Linear Time Backbone Determination in a Wireless Sensor Network

Jake Carlson

April 21, 2018

Abstract

A report on implementing algorithms to partition a random geometric graph into bipartite subgraphs. Three different graph geometries are explored: unit square, unit disk, and unit sphere. Nodes are uniformly distributed in the geometry. Then the edges are determined and the verticies are colored using smallest-last vertex ordering and greedy graph coloring. Once coloring has been used to determine the independent color sets, the combinations of the largest are processed to find the largest backbone. All algorithms used in this report are implemented to run in linear time.

Contents

1 Executive Summary **3**
1.1 Introduction . . . 3
1.2 Environment Description . . . 3

2 Reduction to Practice **4**
2.1 Data Structure Design . . . 4
2.2 Algorithm Descriptions . . . 5
2.2.1 Node Placement . . . 5
2.2.2 Edge Determination . . . 5
2.2.3 Graph Coloring . . . 6
2.2.4 Backbone Determination . . . 7
2.3 Algorithm Engineering . . . 8
2.3.1 Node Placement . . . 8
2.3.2 Edge Determination . . . 9
2.3.3 Graph Coloring . . . 9
2.3.4 Backbone Determination . . . 10
2.4 Verification . . . 10
2.4.1 Node Placement . . . 10
2.4.2 Edge Determination . . . 10
2.4.3 Graph Coloring . . . 13

3 Appendix A - Figures **15**

4 Appendix B - Code Listings **28**

Listings

1 Processing driver . . . 28
2 Topology class and subclasses . . . 30

1 Executive Summary

1.1 Introduction

Random geometric graphs (RGGs) are useful for simulating wireless sensor networks placed in different topologies. This project examines three different geometries: Square, Disk, and Sphere. The user supplies parameters for how many nodes they want in the network and how many connections they want for each node. Then, the simulation finds the average radius needed for that number of connections, determines the edges in the graph, colors the graph to find independent sets, pairs the four largest independent sets to find the largest bipartite subgraphs, and cleans these bipartites to find the major component, or backbone, of each bipartite. The cleaning ensures that there are no singular points of failure that could cause the network to become disconnected. In other words, each backbone exists so that there are multiple paths between any two nodes in the backbone.

This creates network backbones from the random geometric graphs that are highly reliable and allow the largest number of wireless sensors to connect to it in only one hop. Additionally, the linear time implementation of this simulation ensures efficient running time regardless of the input size. The organization of the code base also makes it easy to implement new topologies by subclassing the main Topology class that implements all of the algorithms needed to determine the backbone.

All of the code used for this project, including the graphical display of the generated graphs at each stage in the backbone determination process, can be found here:

`https://github.com/jakecarlson1/sensor-network`

1.2 Environment Description

The data structures and topologies for this simulation are implemented in Python2.7. The graphics are generated using Processing.py [3]. All development and benchmarking has been done on a 2014 MackBook Pro with a 3 GHz Intel Core i7 processor and 16 GB of DDR3 RAM running macOS High Sierra 10.13.3.

Processing offers an easy to use API for drawing and rendering shapes two- and three-dimensions. The Processing.py implementation allows the entire use of the Python programming languages and libraries.

A separate data generation script was used to generate the summary tables (Tables 1, 2, 3). Because these benchmarks were run in a separate script, the timing does not measure the time required to draw the graphs using Processing. The figures were genetared using the matplotlib library [4]. This library, and a variety of others, could not be imported into Processing.py because the jython interpreter used by Processing only accepts libraries written in raw Python.

The different geometries were implemented in a stand alone Python file and imported into the Processing.py script or the data generation script, depending on what was being run. These classes can then be used directly by Processing or the data generation script. Because there is no intermediary file to hold the generated nodes and edges, there is no additional disk space needed to run the simulation. Everything can be done in system memory managed by Processing.

Benchmark	Order	A	Topology	r	Size	Realized A	Max Deg	Min Deg	Run Time (s)
1	1000	32	Square	0.101	14865	29	48	4	0.094
2	8000	64	Square	0.050	245108	61	93	17	1.255
3	16000	32	Square	0.025	250658	31	58	6	1.593
4	64000	64	Square	0.018	2019116	63	98	16	11.124
5	64000	128	Square	0.025	4010430	125	182	35	18.915
6	128000	64	Square	0.013	4051390	63	98	17	21.506
7	128000	128	Square	0.018	8075034	126	175	29	38.348
8	8000	64	Disk	0.045	248036	62	93	16	1.209
9	64000	64	Disk	0.016	2023518	63	104	15	10.547
10	64000	128	Disk	0.022	4015227	125	173	35	18.752
11	16000	64	Sphere	0.126	511920	63	91	35	19.625
12	32000	128	Sphere	0.126	2049089	128	177	84	79.037
13	64000	128	Sphere	0.089	4094059	127	173	87	148.707

Table 1: Benchmarks for generating RGGs. A: input average degree, r: node connection radius

Benchmark	Max Deg Deleted	Color Sets	Largest Color Set	Terminal Clique Size
1	22	22	76	21
2	40	37	320	35
3	25	24	1138	22
4	40	39	2530	38
5	73	64	1373	60
6	40	39	5044	36
7	74	68	2739	62
8	39	37	321	31
9	43	40	2538	36
10	72	64	1371	57
11	39	38	631	36
12	87	66	674	61
13	89	66	1351	61

Table 2: Benchmarks for coloring RGGs

Benchmark	B1 Order	B1 Size	B1 Domination	B1 Faces	B2 Order	B2 Size	B2 Domination	B2 Faces
1	114	296	0.924	-	120	290	0.961	-
2	546	1490	0.955875	-	558	1472	0.97175	-
3	1779	4458	0.9163125	-	1726	4286	0.8928125	-
4	4471	11894	0.977484375	-	4450	11854	0.9753125	-
5	2559	7170	0.991296875	-	2546	7122	0.9895	-
6	9106	24284	0.9815078125	-	8954	23816	0.9807265625	-
7	5169	14476	0.993421875	-	5186	14444	0.9950546875	-
8	572	1504	0.984	-	558	1500	0.980375	-
9	4544	12124	0.9830625	-	4522	12000	0.982625	-
10	2587	7280	0.99525	-	2599	7272	0.993234375	-
11	1176	3166	0.992875	1992	1166	3128	0.9918125	1964
12	1293	3616	0.99875	2325	1284	3596	0.99728125	2314
13	2613	7390	0.99765625	4779	2603	7260	0.997703125	4659

Table 3: Benchmarks for backbone determination

2 Reduction to Practice

2.1 Data Structure Design

The primary data structure used for this project is an adjacency list. However, to allow for constant time lookup of edges of a node, a Python dictionary is used where the keys are nodes and the values are a list of indicies of adjacent nodes in the original list of nodes. The space needed by the adjacency list is $\Theta(|V| + 2|E|)$. Two entries are used for each edge because they are undirected. This is superior to the adjacency matrix data structure which would require $\Theta(|E|^2)$ space.

In order to make this project maintainable as it is developed along the semester, the object-oriented capabilities of Python are used to design the different geometries. First, a Topology class is defined that creates the interface Processing uses to draw the graphs. This base class implements all of the methods needed for node placement and edge detection in 2D graphs. Then, three subclasses are created: Square, Disk, and Sphere.

The Square and Disk topologies simply need to override the methods for generating nodes and calculating the node radius needed for the desired average degree. The Sphere subclass needs to override a few additional functions because it exists in a 3D space. Other than the methods for generating nodes and calculating the node radius, it also needs to override the function used to draw the graph so that Processing will render the graph properly in 3D.

2.2 Algorithm Descriptions

2.2.1 Node Placement

A different node placement algorithm is required for each of the geometries. For the Square, the coordinates for each node are generated as two random numbers taken from a unifrom distribution on the range $[0, 1]$. All of these points are guaranteed to be in the unit square.

For the Disk, a similar method is used. The coordinates for nodes are randomly sampled from a uniform distribution; however, if a node has a distance from the center of the Disk greater than the radius of 1, the coordinates for that node are resampled.

For the Sphere a different method must be used so that all of the nodes are placed on the surface of the Sphere and the volume is vacant. For this geometry, the following equations are used:

$$x = \sqrt{1-u^2}\cos\theta \tag{1}$$
$$y = \sqrt{1-u^2}\sin\theta \tag{2}$$
$$z = u \tag{3}$$

where $\theta \in [0, 2\pi]$ and $u \in [-1, 1]$. This is guaranteed to uniformly distribute nodes on the surface area of the sphere [5].

All of these algorithms can be solved in $\Theta(|V|)$ where because each node only needs to be assigned a position once.

2.2.2 Edge Determination

To calculate the node connection radius needed to acheive the desired average connection, the ratio of node coverage to the total area can be used. This ratio must equal the ratio of the total number of nodes to the average degree, or:

$$\frac{A_{geometry}}{A_{node}} = \frac{Num\,Nodes}{Avg\,Deg} \tag{4}$$

Applying this to each geometry only requires filling in the equation for the area of the geometry and the connection area. This is straight forward for the square and disk. The geometry areas are given by $R^2 = 1$ and $\pi R^2 = \pi$ respectively since these are the unit square and circle. The sphere is slightly more complicated. Since nodes should only be able to connect over the surface of the sphere (following arcs), the connection area is to be taken as the surface area of the spherical cap such that the arc of the cap is twice the length of the connection distance. In other words, a node placed on the surface of the sphere in the center of a spherical cap can connect to any other node that falls in that spherical cap. The equation for the area of the spherical cap is given by

$$S_{cap} = \pi(a^2 + h^2) \tag{5}$$

where a is the distance from the midpoint of the base of the cap to the edge of the base, and h is the distance from the midpoint of the base to the top of the cap (where the node would be) [6]. If we connect these points with a third variable, x, such that x is the actual distance from the node to the edge of its connection area, the Pythagorean theorem can be used to substitute in x^2 for $a^2 + h^2$. The equation for the node connection radius of the unit sphere then looks identical to that of the unit circle. The final list of equations used to calculate node connection radius for a desired average degree are given in Table 4.

Geometry	Geometry Area	Node Area	r
Square	1	πr^2	$r = \sqrt{\frac{Average\ Deg}{\pi \times Num\ Nodes}}$
Disk	π	πr^2	$r = \sqrt{\frac{Average\ Deg}{Num\ Nodes}}$
Sphere	4π	πr^2	$r = 2 \times \sqrt{\frac{Average\ Deg}{Num\ Nodes}}$

Table 4: Equations for node conneciton radius

There are several methods for finding the edges in the graph. The brute force method checks every node, and for each node checks all other nodes to see if they are close enough to form an edge. The brute force method is $\Theta\left(|V|^2\right)$.

The second method to find the edges is the sweep method. This method first sorts the nodes along the x-axis. Then, for any node, we only need to search left and right until the distance along the x-axis is greater than the connection radius for the nodes. This dramatically reduces the search space. The sweep method is $O\left(nlg(n) + 2rn^2\right)$ where $n = |V|$ an r is the connection radius. The $nlg(n)$ portion is for the sorting and the $2rn^2$ portion is for measuring the distance between nodes in a sweep step.

The final method to find edges is the cell method. This method places the nodes into cells of area $r \times r$ based on their position in the topology. When the edge detection runs, each node needs to be visited once, but only the cell the node populates and the neighboring cells need to be searched for connections.

The only method that needs to be adjusted for the Sphere is the cell method. Instead of using a two dimensional grid of cells, a three dimentional mesh is needed to divide the topology. The cells then have volume $r \times r \times r$. Only the current cell and the neighboring cells need to be searched.

2.2.3 Graph Coloring

Two algorithms are used for coloring the graphs. The first is smallest-last vertex ordering, which sorts the verticies based on the number of degrees they have. The second is the greedy graph coloring algorithm.

Smallest-last vertex ordering is used to order the nodes for coloring. The steps to this algorithm are as follows [1]:

1. Initialize a representation of your target graph
2. Find the vertex v_j of minimum degree in your representation
3. Update your representation to simulate deleting v_j
4. If there are still verticies in the representation, return to step 1, otherwise terminate with the sequence of verticies removed

This algorithm is linear if each of the above steps is linear. Step 1 is linear if we can build a represenation of the graph in linear time. For this, we can use an array of buckets, where each bucket holds the verticies that have the same number of edges as the position of the bucket in the array of buckets. To build this data structure, each node only needs to be visited once, making this linear in both space and time. Next, finding the vertex of minimum degree simply requires finding the lowest index bucket that has a node. This is bounded by the number of buckets, which is bounded by the number of nodes, making Step 2 linear. Next, we have to update the representation of the graph. To do this, we have to look at each node that shares an edge with v_j and move it to the bucket for nodes with one fewer degree. This requires traversing the list of edges for v_j which means Step 3 is linear. Since this is repeated for each node, the runtime of this program is $\Theta\left(|E| + |V|\right)$ and the space needed is $\Theta(|V|)$.

After this, a single traversal of the smallest-last vertex ordering is needed to color the graph. As we traverse this list, we check to see if the nodes before it (that are already colored) share an edge with the current node. The node can then be colored with any color it does not share an edge with or, if it shares an edge with all currently used colors, it is assigned a new color. This algorithm is also linear. Each node needs to be visited once and when a node is visited, all previous nodes are checked to see if they are in the edge list of the current node. Because we used smallest last vertex ordering, as we have to check more and more nodes, we get to check fewer and fewer edges. This makes the greedy coloring algorithm $O(|V| + |E|)$.

2.2.4 Backbone Determination

Several algorithms are needed for determining the most suitable backbones for the wireless sensor network. First, the four largest independent sets are paired with each other to generate the largest bipartite subgraphs for the random geometric graph. These bipartites are bound to have minor components that are not connected to the major component, and blocks that are only connected by bridges. These nodes need to be removed in order for the backbone to be considered reliable. Once all of these nodes have been removed from the bipartite, the backbone has been determined. Then, the two backbones with the largest size are selected and their domination (ratio of nodes connected to the backbone) and number of faces (for the sphere topology) are calculated.

The largest independent sets are the largest color sets given by smallest-last vertex coloring. These will be the first four color sets when greedy coloring is used on a sequence of nodes sorted in smallest-last order. The combination of these four independent sets must be taken to find the six largest bipartite subgraphs.

The bipartite subgraphs need to be cleaned up in order to measure the size and coverage area of the backbone. This can be done by first removing all of the tails in the graph, which are sequences of nodes coming off of a component where the end node has degree one, and all nodes in between have degree two. Then, the major component needs to be determined, which is the component with the largest order. Once the largest component is determined, the minor blocks and the bridges connecting them to the major component need to be removed. A bridge is simiar to a tail; it is a chain of edges that, if removed from the graph would increase the number of connected components. These features need to be removed because they do not provide reliability to the wireless sensor network. If a single one of these node were to fail, a portion of the graph would become disconnected from the remaining backbone. This creates a single point of failure that should not occur in a network backbone.

Each of these algorithms can be implemented in linear time. Taking the combinations of the four largest independent color sets can be done by building a bipartite subgraph for each combination where the nodes are copied from the two color sets that make up the bipartite. Each bipartite will then be built in $\Theta(2|V|)$ time and $\Theta(2|V|)$ space where $|V|$ is the number of nodes in each color set. Since there are six ways to choose two items from a set of 4, this runs six times, resulting in $\Theta(12|V|)$ space and time usage for building all of the bipartites.

The tails then need to be removed. This can be done by repeatedly removing all nodes with a degree of one. This will repeatedly remove the last node in the tail until the only remaining node is the node that connected the tail to its component. This will also remove any minor components that consist of a thin chain of nodes with no cycles. This is similar to smallest-last vertex ordering, except the deletion of nodes from the graph stops when there are no more nodes in the bucket for degree one. Since this algorithm is based off of smallest-last vertex ordering, and slvo ran in $\Theta\left(|E| + |V|\right)$, this is bounded above by smallest-last vertex ordering, $O\left(|E| + |V|\right)$. However, since the bipartite could have no tails in it, the lower bound of the runtime is $\Omega(|V|)$ which is the amount of time needed to place nodes in their respective buckets based on how many edges they have in the bipartite. Regardless, this will require $\Theta(|V|)$ space to create a representation of the bipartite that can be deleted from.

Next, the major component needs to be determined. This can be done with breadth-first seach. BFS will traverse the entire graph, counting the number of nodes that can be reached from some start node. If an entire component has been explored from some start node, and there are still unvisited nodes in the graph, BFS will pick a new start node and begin searching from there. By counting the number of nodes connected to each start node, the size of each component can be determined. The major component can be determined by taking the max of these sizes. BFS works with a queue of nodes to search. At the start of an iteration, the current node is removed from the front of the queue, and all of its neighbors are added to the queue, if they have not already been visited. Since each node is only visited once, the runtime for BFS is $\Theta(|V| + |E|)$. BFS operates in-place on the graph, but a parallel array to the array of nodes is needed to remember if a node has been visited or not. This requires $\Theta(|V|)$ space and time to initialize. All together, this algorithm runs $Theta(2|V| + |E|)$ time.

Next, the bridges need to be removed from the major components. This can be done by modifying depth-first search to check for back-edges to nodes. If some node and its edges are being searched, it is a bridge if and only if none of the decendents of the nodes connected to the current node have a back-edge to the current node or any of its ancestors. Back-edges can be checked by maintaining a list of visit times given by the DFS algorithm (tin), and a list of the minimum entry time of any ancestor (fup). If the current node's neighbors have ancestors with earlier entry times, then they must have a back-edge to that node. If they have a back-edge with the current node, the minimum entry time of the ancestors would be the current time. If the miniumum entry time of the neghbor's ancestors is greater than the current time, it must be a bridge. This is codified in the folowing formula [8]:

$$fup[v] = min \begin{cases} tin[v] \\ tin[p] \text{ for all } p \text{ for which } (v,\, p) \text{ is a back edge} \\ fup[to] \text{ for all } to \text{ for which } (v,\, to) \text{ is a tree edge} \end{cases} \tag{6}$$

Given this formula, the current edge $(v,\, to)$ is a bridge if and only if $fup[to] > tin[v]$ in the DFS tree. DFS runs in $\Theta(|V| + |E|)$ and the book-keeping data structures add a total space requirement of $\Theta(2|V|)$.

Once the bridges have been found, the graph needs to be simulated to have them removed, and the resulting connected components need to be searched again for the major component. BFS can be used again, where if an edge is encountered that is in the set of bridge edges, the neighbors to the current node are not pushed into the queue. Using BFS again has a time and space requirements $Theta(2|V| + |E|)$ time and $\Theta(|V|)$ space.

With the bridges removed, the major component in each graph has been determined and all single points of failure that could result in the disconnection of backbone nodes have been removed. It is then time to determine the two largest backbones for further evaluation. The size of the backbones (the number of edges) can be detemined in linear time by traversing all of the nodes in the backbone and counting the edges that are shared with other nodes in the backbone. This runs in-place on the backbone representation in $\Theta(|V| + |E|)$ time for each backbone that needs to have its size calculated.

The domination of the two largest backbones needs to be calculated. Finding the number of nodes connected directly to the backbone is equivalent to finding the number of nodes that are not connected to the backbone. This can be done by traversing all nodes that are not part of the backbone and, for each of their edges, seeing if the adjacent node is a backbone node. This algorithm requires $\Theta(|V|)$ space and $\Theta(|V| + |E|)$ time to run where $|V|$ is the number of nodes not in the backbone.

Finally, if the topology is a sphere, the number of faces can be determined by using Euler's Polyhedral Formula [7], which is given by:

$$2 = V + F - E \tag{7}$$

$$F = 2 - V + E \tag{8}$$

Where V is the number of verticies, E is the number of edges, and F is the number of faces.

2.3 Algorithm Engineering

2.3.1 Node Placement

It is easy to implement the algorithms for placing nodes in the different geometries using Python's math library. This library offers functions for sampling points on a uniform distribution. For the Square, sampling on a range $[0, 1]$ is sufficient for all of the nodes. Since each node only needs to be placed once, this runs at $\Theta(|V|)$ where.

For the Disk, the node needs to be resampled if it is too far from the center. To do this, the distance function is used to find the distance between the node and the center. If the node is further than 1 from the center, node generation falls into a while loop which iterates until the node is within the unit circle. Since nodes are taken from a uniform distribution, the number of nodes that will need to be resampled is approximately equal to the ratio of the area of the square that circumscribes the unit circle which falls outside of the unit circle to the total area of the square. This is given by:

$$\frac{(2r)^2 - \pi r^2}{(2r)^2} = \frac{4 - \pi}{4} = 0.2146 \tag{9}$$

Since the placement algorithm for each node of the Disk will iterate until the node falls within the unit circle, the total number of iterations N can be found as the sum of the geometric series:

$$N = \sum_{k=0}^{\infty} n(0.2146)^k = \frac{n}{1 - 0.2146} = 1.273n \tag{10}$$

where $n = |V|$. This shows this implementation is $\Theta\,(n)$.

For the node placement algorithm of the Sphere, again the math library in Python makes this easy. Each node needs two random values pulled from a uniform distribution, two square root operations, one sine operation, and one cosine operation. Each node only needs to be placed once so the runtime of this algorithm is $\Theta(n)$ where $n = |V|$.

2.3.2 Edge Determination

Each method implemented for finding edges has a different time complexity. The brute force method uses an outer loop and an inner loop, which each iterate over every node in the graph. An edge is saved to the adjacency list if the nodes are not the same and the distance between them is less than or equal to the calculated node radius. This is guaranteed to run in $\Theta\left(n^2\right)$ where $n = |V|$. The number of times the distance needs to be calculated is $n \times (n-1)$ because it will not be calculated when the nodes are the same (distance would be zero, but no edge is drawn here). No additional space is needed for the brute force method so the space complexity is $O(1)$.

The implementation of sweep starts by sorting the nodes along the x-axis. Python lists have a built-in sort function that has $O\left(nlg(n)\right)$ time complexity [9]. After this stage, it iterates over every node building a search space which will be scaned for edges. For each node, the list of nodes is searched right $r \times n$ nodes to find those within one radius length of the current node. With the search space built, the search space is iterated over once to find nodes that have a distance less than or equal the node radius. Then, the indicies of the nodes are added to the adjacency list entry for each other. My implementation of this runs in $O\left(nlg(n) + 2rn\right)$ where $n = |V|$ and r is the node connection radius. Because the list sort method sorts inplace, the only additional space needed is for the search space. This saves $O(rn)$ nodes and is reset after every iteration.

The cell method implementation works in linear time. In the first step of the method, the cells are initialized as a list of empty lists. There are $(1/r + 1)^2$ cells. The nodes are then iterated over and assigned a cell by dividing their x and y coordinates by the node radius. At this point, the cells are iterated over and, for each node in the cell, the nodes in the current cell and the four forward adjacent cells and the are checked to see if they fall within the node radius of the current node. All together, this implementation runs at $O\left(n + n + 5nr^2\right) = O\left((2 + 5r^2)n\right)$ where $n = |V|$. The amount of additional space needed is equal to the number of nodes because they are coppied into their respective cells. This places the space complexity at $\Theta(n)$.

2.3.3 Graph Coloring

Implementing the smallest-last coloring algorithm involves implementing the smallest-last vertex ordering algorithm and the greedy graph coloring algorithm. For smallest-last vertex ordering, the first thing to do is build the data structure used to represent the graph with deleted nodes. This can be done with a list of sets, where each the index in the list represents the degree of the nodes in that set. The number of sets needed is equal to the maximum degree of the nodes. The index of each node is placed in the set corresponding to the number of edges it has then the RGG. Simultaneously, a dictionary is created that maps each node to the number of degrees it has in the graph with deletions. Each value starts at the number of edges the corresponding node has in the RGG. At this point, we have iterated over all of the nodes once and allocated space for twice the number of nodes by copying them into the sets and using them as the keys for the degrees dictionary.

Because Python dictionaries resize at specific numbers of entries, we can determine the number of additional insertions caused by rehashing while the degrees dictionary is built. Python dictionaries start out with space for 8 entries and quadruple in size until the number of entries is above 50,000, at which point it begins to double in size. Clearly the dictionary grows at a logarithmic rate, but the total number of insertions I for an input size of n is given by:

$$I = \begin{cases} n + 8\sum_{k=1}^{\log_4 \lceil n/8 \rceil} 4^k & n \leq 50,000 \\ n + 8\sum_{k=1}^{6} 4^k + 32768\sum_{k=1}^{\log_2 \lceil n/32768 \rceil} 2^k & n > 50,000 \end{cases} \tag{11}$$

Fortunately, because the entire dictionary is built before it is used by the smallest-last vertex ordering algorithm, it will never again be resized once the algorithm starts. Unfortunately, the sets resize at a similar rate and it is more difficult to predict how large the sets will need to be when performing smallest-last vertex ordering. The degree dictionary will also be used to index into the sets, so we gain a speed up here by not having to iterate over all of the edges for a node and determining if the node it shares an edge with are in the remaining graph each time we want to sift nodes down to lower set.

After setting up the graph representation, the smallest-last vertex ordering algorithm runs until every node has been removed from the representation. To delete a node, the first non-empty set is selected. This set must contain the next node to remove becuase it contains all nodes with smallest degree. Before deleting the node from the graph, and moving all adjacent nodes down a set, the current set is checked to see if it has all remaining nodes. If this is the case, the terminal clique has been found, and the size of the terminal clique must be saved. After this check, a node is popped from the end of the current set, and appended to the smallest-last ordering result. Then, all nodes adjacent to the popped node in the original graph are checked to see if they are in the set with its current degree. If it is, the number of degrees for that node can be decremented and the node can be placed into the correct set for its new degree.

The last step is to reverse the order of the smallest-last ordering result because it was built in the opposite order (smallest-first). All together, excluding the initialization of accessory data structures, this implementation runs in $\Theta(2|V|+2|E|)$ time and $\Theta(2|V|)$ space since nodes are removed from the buckets and added to the result.

After this the graph needs to be colored. For this, initially each node is assigned a color of -1 in a node color array that is parallel to the original list of nodes. Then, all of the nodes in the smallest-last vertex ordering are iterated over. At each node, a set of colors that is already used by the neighbors of that node is created by iterating over all of its edge nodes and grabbing their color from the node color array. Then, color just has to be incremented from 0 until it does not exist in the search space set and the color has been determined to assign to the node.

Since the smallest-last odering is used, each time the edges need to be traversed to see if a node is adjacent to the current node, nodes with fewer and fewer edges are being searched. This means that the nodes with the most neighbors are searched first, when the number of other nodes to check is lowest, and the nodes with the fewest neighbors are searched last, when we have the most nodes to check if they share an edge with the current node. All together, this implementation runs in $\Theta(|V| + 2|E|)$ time and $\Theta(|V|)$ space because we need a new array for the colors assigned to each of the nodes.

A setp-by-step walkthough of the smallest-last coloring algorithm is provided to further visualize this algorithm. For this walkthrough, a unit square topology is used with 20 nodes and a node connection radius of 0.4. The smallest-last vertex ordering deletion process is shown in Figure 1. The coloring phase is shown in Figure 2. In the deletion process, the minimum degree node is removed at each step. If there are multiple nodes with the same minimum degree, one is choosen randomly. Once all nodes have been removed, the smallest-last vertex ordering has been detemrined. In the coloring phase, the node that was removed last is assigned a color first. As the smallest-last vertex ordering is traversed, each node's neighbors are checked to see if they have been assigned a color. The first color that has not been used by a neighbor is assigned to the node. To complete this walkthrough, the distribution of the color set sizes and the degrees of nodes when deleted is given in Figure 3.

2.3.4 Backbone Determination

2.4 Verification

2.4.1 Node Placement

The nodes can be verified to be distributed uniformly if the degrees follow a normal distribution. To show that the distribution of degrees for each of the geometries are following a normal distribution, the degree histograms are plotted for each of the benchmarks. The histrograms for Square are given in Figure 5, Disk are given in Figure 6, and Sphere are given in Figure 7. These histograms clearly follow a normal distribution, so the nodes must be placed uniformly.

2.4.2 Edge Determination

The runtime for the edge detection methods can be verified by varying the number of nodes and measuring the runtime of each algorithm. By looking at how the runtime grows, we can calculate the trendline that best fits the growth rate. For the first comparison, the number of nodes is varied from 4,000 to 64,000 in steps of 4,000, while holding the desired average dgree constant at 16. As we can see in Figure 4, the growth rates of the brute force and sweep methods are quadratic, while the growth rate of the cell method. The trendline functions are given on the graph.

For the second metric, the number of nodes is held constant at 32,000 and varied the desired average degree from 2 to 32 in steps of 2. The graph is given in Figure 4. The cell method clearly grows linearly, but the sweep method is harder to gauge. Since varying the desired average degree should only change

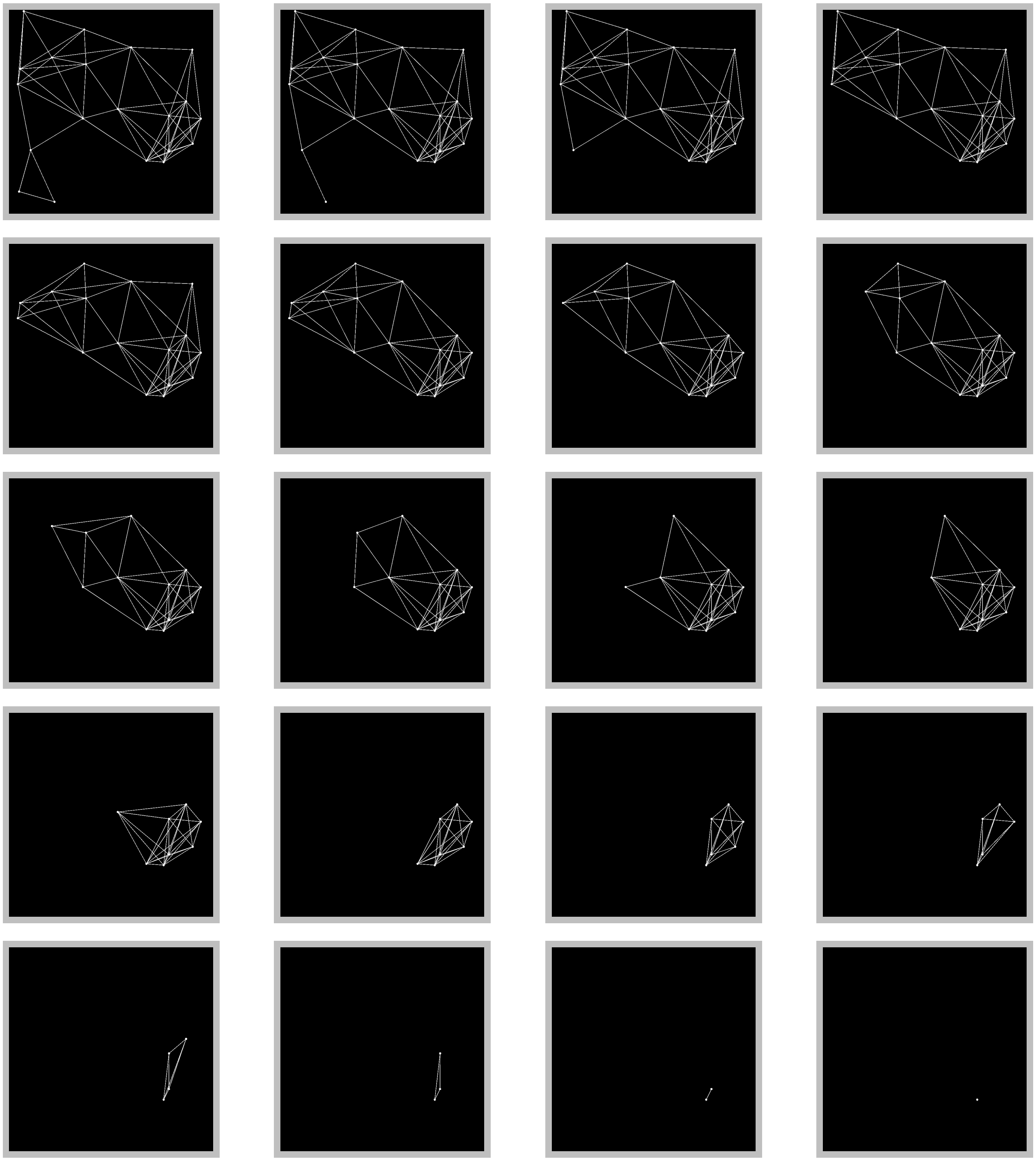


Figure 1: Smallest-last vertex ordering deletion process

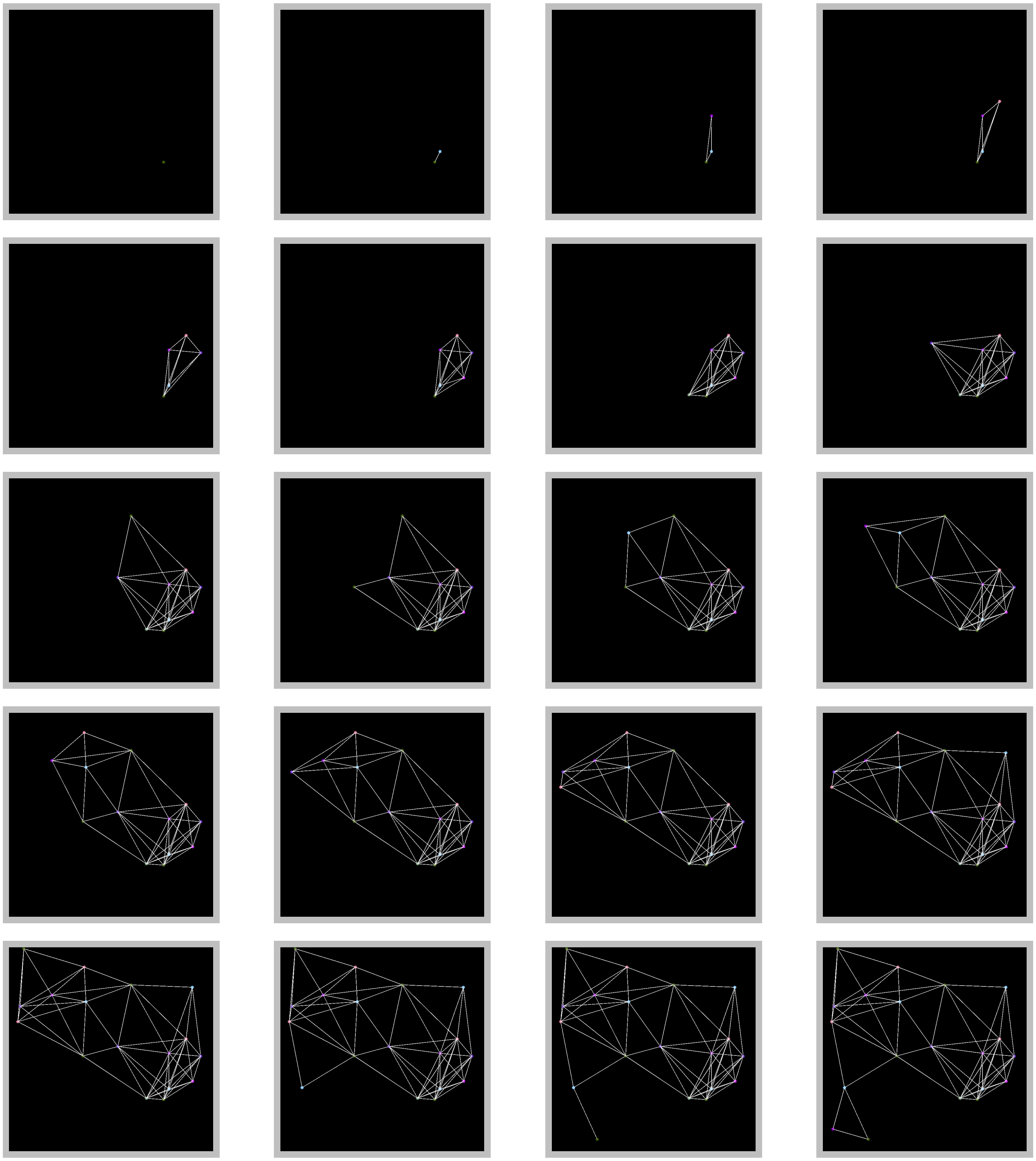
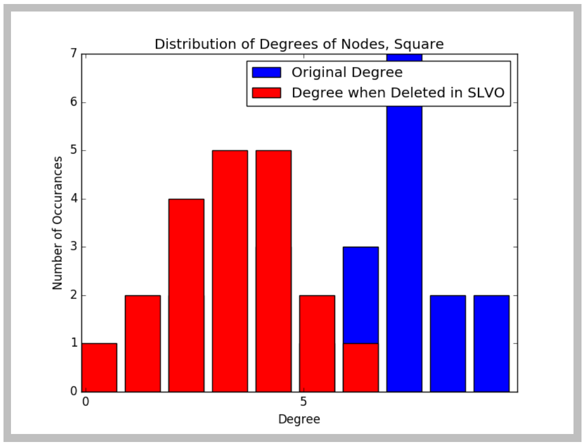


Figure 2: Smallest-last vertex ordering coloring process



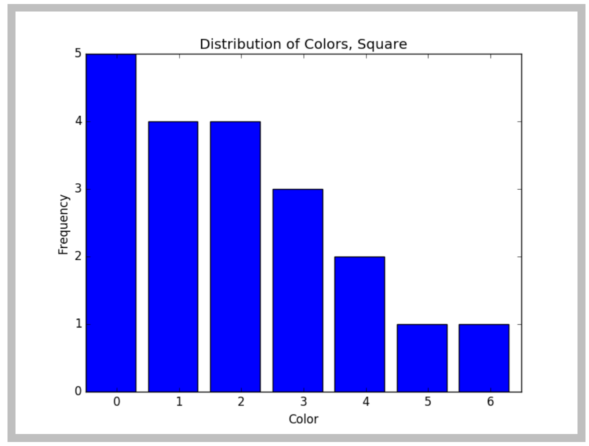


Figure 3: Distribution of degree when deleted and color set size for the 20 node walkthrough

the node radius, this should grow linearly as well. However, because each graph is randomly generated, some graphs can have nodes that are closer to sorted order than others. This can effect the measured runtime.

2.4.3 Graph Coloring

Smallest-last vertex ordering can be verified by looking at the distribution of the degrees of nodes when deleted. Since this algorithm repeatedly removes the node with the fewest connections, and because the removal of that node will cause the fewest number of nodes to move to the next lowest bucket, we would expect the bulk of the nodes to have a large degree when they are deleted. This would be indicated by a negative skew in the distribution of degrees when deleted. Additionally, since the nodes are only removed when they satisfy the criteria of being the node with the minimum degree, we should see the standard deviation of the distribution of nodes to be much smaller than in the original distribution of degrees. Both of these features can be found in Figures 8, 9, and 10 which plot the original distribution of degrees alongside the distribution of degrees when deleted. We see that the distribution of degrees when deleted follows a normal distrbution with a negative skew and a relatively small standard deviation compared to the original distribution of degrees.

The color sets can be verified by looking at the distribution of colors used to color the graph. The number of items in each color should follow a trend where the first colors used have the most members, and the last colors have the fewest items because they are used to accommodate nodes where the earlier colors are all used by a node's neighbors. This trend is shown in Figures 11, 12, and 13.

To further verify the accuracy of the smallest-last coloring implementation additional code was used to verify that the coloring result was correct while running benchmarks. All of the nodes in the smallest-last vertex ordering are traversed, and for each node, the edges are visited to see if any adjacent nodes have the same color as the node being checked. If any of these neighbors have the same color, the coloring is not correct and our independent sets cannot be used for backbone determination. All of the benchmarks ran and returned valid colorings.

References

[1] Matula, David; Beck, Leland, Smallest-Last Ordering and Clustering and Graph Coloring Algorithms, 1983

[2] Johnson, Ian, Linear-Time Computation of High-Converage Backbones for Wireless Sensor Networks, https://github.com/ianjjohnson/SensorNetwork/blob/master/Report/Report.pdf, 2016

[3] Fry, Ben; Reas Casey, Processing, https://processing.org, 2018 v3.3.7

[4] The Matplotlib Development, matplotlib, https://matplotlib.org, 2018

[5] Weisstein, Eric W., Wolfram MathWorld Sphere Point Picking, http://mathworld.wolfram.com/SpherePointPicking.html

[6] Weisstein, Eric W., Wolfram MathWorld Spherical Cap, http://mathworld.wolfram.com/SphericalCap.html

[7] Weisstein, Eric W., Wolfram MathWorld Polyhedral Formula, http://mathworld.wolfram.com/PolyhedralFormula.html

[8] Kogler, Jakob, Finding bridges in a graph in $O(N + M)$, https://e-maxx-eng.appspot.com/graph/bridge-searching.html, 2018

[9] Peters, Tim, Timsort, http://svn.python.org/projects/python/trunk/Objects/listsort.txt

[10] Rees, Gareth, Python's underlying hash data structure for dictionaries, https://stackoverflow.com/questions/4279358/pythons-underlying-hash-data-structure-for-dictionaries, 2010

[11] Thomas, Alec, Why is tuple faster than list?, https://stackoverflow.com/questions/3340539/why-is-tuple-faster-than-list, 2010

[12] Kruse, Lars, Python Speed, Performance Tips, https://wiki.python.org/moin/PythonSpeed/PerformanceTips, 2016

3 Appendix A - Figures

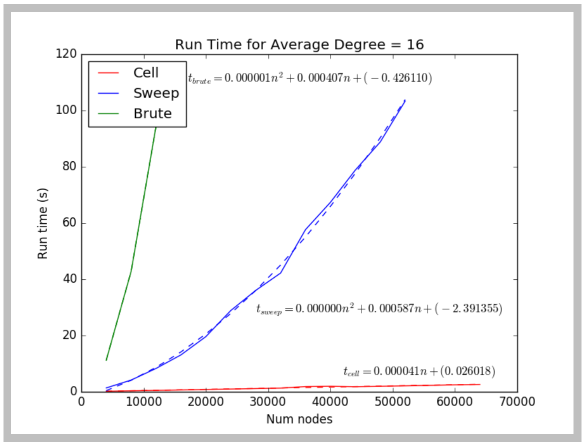


Figure 4: Runtime for edge detection methods. left: constant average degree of 16, right: variable average degree

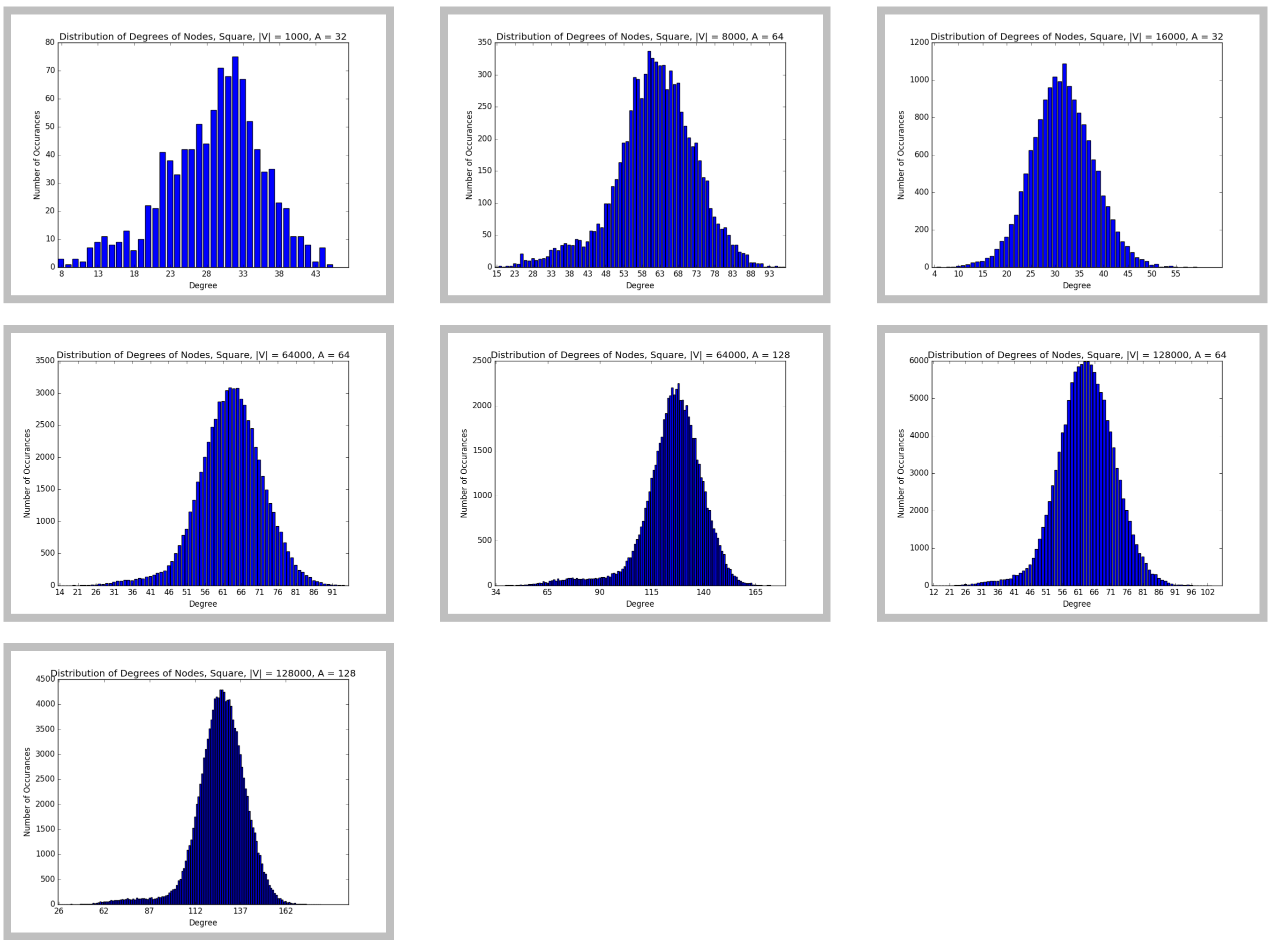


Figure 5: Square benchmarks distribution of degree graphs

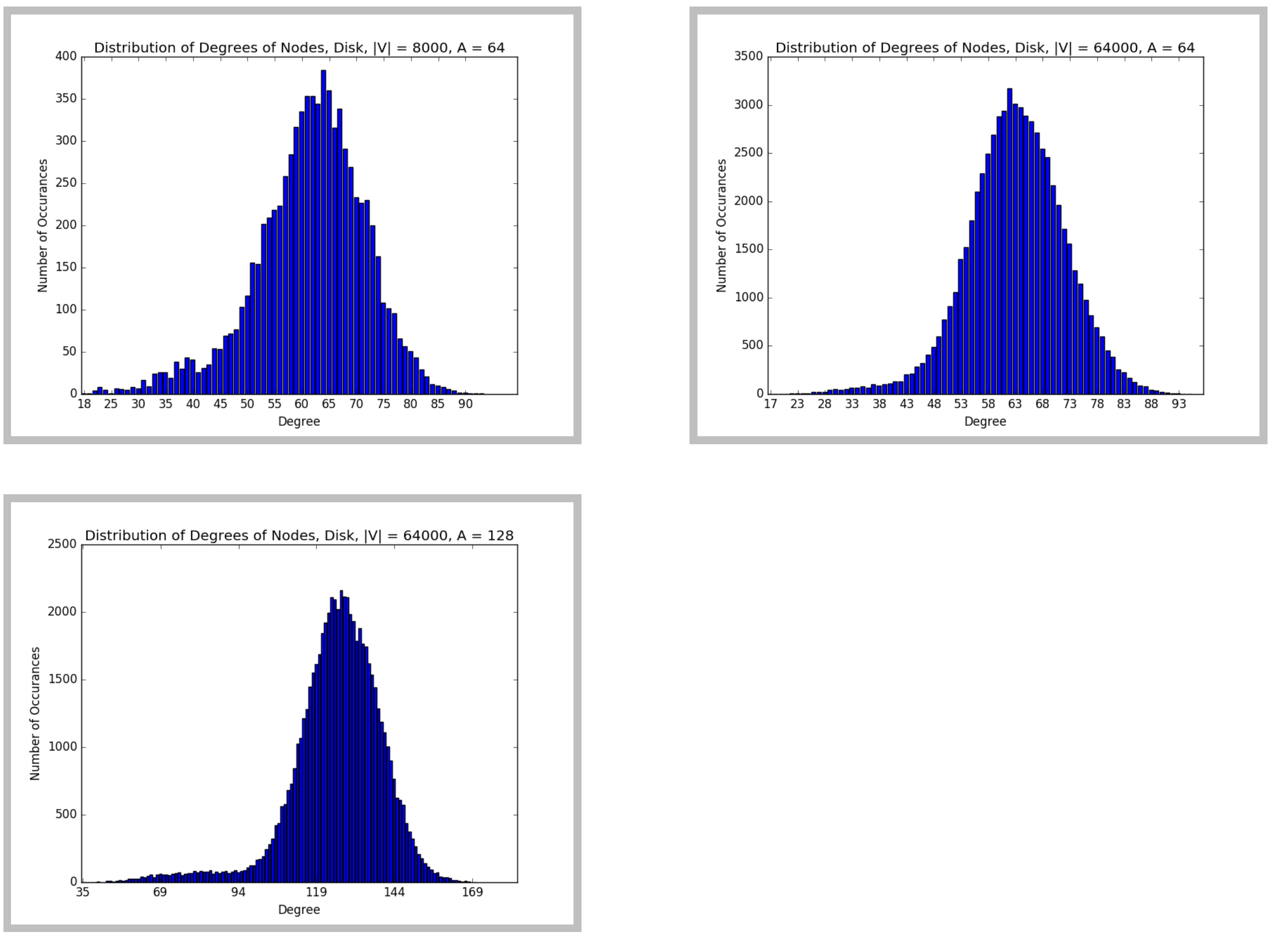


Figure 6: Disk benchmarks distribution of degree graphs

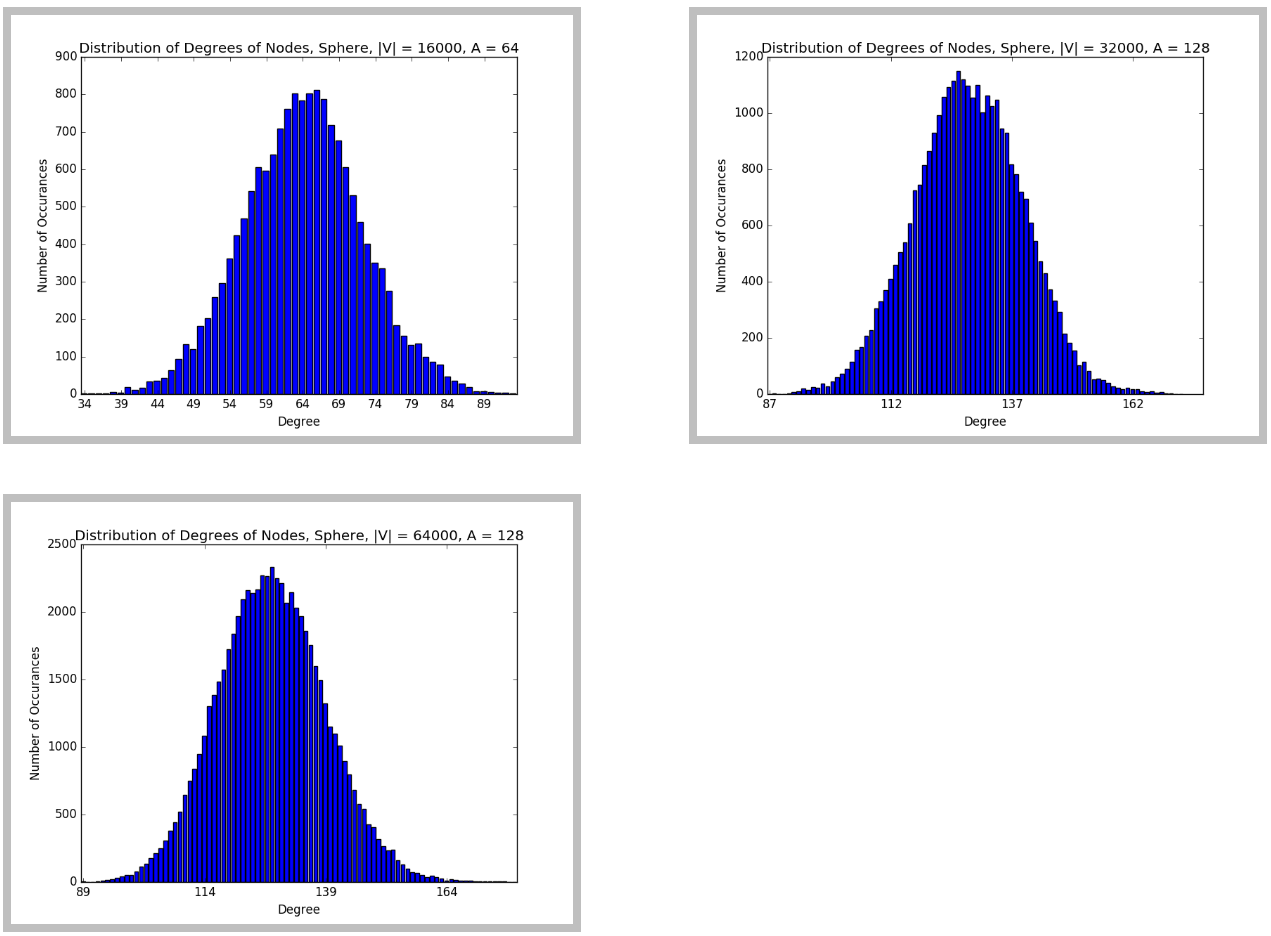


Figure 7: Sphere benchmarks distribution of degree graphs

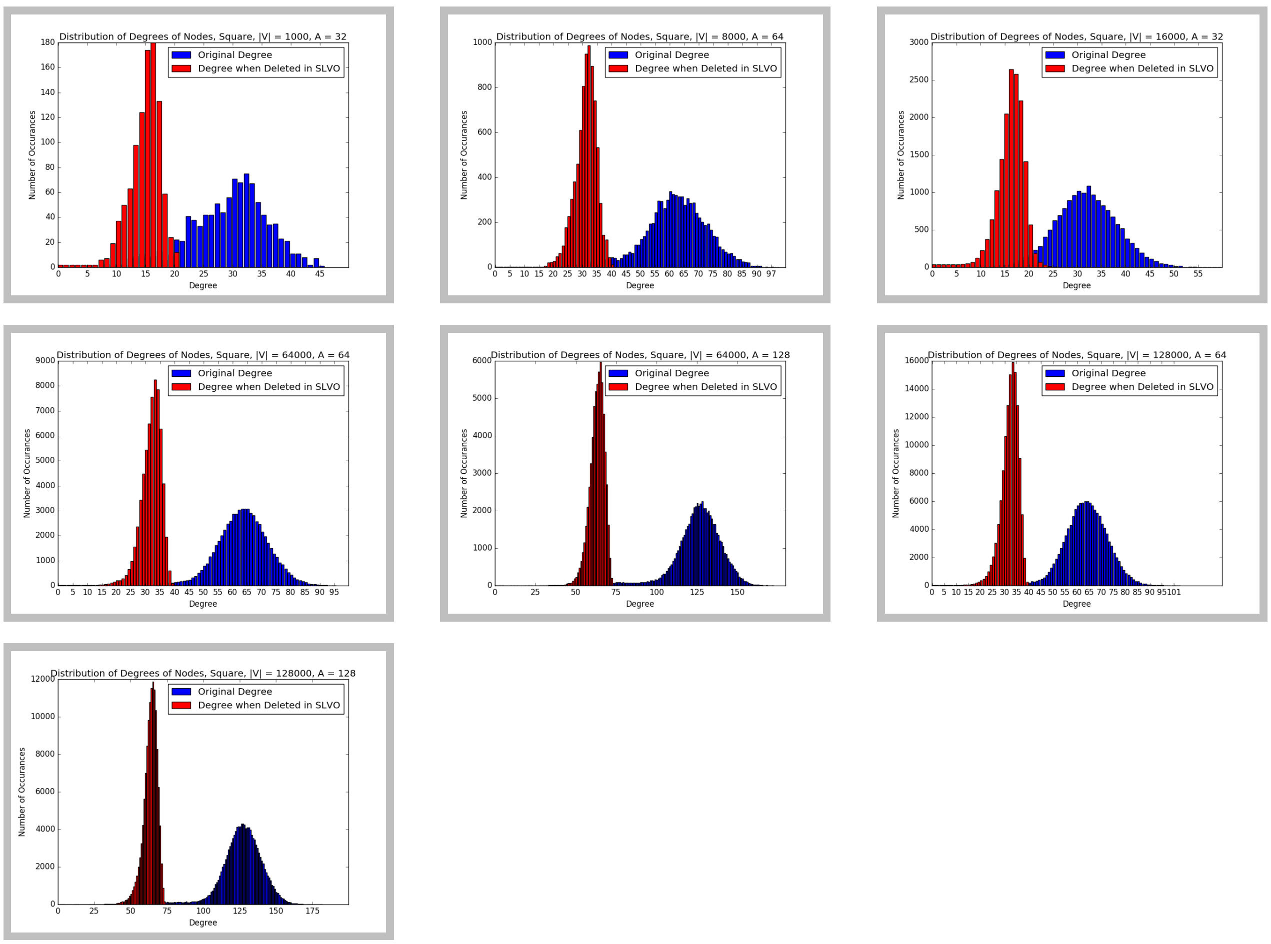


Figure 8: Square benchmarks distribution of degree when deleted graphs

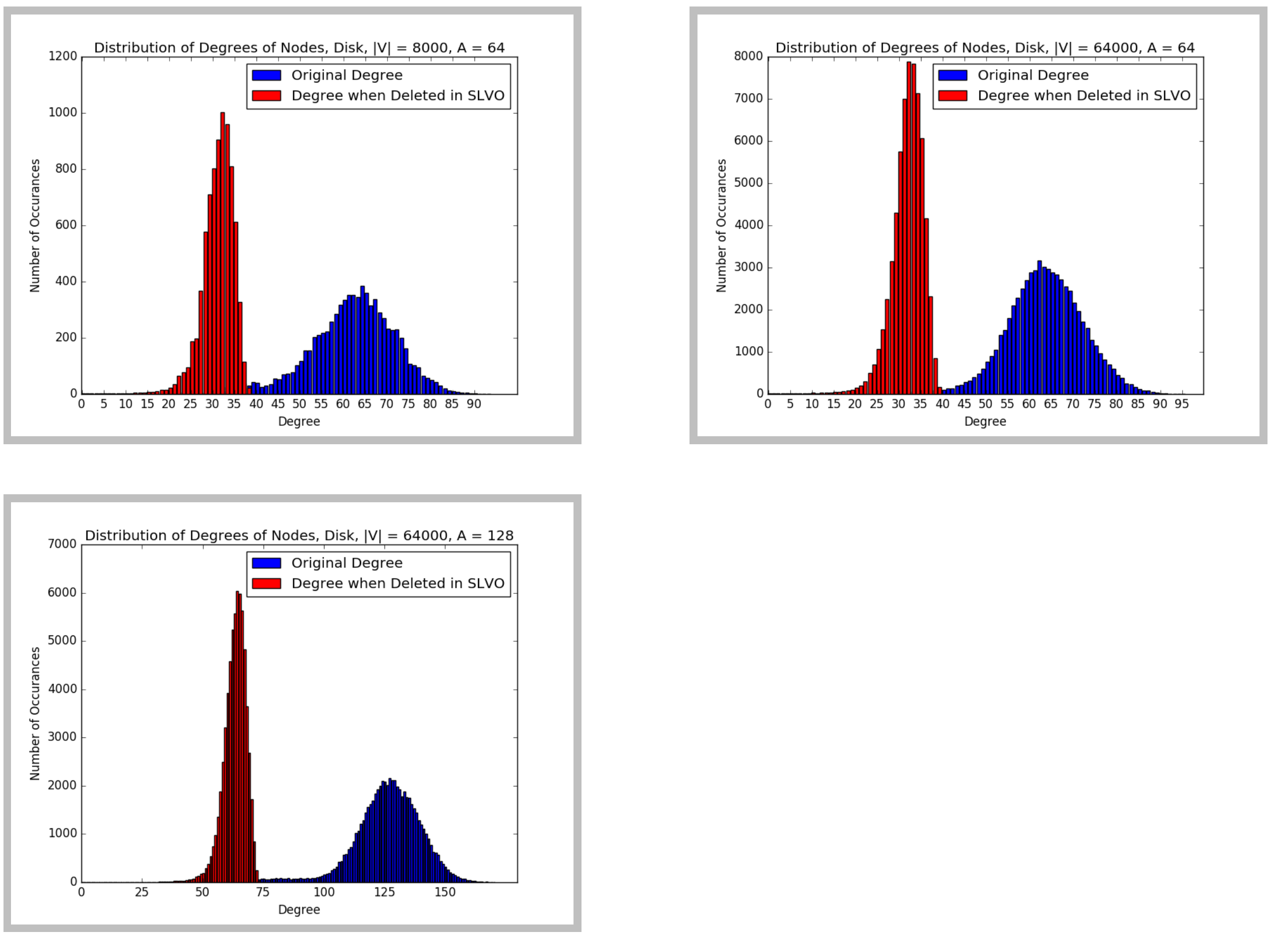


Figure 9: Disk benchmarks distribution of degree when deleted graphs

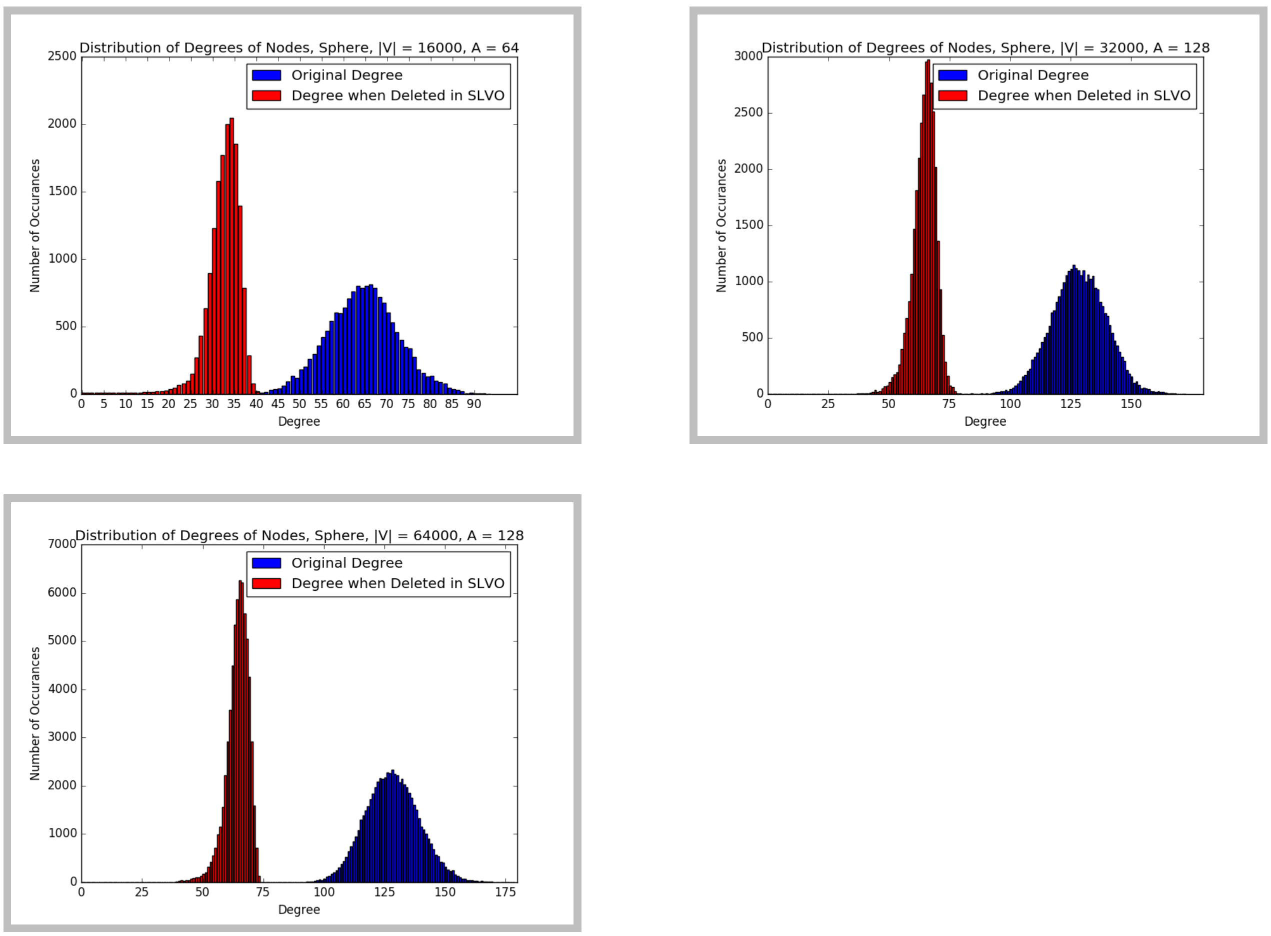


Figure 10: Sphere benchmarks distribution of degree when deleted graphs

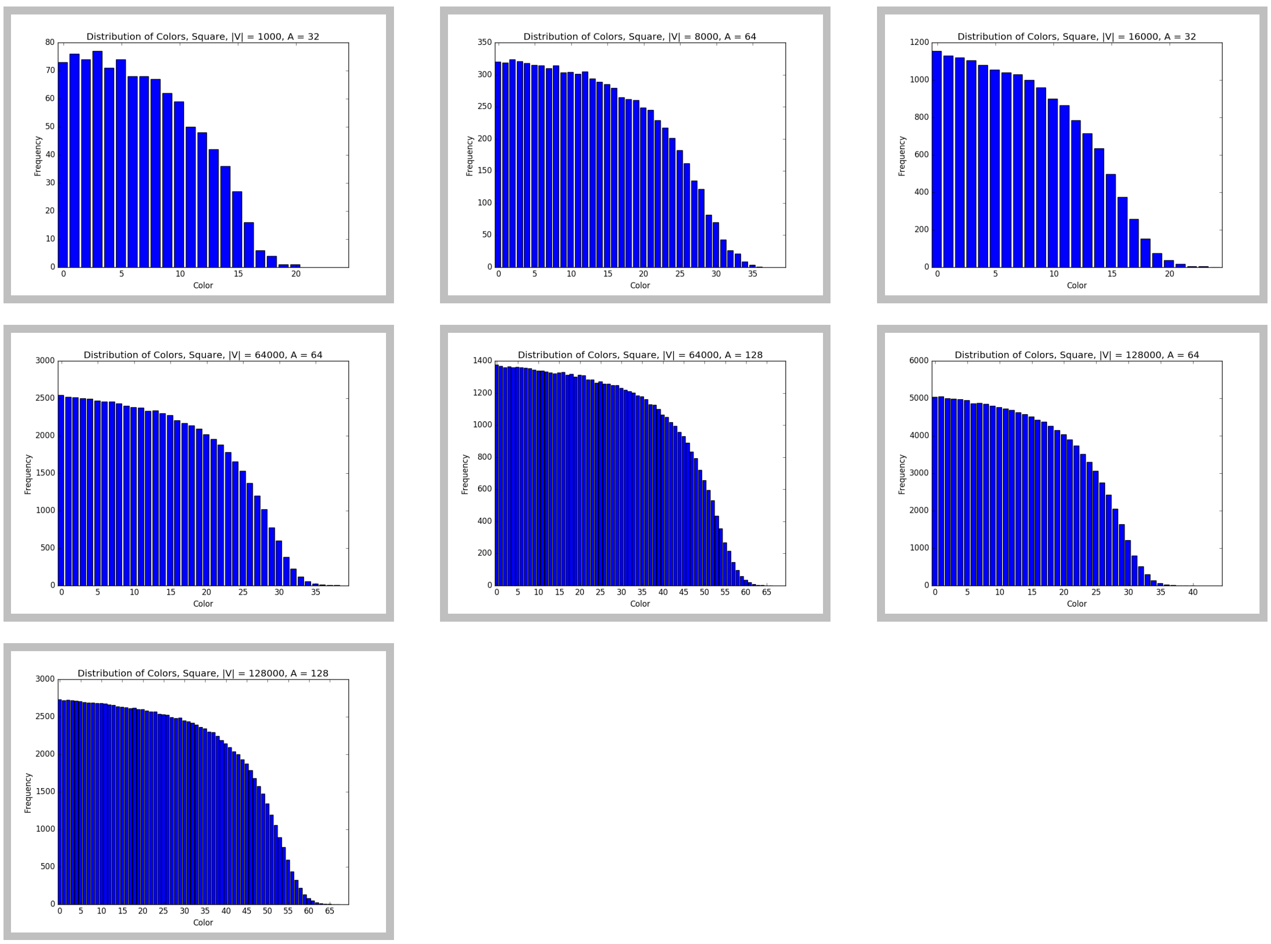


Figure 11: Square benchmarks distribution of colors graphs

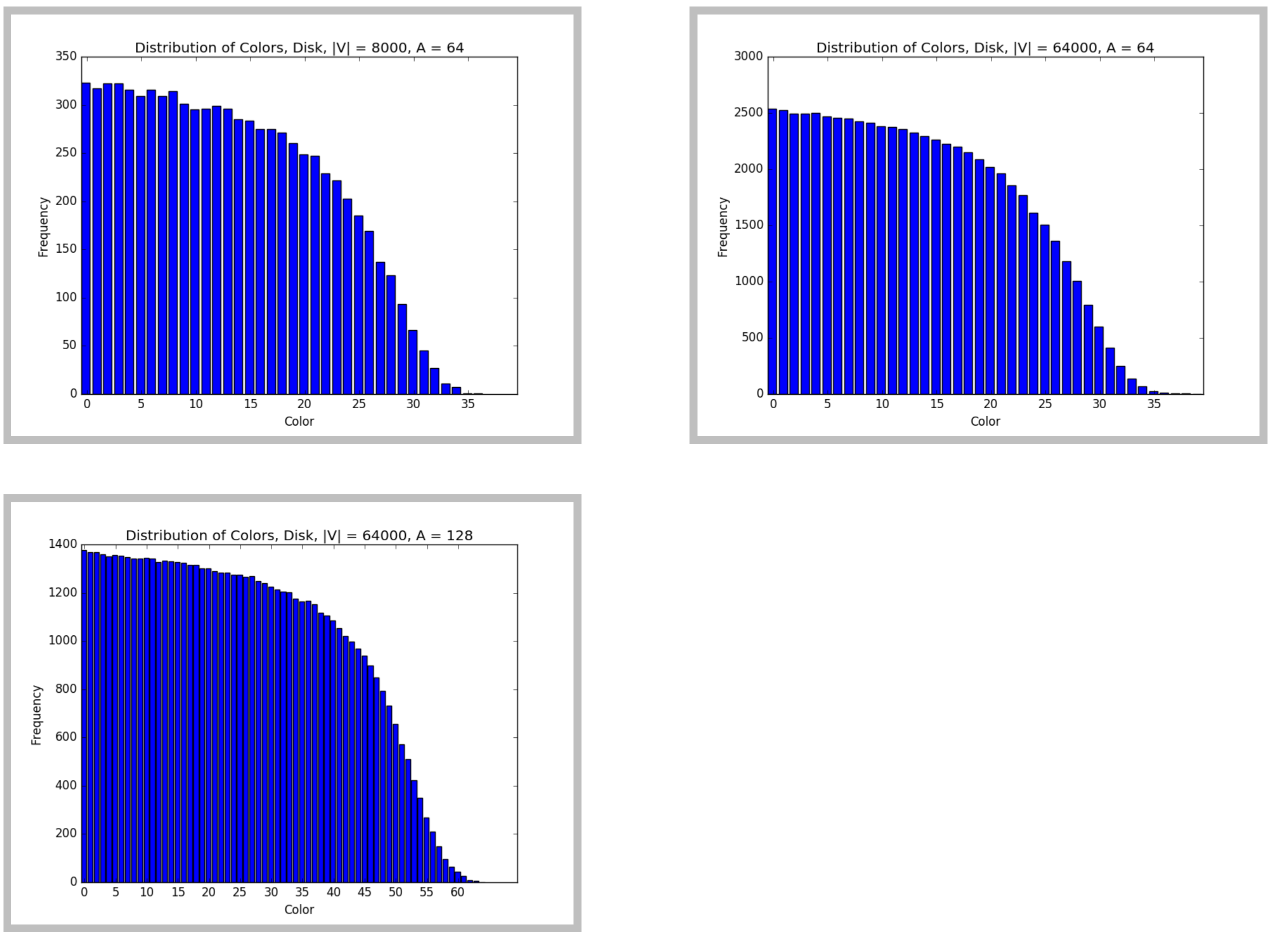


Figure 12: Disk benchmarks distribution of colors graphs

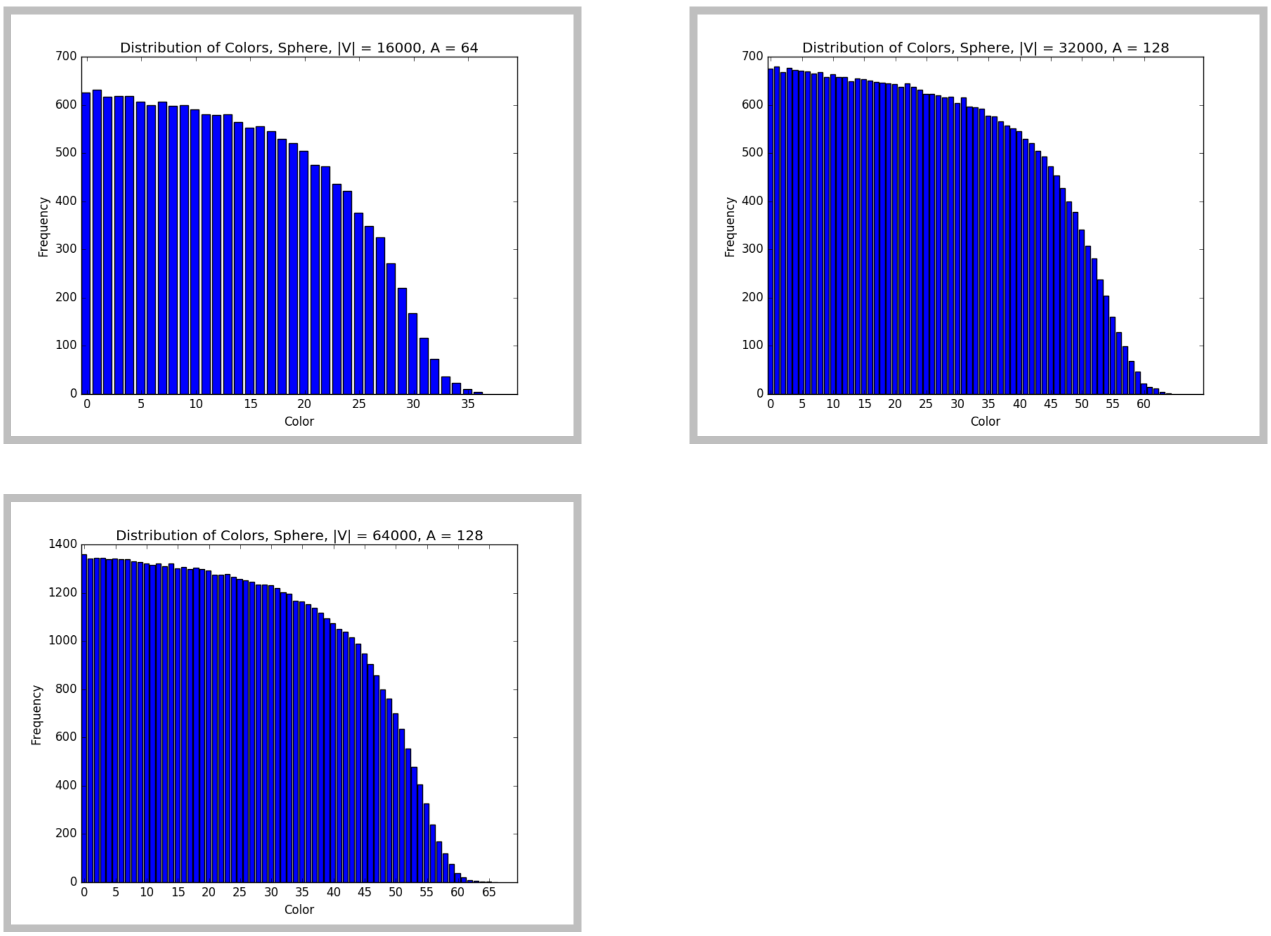


Figure 13: Sphere benchmarks distribution of colors graphs

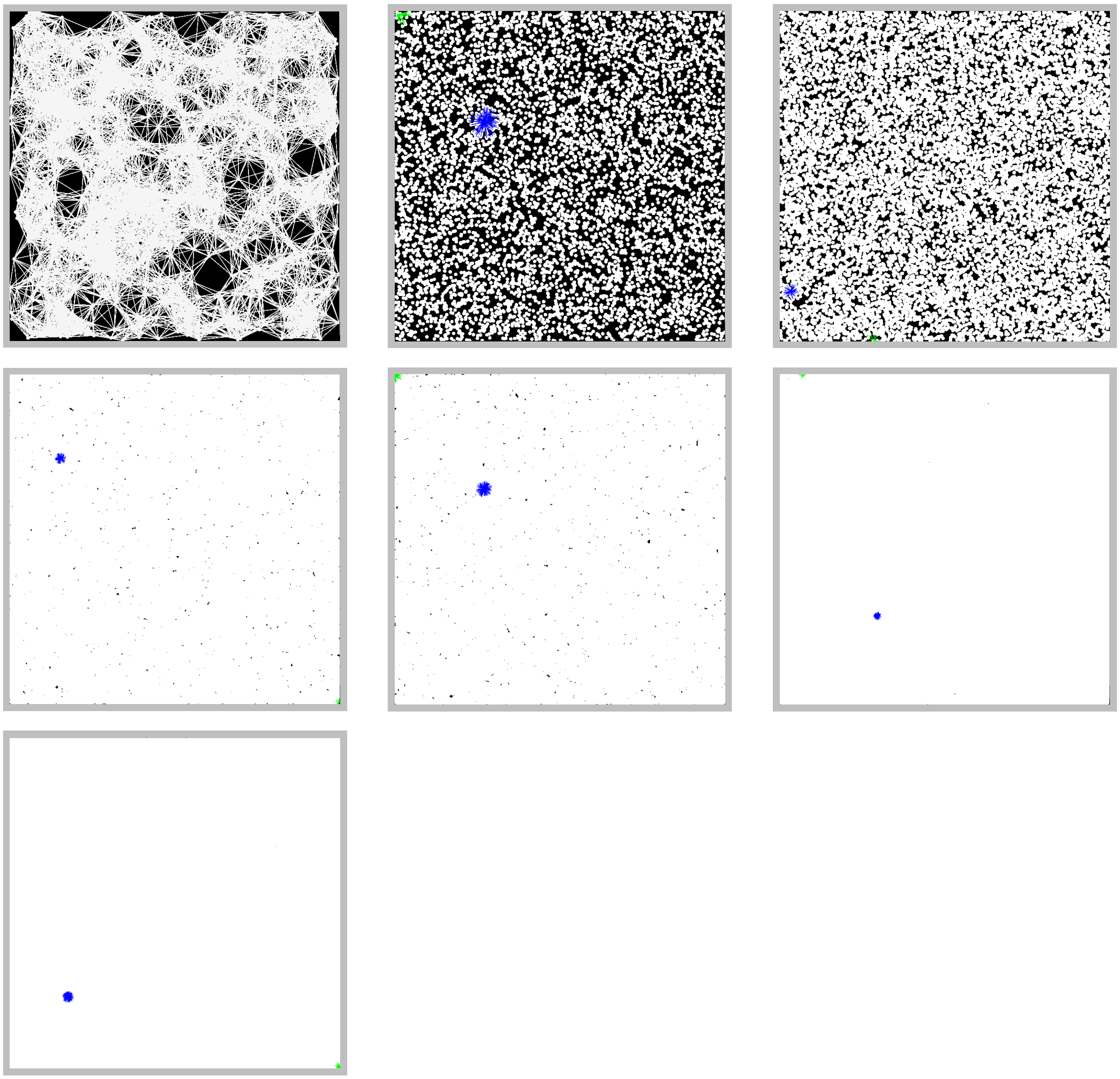


Figure 14: Square benchmark graphs

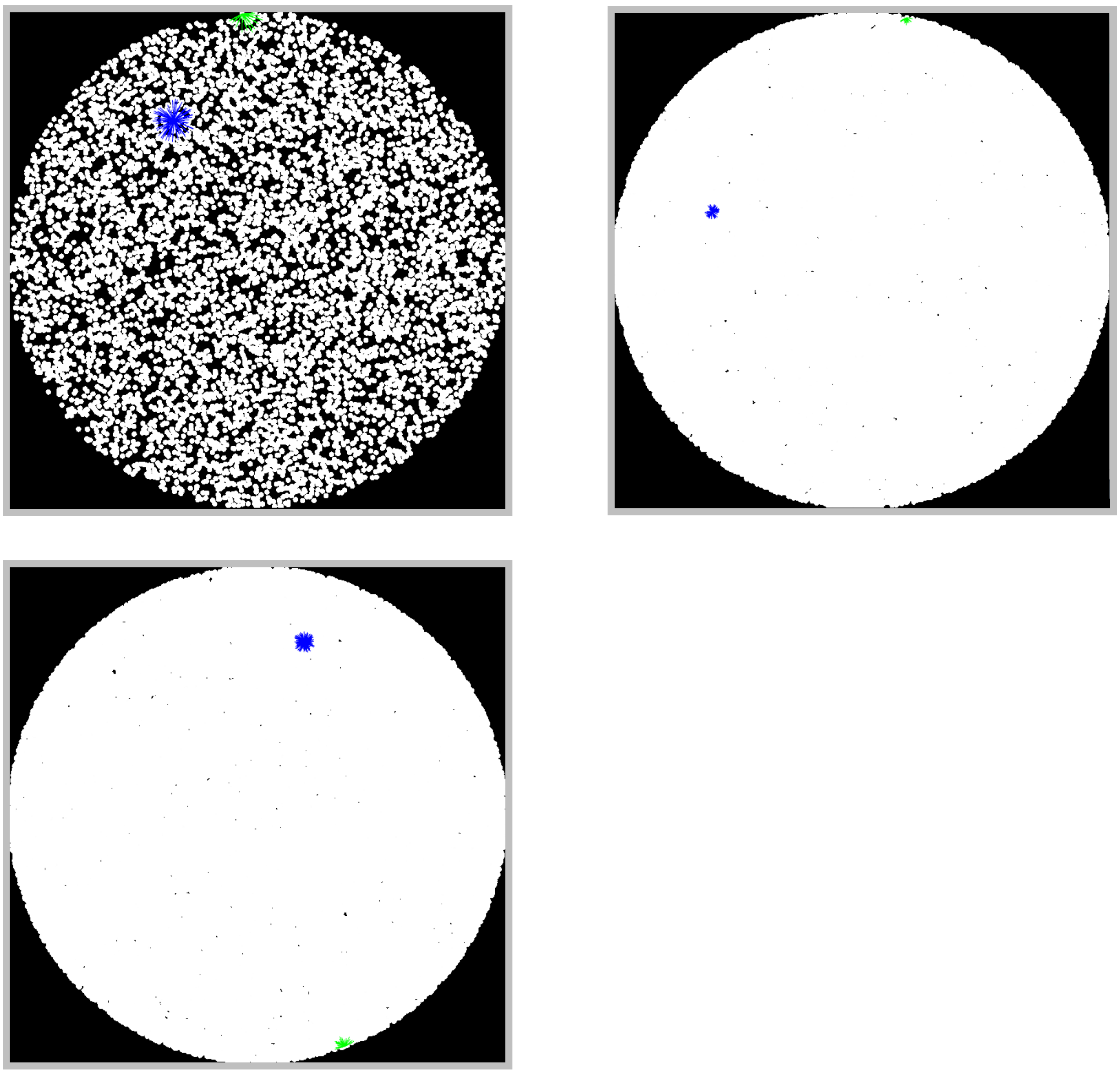


Figure 15: Disk benchmark graphs

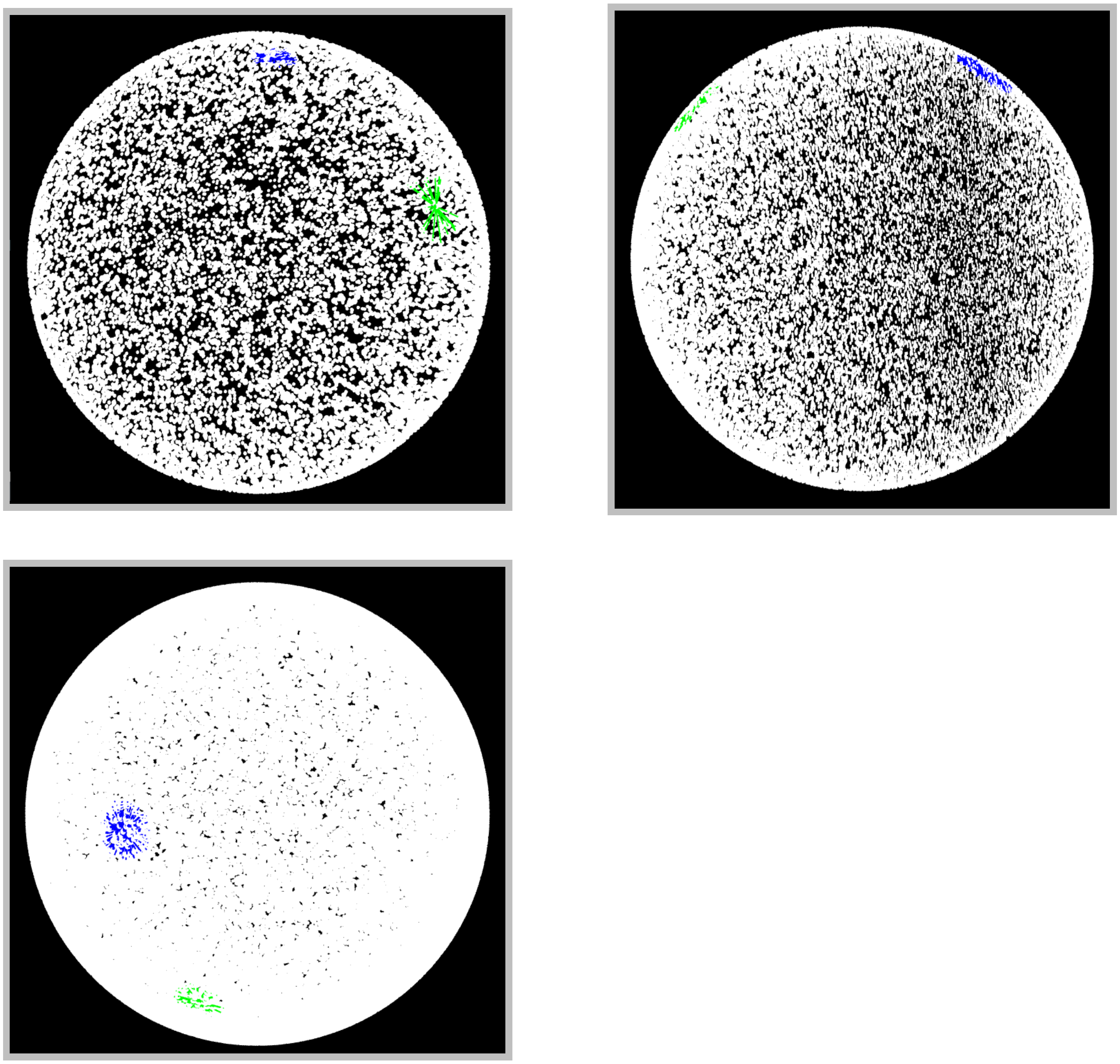


Figure 16: Sphere benchmark graphs

4 Appendix B - Code Listings

Listing 1: Processing driver

```
import random
import sys
import time
import math
from collections import Counter
from objects.topology import Square, Disk, Sphere

CANVAS_HEIGHT = 720
CANVAS_WIDTH = 720

NUM_NODES = 1000
AVG_DEG = 16

MAX_NODES_TO_DRAW_EDGES = 8000

RUN_BENCHMARK = False

def setup():
    size(CANVAS_WIDTH, CANVAS_HEIGHT, P3D)
    background(0)

def draw():
    global curr_vis

    if curr_vis == 0:
        topology.drawGraph(MAX_NODES_TO_DRAW_EDGES)
    elif curr_vis == 1:
        topology.drawSlvo()
    elif curr_vis == 2:
        topology.drawColoring()
    elif curr_vis == 3:
        topology.drawPairs(0)
    elif curr_vis == 4:
        topology.drawPairs(1)
    elif curr_vis == 5:
        topology.drawPairs(2)
    elif curr_vis == 6:
        topology.drawPairs(3)
    elif curr_vis == 7:
        topology.drawBackbones()

def keyPressed():
    global curr_vis
    global step_size

    if key == ' ':
        toggleLooping()
    elif key == 'i':
        topology.switchFgBg()
    elif key == 'l':
        incrementVis()
        topology.mightResetCurrNode()
    elif key == 'h':
        decrementVis()
        topology.mightResetCurrNode()
    elif key == 'k':
        if curr_vis > 2 and curr_vis < 7:
            topology.incrementCurrPair()
        elif curr_vis == 7:
            topology.incrementCurrBackbone()
        else:
            topology.incrementCurrNode(step_size)
    elif key == 'j':
        if curr_vis > 2 and curr_vis < 7:
```

```
            topology.decrementCurrPair()
        elif curr_vis == 7:
            topology.decrementCurrBackbone()
        else:
            topology.decrementCurrNode(step_size)
    elif key == 'y':
        saveFrame("../report/images/{}-{}.png".format(
                   "slvo" if curr_vis == 1 else "color", topology.curr_node))
    elif key >= '0' and key <= '9':
        step_size = 2**int(key)
        print "New step size:", step_size
    elif key == ']':
        step_size = 2*step_size
        print "New step size:", step_size
    elif key == '[':
        step_size = step_size/2
        print "New step size:", step_size
    elif key == 'm':
        print "\n---- Help Menu ----"
        print "Use 'hjkl' to move between visualizations"
        print "Press 'i' to invert the color scheme"
        print "Press space to pause rotation of the sphere"
        print "Press 'y' to take a screenshot of the current frame"
        print "Entering a number n between 0 and 9 will set the step size to 2^n nodes"
        print "Using ']' will double the step size, '[' will half it"

# def mouseDragged():
#     global topology
#     topology.updateRotation(mouseX, mouseY)

def toggleLooping():
    global is_looping
    if is_looping:
        noLoop()
        is_looping = False
    else:
        loop()
        is_looping = True

def incrementVis():
    global curr_vis
    global topology
    if curr_vis < 7:
        curr_vis += 1
    background(topology.color_bg)

def decrementVis():
    global curr_vis
    global topology
    if curr_vis > 0:
        curr_vis -= 1
    background(topology.color_bg)

def main():
    # sys.setrecursionlimit(32000)

    global is_looping
    global curr_vis
    global step_size
    is_looping = True
    curr_vis = 0
    step_size = 1

    global topology
    topology = Square()
    # topology = Disk()
    # topology = Sphere()
```

```
    topology.num_nodes = NUM_NODES
    topology.avg_deg = AVG_DEG
    topology.canvas_height = CANVAS_HEIGHT
    topology.canvas_width = CANVAS_WIDTH

    if RUN_BENCHMARK:
        n_benchmark = 1
        topology.prepBenchmark(n_benchmark)

    run_time = time.clock()

    topology.generateNodes()
    topology.findEdges(method="cell")
    topology.colorGraph()
    topology.generateBackbones()

    print "Average degree: {}".format(topology.findAvgDegree())
    print "Min degree: {}".format(topology.getMinDegree())
    print "Max degree: {}".format(topology.getMaxDegree())
    print "Num edges: {}".format(topology.findNumEdges())
    print "Node r: {0:.3f}".format(topology.node_r)
    print "Terminal clique size: {}".format(topology.term_clique_size)
    print "Number of colors: {}".format(len(set(topology.node_colors)))
    print "Max degree when deleted: {}".format(max(topology.deg_when_del.values())
    )
    color_cnt = Counter(topology.node_colors)
    print "Max color set size: {}  color: {}".format(color_cnt.most_common(1)
    [0][1],
                                                     color_cnt.most_common(1)
    [0][0])

    run_time = time.clock() - run_time
    print "Run time: {0:.3f} s".format(run_time)

    print "\nPress 'm' for the menu"

main()
```

Listing 2: Topology class and subclasses

```
import random
import math
import time
from collections import deque

# benchmarks (num_nodes, avg_deg)
SQUARE_BENCHMARKS = [(1000,32), (8000,64), (16000,32), (64000,64), (64000,128),
                     (128000,64), (128000, 128)]
DISK_BENCHMARKS = [(8000,64), (64000,64), (64000,128)]
SPHERE_BENCHMARKS = [(16000,64), (32000,128), (64000,128)]

"""
Topology - super class for the shape of the random geometric graph
"""
class Topology(object):

    num_nodes = 100
    avg_deg = 0
    canvas_height = 720
    canvas_width = 720

    def __init__(self):
        self.nodes = []
        self.edges = {}
        self.node_r = 0.0
        self.minDeg = ()
        self.maxDeg = ()
```

```
        self.slvo = []
        self.deg_when_del = {}
        self.node_colors = []
        self.pairs = []
        self.no_tails = []
        self.major_comps = []
        self.clean_pairs = []
        self.backbones = []
        self.backbones_meta = []
        self.curr_node = 0
        self.curr_pair = 0
        self.curr_backbone = 0

        self.rot = (0,0,0)
        self.color_bg = 0
        self.color_fg = 255

    # public funciton for generating nodes of the graph, must be subclassed
    def generateNodes(self):
        print "Method for generating nodes not subclassed"

    # public function for finding edges
    def findEdges(self, method="brute"):
        self._getRadiusForAverageDegree()
        self._addNodesAsEdgeKeys()

        if method == "brute":
            self._bruteForceFindEdges()
        elif method == "sweep":
            self._sweepFindEdges()
        elif method == "cell":
            self._cellFindEdges()
        else:
            print "Find edges method not defined: {}".format(method)

        self._findMinAndMaxDegree()

    # brute force edge detection
    def _bruteForceFindEdges(self):
        for i, n in enumerate(self.nodes):
            for j, m in enumerate(self.nodes):
                if i != j and self._distance(n, m) <= self.node_r:
                    self.edges[n].append(j)

    # sweep edge detection
    def _sweepFindEdges(self):
        self.nodes.sort(key=lambda x: x[0])

        for i, n in enumerate(self.nodes):
            search_space = []
            for j in range(1,self.num_nodes-i):
                if abs(n[0] - self.nodes[i+j][0]) <= self.node_r:
                    search_space.append(i+j)
                else:
                    break
            for j in search_space:
                if self._distance(n, self.nodes[j]) <= self.node_r:
                    self.edges[n].append(j)
                    self.edges[self.nodes[j]].append(i)

    # cell edge detection
    def _cellFindEdges(self):
        num_cells = int(1/self.node_r) + 1
        cells = []
        for i in range(num_cells):
            cells.append([[] for j in range(num_cells)])

        for i, n in enumerate(self.nodes):
```

```
            cells[int(n[0]/self.node_r)][int(n[1]/self.node_r)].append(i)

        for i in range(num_cells):
            for j in range(num_cells):
                for n_i in cells[i][j]:
                    for c in self._findAdjCells(i, j, num_cells):
                        for m_i in cells[c[0]][c[1]]:
                            if self._distance(self.nodes[n_i], self.nodes[m_i]) <= self.node_r:
                                self.edges[self.nodes[n_i]].append(m_i)
                                self.edges[self.nodes[m_i]].append(n_i)
                    for m_i in cells[i][j]:
                        if self._distance(self.nodes[n_i], self.nodes[m_i]) <= self.node_r and n_i != m_i:
                            self.edges[self.nodes[n_i]].append(m_i)

    # cell edge detection helper function
    def _findAdjCells(self, i, j, n):
        adj_cells = [(1,-1), (0,1), (1,1), (1,0)]
        return (((i+x[0])%n,(j+x[1])%n) for x in adj_cells)

    # function for finding the radius needed for the desired average degree
    # must be subclassed
    def _getRadiusForAverageDegree(self):
        print "Method for finding necessary radius for average degree not subclassed"

    # helper function for findEdges, initializes edges dict
    def _addNodesAsEdgeKeys(self):
        self.edges = {n:[] for n in self.nodes}

    # claculates the distance between two nodes (2D)
    def _distance(self, n, m):
        return math.sqrt((n[0] - m[0])**2+(n[1] - m[1])**2)

    # public function for finding the number of edges
    def findNumEdges(self):
        sigma_edges = 0
        for k in self.edges.keys():
            sigma_edges += len(self.edges[k])

        return sigma_edges/2

    # public function for finding the average degree of nodes
    def findAvgDegree(self):
        return 2*self.findNumEdges()/self.num_nodes

    # helper funciton for finding nodes with min and max degree
    def _findMinAndMaxDegree(self):
        self.minDeg = self.edges.keys()[0]
        self.maxDeg = self.edges.keys()[0]

        for k in self.edges.keys():
            if len(self.edges[k]) < len(self.edges[self.minDeg]):
                self.minDeg = k
            if len(self.edges[k]) > len(self.edges[self.maxDeg]):
                self.maxDeg = k

    # public function for getting the minimum degree
    def getMinDegree(self):
        return len(self.edges[self.minDeg])

    # public functino for getting the maximum degree
    def getMaxDegree(self):
        return len(self.edges[self.maxDeg])

    # public function for setting up the benchmark to run, must be subclassed
    def prepBenchmark(self, n):
```

```
        print "Method for preparing benchmark not subclassed"

    # public function for drawing the graph
    def drawGraph(self, n_limit):
        self._drawNodes(self.nodes)
        if self.num_nodes <= n_limit:
            self._drawEdges(self.nodes)
        else:
            self._drawMinMaxDegNodes()

    # responsible for drawing the nodes in the canvas
    def _drawNodes(self, node_list):
        strokeWeight(2)
        stroke(self.color_fg)
        fill(self.color_fg)

        for n in node_list:
            ellipse(n[0]*self.canvas_width, n[1]*self.canvas_height, 5, 5)

    # responsible for drawing the edges in the canavas
    def _drawEdges(self, node_list):
        strokeWeight(1)
        stroke(245)
        fill(self.color_fg)

        s = set(node_list)

        for n in node_list:
            for m_i in self.edges[n]:
                if self.nodes[m_i] in s:
                    line(n[0]*self.canvas_width, n[1]*self.canvas_height, self.nodes[m_i][0]*self.canvas_width, self.nodes[m_i][1]*self.canvas_height)

    # responsible for drawing the edges of the min and max degree nodes
    def _drawMinMaxDegNodes(self):
        strokeWeight(1)
        stroke(0,self.color_fg,0)
        fill(self.color_fg)
        for n_i in self.edges[self.minDeg]:
            line(self.minDeg[0]*self.canvas_width, self.minDeg[1]*self.canvas_height, self.nodes[n_i][0]*self.canvas_width, self.nodes[n_i][1]*self.canvas_height)

        stroke(0,0,self.color_fg)
        for n_i in self.edges[self.maxDeg]:
            line(self.maxDeg[0]*self.canvas_width, self.maxDeg[1]*self.canvas_height, self.nodes[n_i][0]*self.canvas_width, self.nodes[n_i][1]*self.canvas_height)

    # uses smallest last vertex ordering to color the graph
    def colorGraph(self):
        self.slvo, self.deg_when_del = self._smallestLastVertexOrdering()
        self.node_colors = self._assignNodeColors(self.slvo)
        self.color_map = self._mapColorsToRGB(self.node_colors)

    # constructs a degree structure and determines the smallest last vertex ordering
    def _smallestLastVertexOrdering(self):
        deg_sets = {l:set() for l in range(len(self.edges[self.maxDeg])+1)}
        deg_when_del = {n:len(self.edges[n]) for n in self.nodes}

        for i, n in enumerate(self.nodes):
            deg_sets[deg_when_del[n]].add(i)

        smallest_last_ordering = []

        clique_found = False
        j = len(self.nodes)
```

```python
        while j > 0:
            # get the current smallest bucket
            curr_bucket = 0
            while len(deg_sets[curr_bucket]) == 0:
                curr_bucket += 1

            # if all the remaining nodes are connected we have the terminal clique
            if not clique_found and len(deg_sets[curr_bucket]) == j:
                clique_found = True
                self.term_clique_size = curr_bucket

            # get node with smallest degree
            v_i = deg_sets[curr_bucket].pop()
            smallest_last_ordering.append(v_i)

            # decrement position of nodes that shared an edge with v
            for n_i in (n_i for n_i in self.edges[self.nodes[v_i]] if n_i in deg_sets[deg_when_del[self.nodes[n_i]]]):
                deg_sets[deg_when_del[self.nodes[n_i]]].remove(n_i)
                deg_when_del[self.nodes[n_i]] -= 1
                deg_sets[deg_when_del[self.nodes[n_i]]].add(n_i)

            j -= 1

        # reverse list since it was built shortest-first
        return smallest_last_ordering[::-1], deg_when_del

    # assigns the colors to nodes given in a smallest-last vertex ordering as a parallel array
    def _assignNodeColors(self, slvo):
        colors = [-1 for _ in range(len(slvo))]
        for i in slvo:
            adj_colors = set([colors[j] for j in self.edges[self.nodes[i]]])
            color = 0
            while color in adj_colors:
                color += 1
            colors[i] = color

        return colors

    # generates random color codes for each color set and returns them in a dictionary
    def _mapColorsToRGB(self, color_list):
        s = set(color_list)
        color_map = {}
        while len(s) > 0:
            c = s.pop()
            color_map[c] = (random.randint(0,255), random.randint(0,255), random.randint(0,255))

        return color_map

    # draw nodes as they are removed in smallest-last vertex ordering
    def drawSlvo(self):
        l = [self.nodes[i] for i in self.slvo[0:self.num_nodes - self.curr_node]]
        self._drawNodes(l)
        self._drawEdges(l)

    # increments curr_node, used to limit the number of nodes drawn
    def incrementCurrNode(self, s):
        if self.curr_node + s <= self.num_nodes:
            self.curr_node += s
            background(self.color_bg)
        elif self.curr_node != self.num_nodes:
            self.curr_node = self.num_nodes
            background(self.color_bg)

    # decrements curr_node, used to limit the number of nodes drawn
```

```
    def decrementCurrNode(self, s):
        if self.curr_node - s >= 0:
            self.curr_node -= s
            background(self.color_bg)
        elif self.curr_node != 0:
            self.curr_node = 0
            background(self.color_bg)

    # used to reset curr node if all nodes have been drawn and the method changes
    def mightResetCurrNode(self):
        if self.curr_node == self.num_nodes:
            curr_node = 0
            background(self.color_bg)

    # increments curr_backbone, used to draw different backbones
    def incrementCurrPair(self):
        if self.curr_pair < len(self.pairs) - 1:
            self.curr_pair += 1
            background(self.color_bg)

    # decrements curr_backbone, used to draw different backbones
    def decrementCurrPair(self):
        if self.curr_pair > 0:
            self.curr_pair -= 1
            background(self.color_bg)

    # increments curr_backbone, used to draw different backbones
    def incrementCurrBackbone(self):
        if self.curr_backbone < len(self.backbones) - 1:
            self.curr_backbone += 1
            background(self.color_bg)

    # decrements curr_backbone, used to draw different backbones
    def decrementCurrBackbone(self):
        if self.curr_backbone > 0:
            self.curr_backbone -= 1
            background(self.color_bg)

    # switch foreground and background colors
    def switchFgBg(self):
        self.color_fg, self.color_bg = self.color_bg, self.color_fg

    # # update the rotation of the drawing
    # def updateRotation(self, x, y):
    #     # self.rot = (self.rot[0], self.rot[1]-math.pi/100, self.rot[2])
    #     # self.rot = (x*math.cos(self.rot[0])*math.pi/500, self.rot[1], self.rot
    [2])
    #     self.rot = (self.rot[0], x*math.cos(self.rot[1])*math.pi/1000, self.rot
    [2])
    #     # rotateX(self.rot[0])
    #     # rotateZ(self.rot[2])
    #     # rotateY(-1*self.rot[1])

    # used to draw the graph with the nodes colored
    def drawColoring(self):
        l = [self.nodes[i] for i in self.slvo[0:self.curr_node]]
        self._drawNodes(l)
        self._applyColors(self.slvo[0:self.curr_node])
        self._drawEdges(l)

    # places colors on the nodes
    def _applyColors(self, node_i_list):
        strokeWeight(5)

        num_colors = max(self.node_colors)

        for n_i in node_i_list:
            c = self.color_map[self.node_colors[n_i]]
```

```python
            stroke(c[0], c[1], c[2])
            fill(c[0], c[1], c[2])
            ellipse(self.nodes[n_i][0]*self.canvas_width, self.nodes[n_i][1]*self.canvas_height, 5, 5)

    # public function for pairing the independent sets and picking the largest backbones
    def generateBackbones(self):
        # pair four largest independent sets
        self.pairs = self._pairIndependentSets(self.node_colors)

        # delete minor components and tails
        self.no_tails, self.major_comps, self.clean_pairs = self._cleanPairs(self.pairs)

        # pick two backbones of largest size
        self.backbones, self.backbones_meta = self._getLargestBackbones(self.clean_pairs)

        # calculate domination
        self.backbones_meta = self._getDonimations(self.backbones, self.backbones_meta)

    # pairs the four largest independent color sets
    def _pairIndependentSets(self, color_list):
        # the first four color sets should be the largest (slvo)
        indep_sets = [set() for _ in range(4)]

        for i, n in enumerate(self.nodes):
            if self.node_colors[i] < 4:
                indep_sets[self.node_colors[i]].add(i)

        # return combinations of sets (union)
        return [s1 | s2 for i, s1 in enumerate(indep_sets) for s2 in indep_sets[i+1:]]

    # removes the minor components and tails from the bipartite subgraphs
    def _cleanPairs(self, bipartites):
        no_tails = []
        major_comps = []
        results = []
        for b in bipartites:
            # remove the tails and save the graph for visualization
            b = self._removeTails(b)
            no_tails.append(b)

            # use BFS to get the major component
            major_comp = self._bfs(b)
            major_comps.append(major_comp)

            # use DFS to remove bridges
            backbone = self._removeBridges(major_comp)
            results.append(backbone)

        return no_tails, major_comps, results

    # remove tails from bipartite, very similar to smallest-last vertex ordering
    def _removeTails(self, bipartite):
        bipartite = bipartite.copy()
        # build graph representation
        points = list(bipartite)
        deg_sets = {l:set() for l in range(len(self.edges[self.maxDeg])+1)}
        deg_map = {n_i:len([e_i for e_i in self.edges[self.nodes[n_i]] if e_i in bipartite]) for n_i in points}

        for i, n in enumerate(self.nodes):
            if i in bipartite:
                deg_sets[deg_map[i]].add(i)
```

```
        # remove nodes with zero or one edge until there are no tails
        while len(deg_sets[0]) > 0 or len(deg_sets[1]) > 0:
            to_remove = deg_sets[0] | deg_sets[1]
            deg_sets[0] = set()
            deg_sets[1] = set()

            for n_i in list(to_remove):
                for e_i in [e_i for e_i in self.edges[self.nodes[n_i]] if e_i in
    bipartite]:
                    if e_i in deg_sets[deg_map[e_i]]:
                        deg_sets[deg_map[e_i]].remove(e_i)
                        deg_map[e_i] -= 1
                        deg_sets[deg_map[e_i]].add(e_i)

                bipartite.remove(n_i)

        return bipartite

    # use BFS to find the major component
    def _bfs(self, bipartite, rm_edges=None):
        points = list(bipartite)
        # used to index into the points array
        index_to_local = {n_i:i for i, n_i in enumerate(points)}
        # used to index into the nodes array
        index_to_global = {i:n_i for i, n_i in enumerate(points)}
        visited = [0 for _ in points]
        visits = []
        components = []

        while 0 in visited:
            visit = 1

            queue = deque()
            root = visited.index(0)
            queue.append(root)
            visited[root] = 1
            # builds a set for the points in each component
            components.append(set([index_to_global[root]]))

            while len(queue) > 0:
                curr = queue.pop()

                for e in [index_to_local[e] for e in self.edges[self.nodes[points[
    curr]]] if e in bipartite]:
                    if rm_edges != None and (e in rm_edges and curr in rm_edges):
                        continue
                    if visited[e] == 0:
                        visit += 1
                        queue.append(e)
                        components[-1].add(index_to_global[e])
                        visited[e] = 1

            visits.append(visit)

        return components[visits.index(max(visits))]

    # removes all bridges and minor blocks from major component
    # algorithm: https://e-maxx-eng.appspot.com/graph/bridge-searching.html
    def _removeBridges(self, major_comp):
        points = list(major_comp)
        # used to index into the points array
        index_to_local = {n_i:i for i, n_i in enumerate(points)}
        # used to index into the nodes array
        index_to_global = {i:n_i for i, n_i in enumerate(points)}
        visited = [0 for _ in points]
        bridge_nodes = set()
        tin = [-1 for _ in points]
```

```python
        fup = [-1 for _ in points]
        visit = 0

        for i, p in enumerate(points):
            if visited[i] == 0:
                self._dfs(major_comp, points, i, p, index_to_local, visited, bridge_nodes, tin, fup, visit)

        return self._bfs(major_comp, bridge_nodes)

    # use DFS to find bridges
    def _dfs(self, comp, points, i, p, index_to_local, visited, bridge_nodes, tin, fup, visit, to=-1):
        visited[i] = 1
        tin[i] = visit
        fup[i] = visit
        visit += 1
        for e in [index_to_local[e] for e in self.edges[self.nodes[p]] if e in comp]:
            if e == to:
                continue
            if visited[e] == 1:
                fup[i] = min(fup[i], tin[e])
            else:
                self._dfs(comp, points, e, points[e], index_to_local, visited, bridge_nodes, tin, fup, visit, to=i)
                fup[i] = min(fup[i], fup[e])
                if fup[e] > tin[i]:
                    if i not in bridge_nodes:
                        bridge_nodes.add(i)
                    if e not in bridge_nodes:
                        bridge_nodes.add(e)

    # public function for drawing the color set pairs
    def drawPairs(self, mode=0):
        l_i = []
        if mode == 0:
            l_i = list(self.pairs[self.curr_pair])
        elif mode == 1:
            l_i = list(self.no_tails[self.curr_pair])
        elif mode == 2:
            l_i = list(self.major_comps[self.curr_pair])
        elif mode == 3:
            l_i = list(self.clean_pairs[self.curr_pair])

        l_n = [self.nodes[i] for i in l_i]
        self._drawNodes(l_n)
        self._applyColors(l_i)
        self._drawEdges(l_n)

    # returns the two major components with the largest size
    def _getLargestBackbones(self, c_pairs):
        sizes = [0, 0]
        result = [None, None]
        for p in c_pairs:
            size = self._calcSize(p)

            if size > min(sizes):
                min_i = sizes.index(min(sizes))
                sizes[min_i] = size
                result[min_i] = p

        # saves backbone meta data (order, size)
        meta = [(len(result[i]), sizes[i]) for i in range(len(result))]
        if sizes[1] > sizes[0]:
            return result[::-1], meta[::-1]

        return result, meta
```

```python

    # calculates the size of a graph
    def _calcSize(self, graph):
        size = 0
        for n_i in list(graph):
            size += len([e for e in self.edges[self.nodes[n_i]] if e in graph])

        return size

    # calculates the percentage of nodes covered by each backbone
    def _getDonimations(self, b_bones, meta):
        for i, b in enumerate(b_bones):
            # find the number of nodes that do not share an edge with a backbone node
            # search all nodes not in backbone
            search_space = set(range(self.num_nodes)) - b
            for n_i in list(search_space):
                for e in self.edges[self.nodes[n_i]]:
                    if e in b:
                        search_space.remove(n_i)
                        break

            meta[i] = (meta[i][0], meta[i][1], (self.num_nodes - len(search_space) + 0.0)/self.num_nodes)

        return meta

    # public function for drawing the backbones
    def drawBackbones(self):
        l_i = list(self.backbones[self.curr_backbone])
        l_n = [self.nodes[i] for i in l_i]
        self._drawNodes(l_n)
        self._applyColors(l_i)
        self._drawEdges(l_n)

"""
Square - inherits from Topology, overloads generateNodes and _getRadiusForAverageDegree
for a unit square topology
"""
class Square(Topology):

    def __init__(self):
        super(Square, self).__init__()

    # places nodes uniformly in a unit square
    def generateNodes(self):
        for i in range(self.num_nodes):
            self.nodes.append((random.uniform(0,1), random.uniform(0,1)))

    # calculates the radius needed for the requested average degree in a unit square
    def _getRadiusForAverageDegree(self):
        self.node_r = math.sqrt(self.avg_deg/(self.num_nodes * math.pi))

    # gets benchmark setting for square
    def prepBenchmark(self, n):
        self.num_nodes = SQUARE_BENCHMARKS[n][0]
        self.avg_deg = SQUARE_BENCHMARKS[n][1]

"""
Disk - inherits from Topology, overloads generateNodes and _getRadiusForAverageDegree
for a unit circle topology
"""
class Disk(Topology):

    def __init__(self):
```

```
        super(Disk, self).__init__()

    # places nodes uniformly in a unit square and regenerates the node if it falls
    # outside of the circle
    def generateNodes(self):
        for i in range(self.num_nodes):
            p = (random.uniform(0,1), random.uniform(0,1))
            while self._distance(p, (0.5,0.5)) > 0.5:
                p = (random.uniform(0,1), random.uniform(0,1))
            self.nodes.append(p)

    # calculates the radius needed for the requested average degree in a unit circle
    def _getRadiusForAverageDegree(self):
        self.node_r = math.sqrt((self.avg_deg + 0.0)/self.num_nodes)/2

    # gets benchmark setting for disk
    def prepBenchmark(self, n):
        self.num_nodes = DISK_BENCHMARKS[n][0]
        self.avg_deg = DISK_BENCHMARKS[n][1]

"""
Sphere - inherits from Topology, overloads generateNodes, _getRadiusForAverageDegree,
and _distance for a unit sphere topology. Also updates the drawGraph function for
a 3D canvas
"""
class Sphere(Topology):

    # adds rotation and node limit variables
    def __init__(self):
        super(Sphere, self).__init__()
        self.rot = (0,math.pi/4,0) # this may move to Topology if rotation is given to the 2D shapes
        # used to control _drawNodes functionality
        self.n_limit = 8000
        self.num_faces = []

    # places nodes in a unit cube and projects them onto the surface of the sphere
    def generateNodes(self):
        for i in range(self.num_nodes):
            # equations for uniformly distributing nodes on the surface area of
            # a sphere: http://mathworld.wolfram.com/SpherePointPicking.html
            u = random.uniform(-1,1)
            theta = random.uniform(0, 2*math.pi)
            p = (
                math.sqrt(1 - u**2) * math.cos(theta),
                math.sqrt(1 - u**2) * math.sin(theta),
                u
            )
            self.nodes.append(p)

    # calculates the radius needed for the requested average degree in a unit sphere
    def _getRadiusForAverageDegree(self):
        self.node_r = math.sqrt((self.avg_deg + 0.0)/self.num_nodes)*2

    # calculates the distance between two nodes (3D)
    def _distance(self, n, m):
        return math.sqrt((n[0] - m[0])**2+(n[1] - m[1])**2+(n[2] - m[2])**2)

    # gets benchmark setting for sphere
    def prepBenchmark(self, n):
        self.num_nodes = SPHERE_BENCHMARKS[n][0]
        self.avg_deg = SPHERE_BENCHMARKS[n][1]

    # public function for drawing graph, updates node limit if necessary
    def drawGraph(self, n_limit):
```

```
        self.n_limit = n_limit
        self._drawNodesAndEdges(self.nodes)

    # responsible for drawing nodes and edges in 3D space
    def _drawNodesAndEdges(self, node_list):
        # positions camera
        camera(self.canvas_width/2, self.canvas_height/2, self.canvas_width*-2, 0.5,0.5,0, 0,1,0)

        # updates rotation
        self.rot = (self.rot[0], self.rot[1]-math.pi/100, self.rot[2])

        background(self.color_bg)
        strokeWeight(2)
        stroke(self.color_fg)
        fill(self.color_fg)

        s = set(node_list)

        for n in node_list:
            pushMatrix()

            # sets new rotation
            rotateZ(self.rot[2])
            rotateY(-1*self.rot[1])

            # sets drawing origin to current node
            translate(n[0]*self.canvas_width, n[1]*self.canvas_height, n[2]*self.canvas_width)

            # places ellipse at origin
            ellipse(0, 0, 10, 10)

            # draw all edges
            if len(node_list) <= self.n_limit:
                for e_i in self.edges[n]:
                    if self.nodes[e_i] in s:
                        e = self.nodes[e_i]
                        # draws line from origin to neighboring node
                        line(0,0,0, (e[0] - n[0])*self.canvas_width, (e[1] - n[1])*self.canvas_height, (e[2] - n[2])*self.canvas_width)
            # draw edges for min degree node
            elif n == self.minDeg:
                stroke(0,self.color_fg,0)
                for e_i in self.edges[n]:
                    e = self.nodes[e_i]
                    # draws line from origin to neighboring node
                    line(0,0,0, (e[0] - n[0])*self.canvas_width, (e[1] - n[1])*self.canvas_height, (e[2] - n[2])*self.canvas_width)
                stroke(self.color_fg)
            # draw edges for max degree node
            elif n == self.maxDeg:
                stroke(0,0,self.color_fg)
                for e_i in self.edges[n]:
                    e = self.nodes[e_i]
                    # draws line from origin to neighboring node
                    line(0,0,0, (e[0] - n[0])*self.canvas_width, (e[1] - n[1])*self.canvas_height, (e[2] - n[2])*self.canvas_width)
                stroke(self.color_fg)

            popMatrix()

    # draw nodes as they are removed in smallest-last vertex ordering
    def drawSlvo(self):
        l = [self.nodes[i] for i in self.slvo[0:self.num_nodes - self.curr_node]]
        self._drawNodesAndEdges(l)

    # used to draw the graph with the nodes colored
```

```
    def drawColoring(self):
        l = [self.nodes[i] for i in self.slvo[0:self.curr_node]]
        self._drawNodesAndEdges(l)
        self._applyColors(self.slvo[0:self.curr_node])

    # places colors on the nodes
    def _applyColors(self, node_i_list):
        strokeWeight(2)

        num_colors = max(self.node_colors)

        for n_i in node_i_list:
            c = self.color_map[self.node_colors[n_i]]
            stroke(c[0], c[1], c[2])
            fill(c[0], c[1], c[2])

            pushMatrix()

            # sets new rotation
            rotateZ(self.rot[2])
            rotateY(-1*self.rot[1])

            # sets drawing origin to current node
            translate(self.nodes[n_i][0]*self.canvas_width, self.nodes[n_i][1]*self.canvas_height, self.nodes[n_i][2]*self.canvas_width)

            # places ellipse at origin
            ellipse(0, 0, 10, 10)

            popMatrix()

    # public function for pairing the independent sets and picking the largest backbones
    def generateBackbones(self):
        # uses base class method for generating backbones and meta data
        super(Sphere, self).generateBackbones()

        # calculate faces
        self.num_faces = self._countFaces(self.backbones_meta)

    # calcualtes the number of faces in the backbones of sphere topology
    def _countFaces(self, b_meta):
        # Euler's polyhedral formula
        # http://mathworld.wolfram.com/PolyhedralFormula.html
        return [2 - m[0] + m[1] for m in b_meta]

    # public function for drawing the color set pairs
    def drawPairs(self, mode=0):
        l_i = []
        if mode == 0:
            l_i = list(self.pairs[self.curr_pair])
        elif mode == 1:
            l_i = list(self.no_tails[self.curr_pair])
        elif mode == 2:
            l_i = list(self.major_comps[self.curr_pair])
        elif mode == 3:
            l_i = list(self.clean_pairs[self.curr_pair])

        l_n = [self.nodes[i] for i in l_i]
        self._drawNodesAndEdges(l_n)
        self._applyColors(l_i)

    # public function for drawing the backbones
    def drawBackbones(self):
        l_i = list(self.backbones[self.curr_backbone])
        l_n = [self.nodes[i] for i in l_i]
        self._drawNodesAndEdges(l_n)
        self._applyColors(l_i)
```