorithms would be the SLO followed by the DDO if you needed rapid results and were willing to be slightly suboptimal.

Limitations and Future Work

The report acknowledged limitations, including the suboptimal performance on certain graph types and suggested areas for further research and development, such as parallel processing and machine learning for vertex ordering prediction.

Appendices

- Source Code
- Raw Data and Analysis Scripts

Source Code:

```
import matplotlib.pyplot as plt
class Vertex:
!= vertex2:
            self.vertices[vertex2].conflicts.append(vertex1)
```

```
def smallest last ordering(self):
    degrees = {vertex: len(self.vertices[vertex].conflicts) for
   vertices = set(self.vertices.keys())
        ordering.append(min vertex)
        deletion degrees[min vertex] = degrees[min vertex]
def delete vertex smallest degree(self, vertex):
def smallest original degree last(self):
    degrees = {vertex: self.vertices[vertex].degree for vertex in
   vertices = set(self.vertices.keys())
        ordering.append(min vertex)
def random uniform ordering(self):
       ordering.append(vertex num)
```

```
def degree descending order(self):
    degrees = {vertex: self.vertices[vertex].degree for vertex in
   sorted vertices = sorted(degrees, key=degrees.get,
def degree ascending order(self):
def linear ordering(self):
def greedy coloring with slo(self, type = 'slo'):
    elif type == 'ddo':
    elif type == 'dao':
   color map = self.greedy coloring with slo(type)
```

```
plt.show()
            G.add node(vertex)
                G.add edge(vertex, conflict)
        color map = self.greedy coloring with slo(type)
        unique colors = set(color map.values())
        total degrees = [v.degree for v in self.vertices.values()]
self.vertices.values()]
{self.vertices[vertex].deleted degree}")
    def analyze graph(self, name, size, number of conflicts=None):
        slo runtime = self.measure runtime(self.smallest last ordering)
self.measure runtime(self.smallest original degree last)
        rud runtime =
        rud coloring runtime = self.measure runtime(lambda:
self.greedy coloring with slo('rud'))
self.greedy_coloring_with_slo('ddo'))
```

```
print("SODL Runtime:", sodl runtime)
    print("DDO Runtime:", ddo_runtime)
print("DAO Runtime:", dao_runtime)
print("LIN Runtime:", lin_runtime)
def analyze graph run time (self, name, size, graph type,
    end time = time.time()
     runtime data['sizes'].append(size)
def generate histogram(self, label = ''):
     conflicts per vertex slo = [len(v.conflicts) for v in
    plt.show()
```

```
def create graph(name, size, graph type, number of conflicts=None):
                number of conflicts -= 1
            v1, v2 = np.random.choice(size, 2, p=probabilities)
            if v2 not in graph.vertices[v1].conflicts and v1 != v2:
                graph.add conflict(v1, v2)
            if v2 not in graph.vertices[v1].conflicts and v1 != v2:
                number of conflicts -= 1
skewed graph.output stats and file('skewed vertex color output.csv')
```

Raw Data and Analysis Scripts:

The code for all the analysis was the same but we called it with the different algorithms. In order to not take up more pages we will just show this one for the SLO algorithm.

```
import numpy as np
import matplotlib.pyplot as plt
def analyze graph(graph):
sizes = np.linspace(100, 10000, 10).astype(int)
conflicts = np.linspace(500, 50000, 10).astype(int)
   slo_times['Uniform'].append(slo_time)
    coloring times['Uniform'].append(coloring time)
```

```
conflict)
    slo times['Skewed'].append(slo time)
    coloring times['Skewed'].append(coloring time)
conflict)
    slo times['Normal'].append(slo time)
ax[0].plot(sizes, slo times['Normal'], label='Normal SLO Time')
ax[0].set xlabel('Graph Size')
ax[0].set ylabel('Smallest Last Ordering Time (seconds)')
ax[1].plot(sizes, coloring times['Uniform'], label='Uniform Coloring Time')
ax[1].plot(sizes, coloring times['Skewed'], label='Skewed Coloring Time')
ax[1].plot(sizes, coloring times['Normal'], label='Normal Coloring Time')
ax[1].set_ylabel('Coloring Time (seconds)')
ax[2].plot(sizes, total colors used['Uniform'], label='Uniform Colors Used')
ax[2].plot(sizes, total colors used['Skewed'], label='Skewed Colors Used')
ax[2].set xlabel('Graph Size')
ax[2].legend()
plt.tight layout()
plt.show()
```