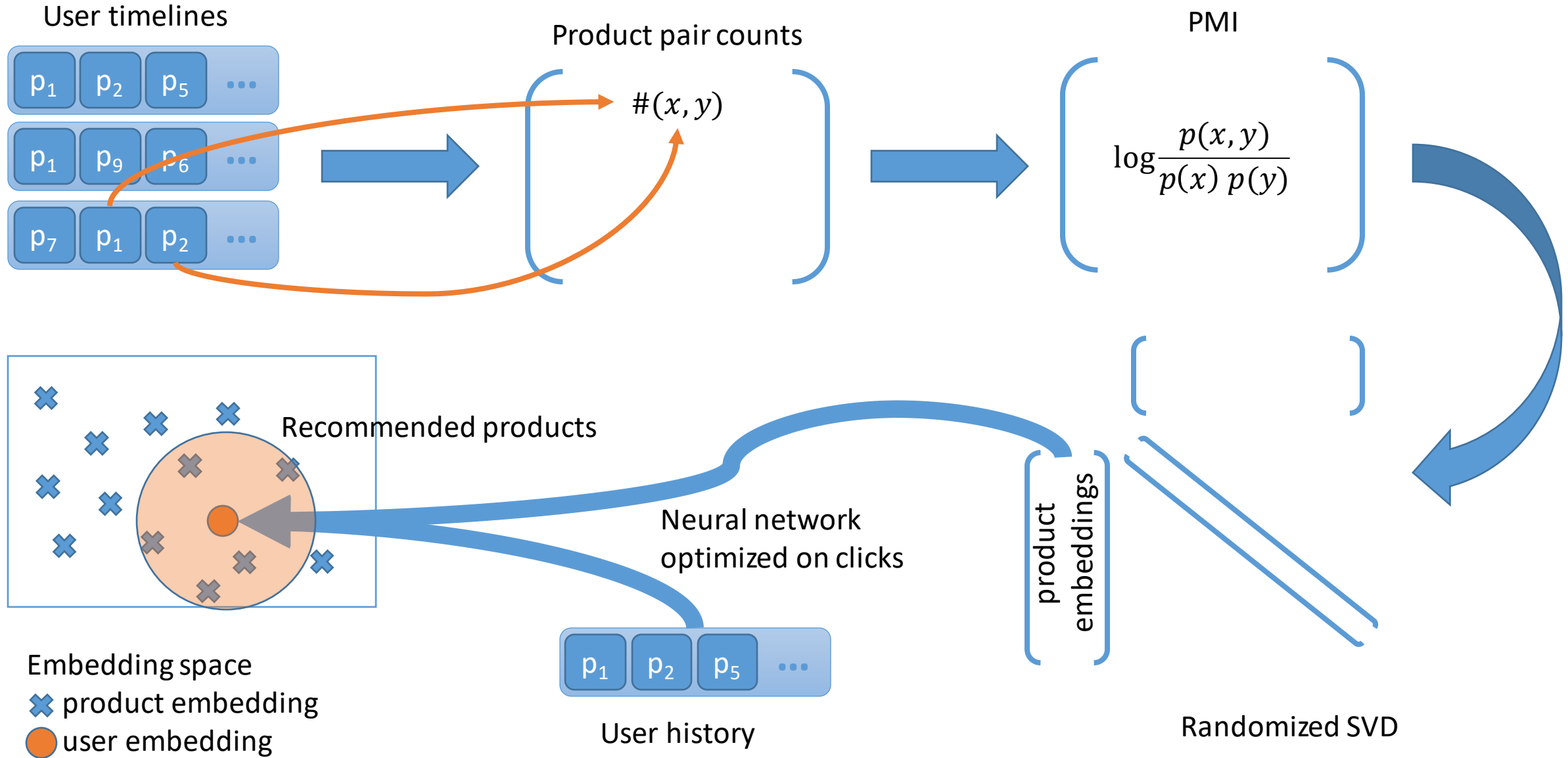


Approximate Nearest Neighbors Algorithms



David Diebold
Martin Bompairé

Where did we stop ?



Outline

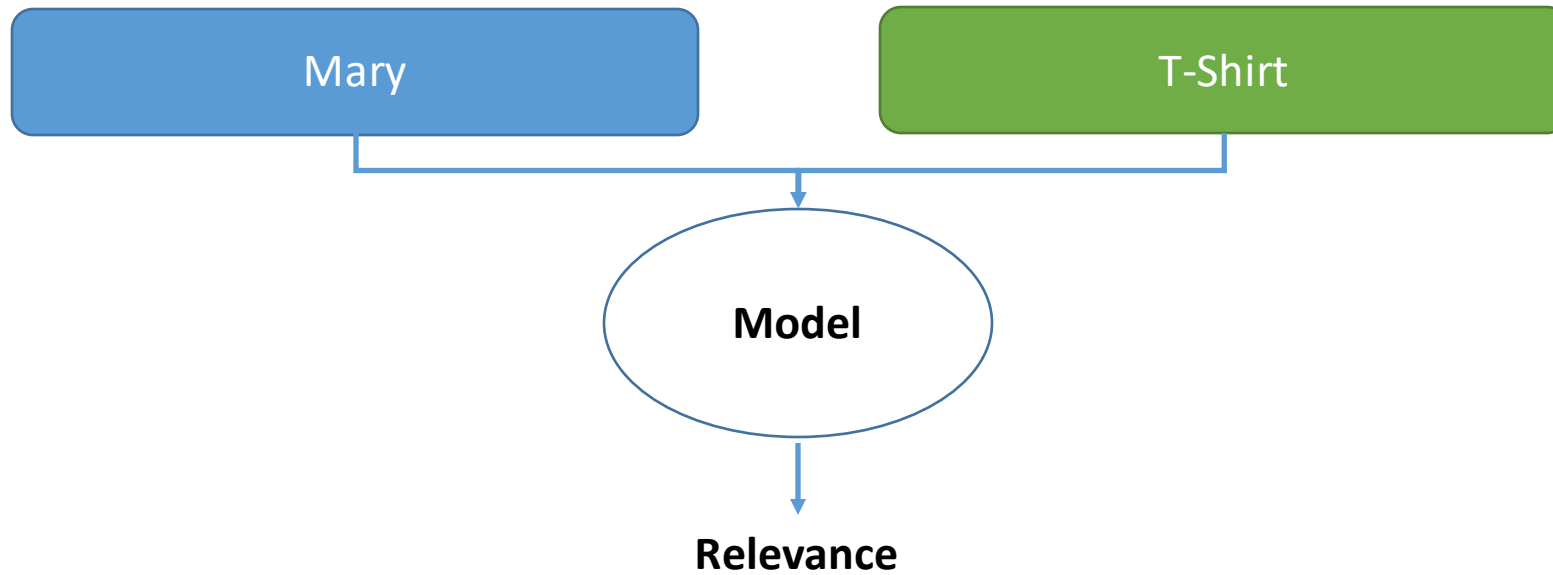
Motivation

Overview of KNN algorithms

Hnsw deep dive

Quantization

Problem Statement (1/2)



Problem Statement (2/2)



Naive solution : pooling

Need to work on a shortlist of candidates, picked carefully : Pooling

Similarities

Complementarities

Best of per
categories

...

Random



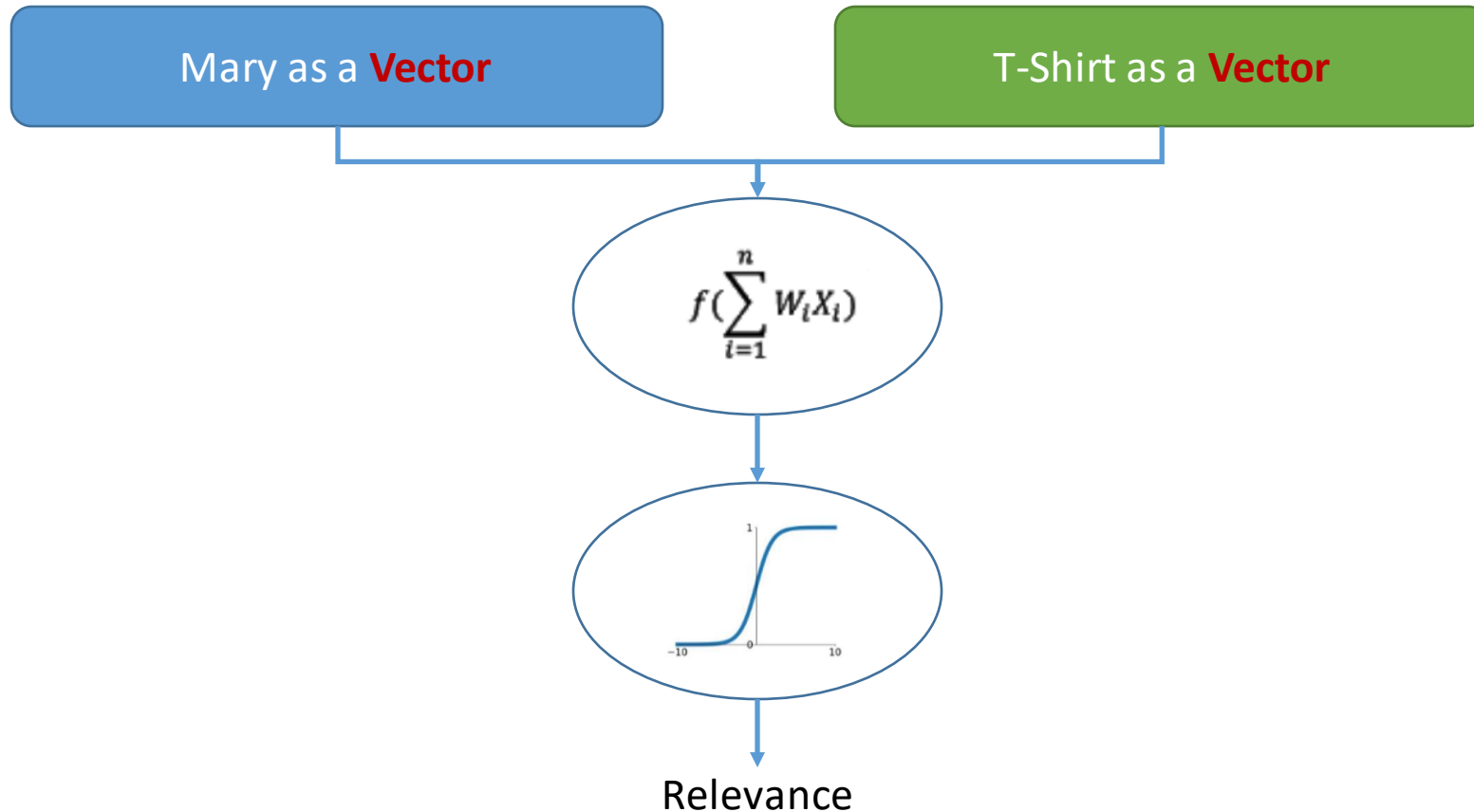
Not very great

We may miss interesting items

We may evaluate some items
for nothing

When adding a new source
sorting against a given score,
we need to pass this new kind
of score to the oracle (new
dimension)

Other solution – Representation learning



Finding best item becomes a nearest neighbor problem

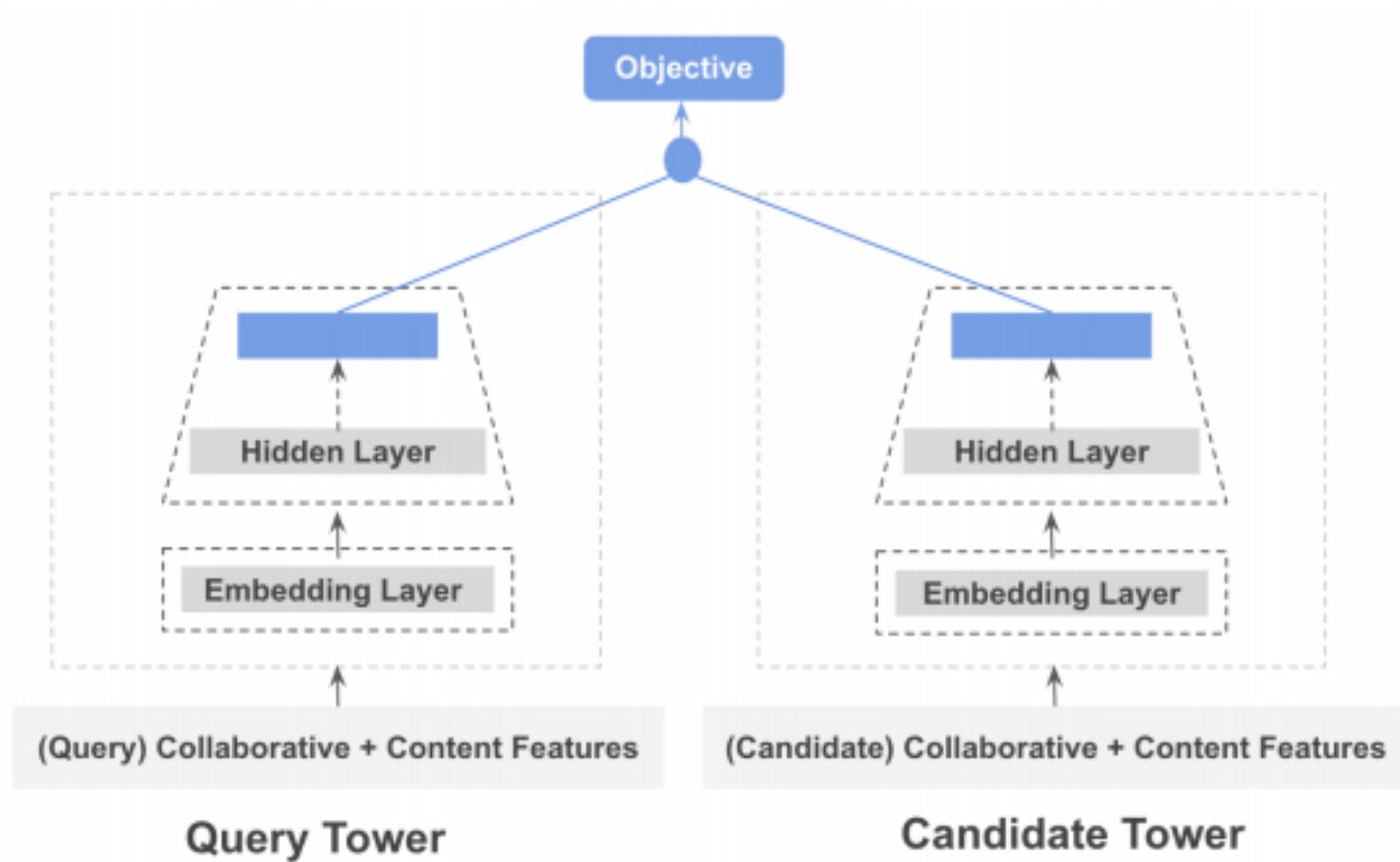
Logistic Regression : $f(x, y) = \frac{1}{1 + e^{-x \cdot y}}$

We want to maximize vectors dot product

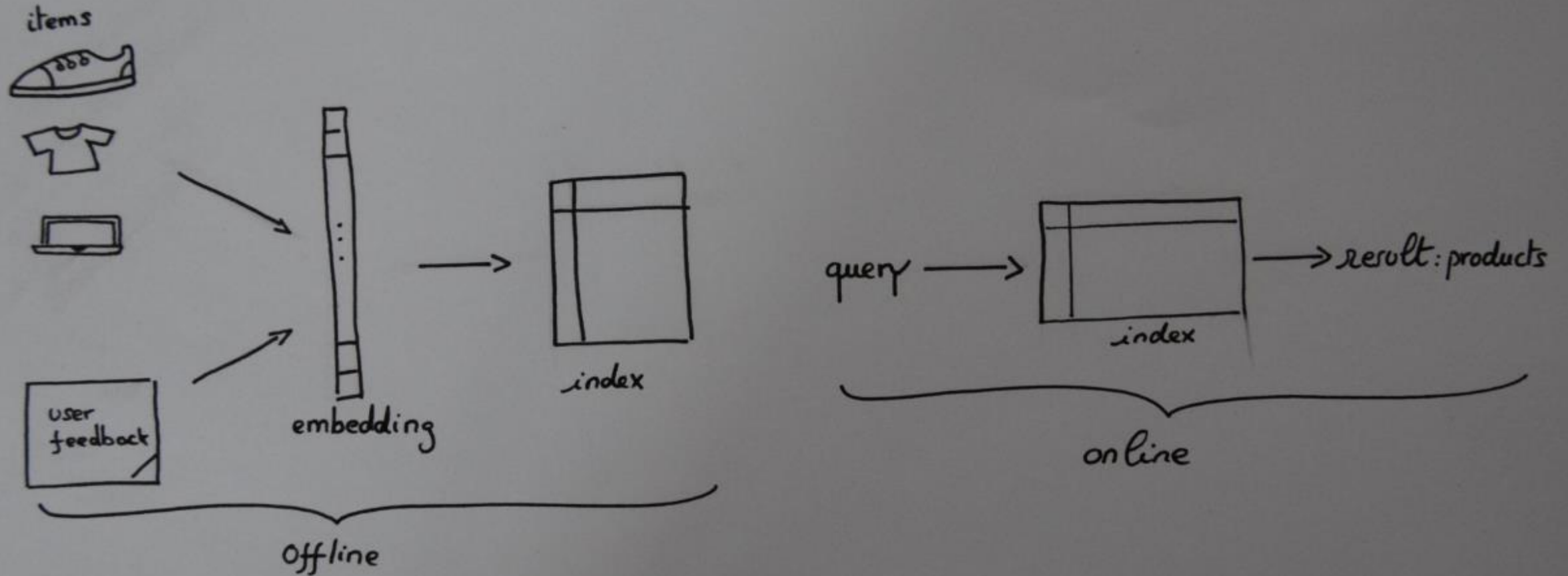
If each vector is normalized we have : $d(x, y) = \|x\| + \|y\| - 2x \cdot y$

We want to minimize a distance.

Logistic Regression can be replaced too !

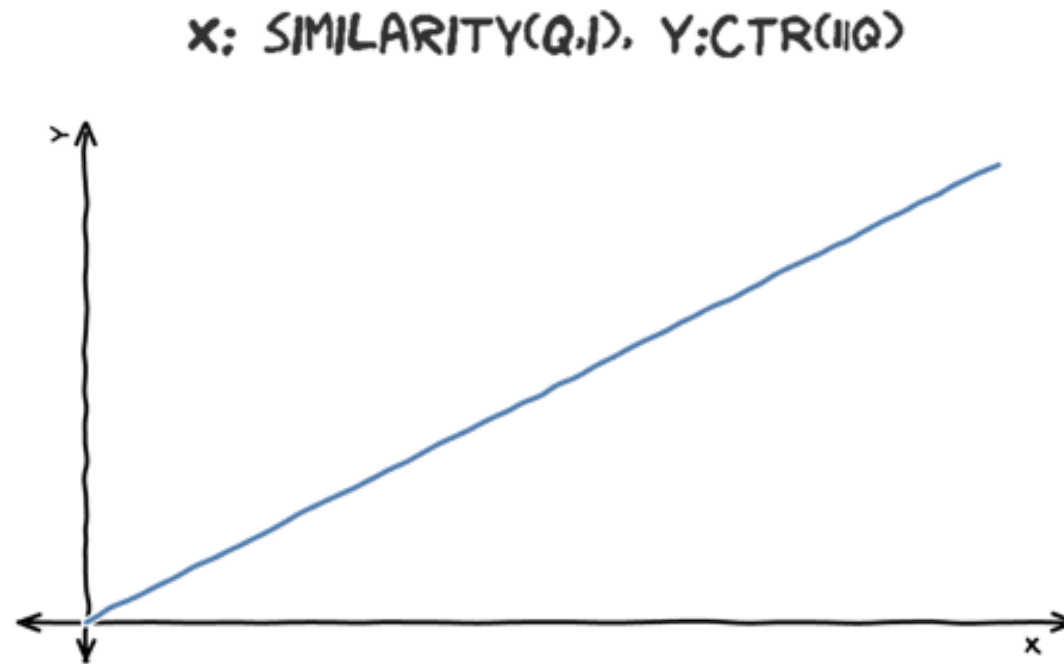


Recommendation System Pipeline



Cold start problem – new user, with particular set of item affinities

We can learn user representation model, such that $\langle \text{user} | \text{item} \rangle$ is proportionnal to user-item relevance.



Brute-force complexity is linear ☹️

Compare the key item with all existing ones

We want something better !

Overview of KNN methods

Tree based

Hash based

Graph based

Hierarchical Trees

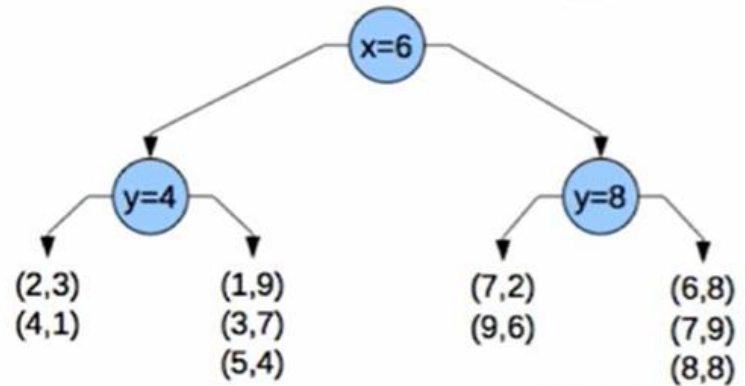
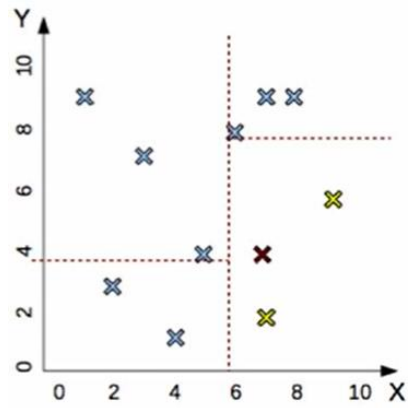


KD Tree

Construction

While cell contains
'too many' items:

- Pick random axis
- Find median
- Split cell

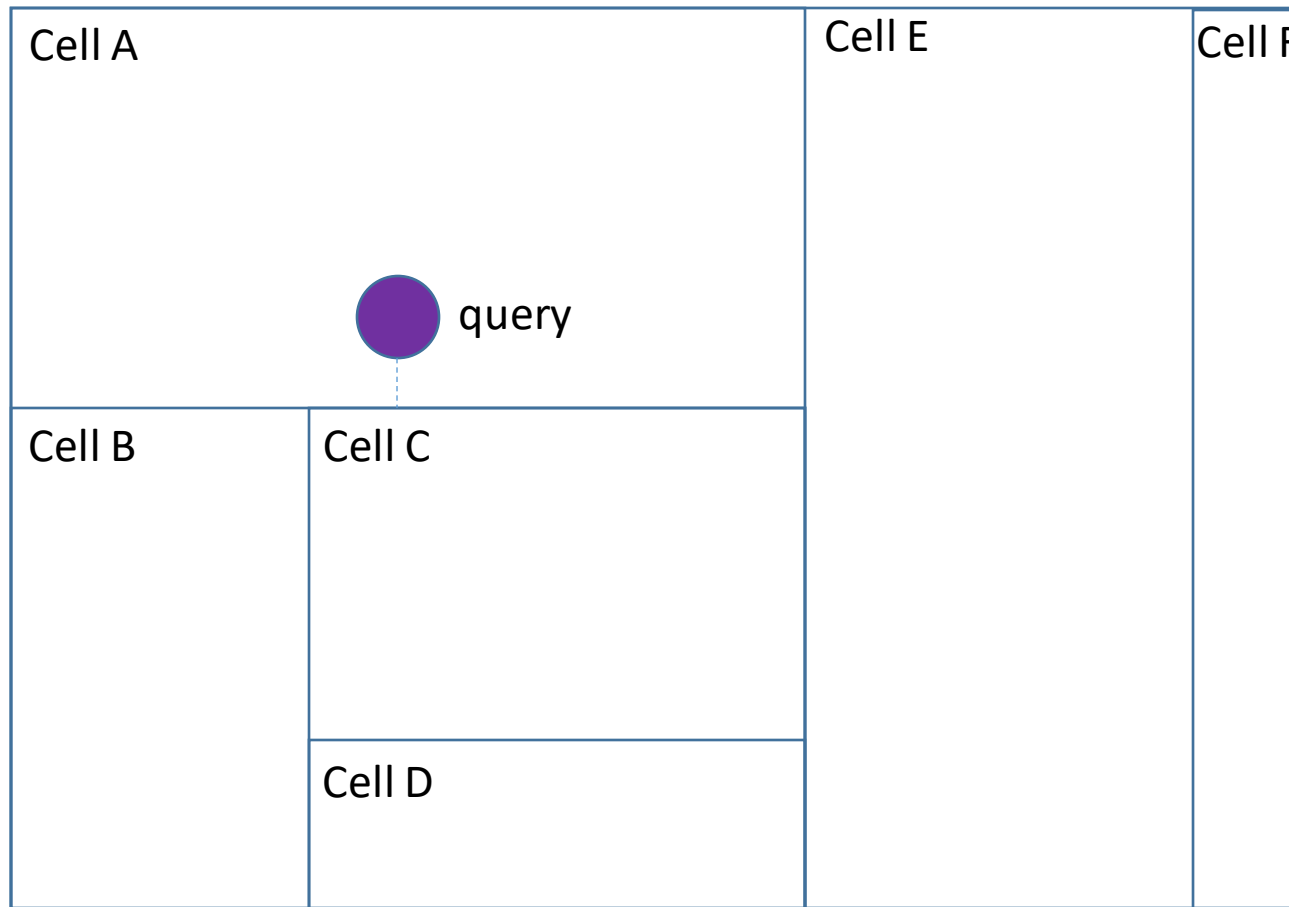


Search

Naïve way: Walk the tree, then look for the nearest neighbor inside the leaf.

No guarantee to find THE nearest neighbor.

KD Tree – Exact search



Ideas

It is fast to compute min distance between cell and queried point.

Don't explore a cell when we know it's min distance is larger than nearest point found so far.

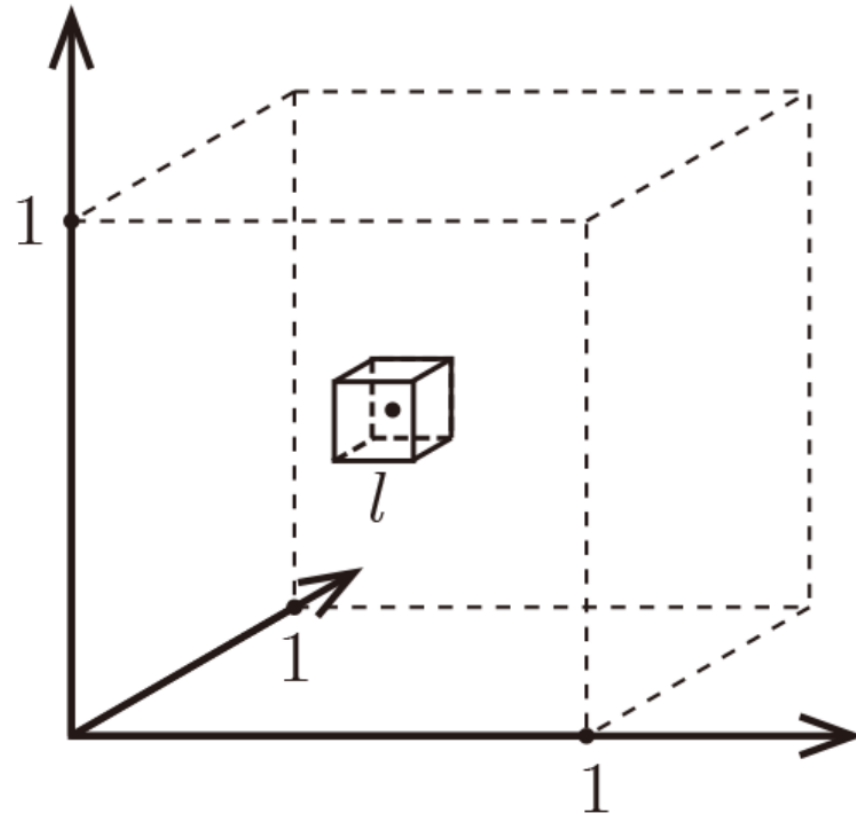
We use branch-and-bound to find nearest neighbors.

Start with nearest cells first to increase chances to minimize nearest point distance early.

Can be made approximate with some guarantees. Multiply $d(\text{query}, \text{cell})$ by $(1 + \text{Epsilon})$

Still, not efficient with high dimension vector

Curse of dimensionality



N points placed uniformly inside the space.

Let ℓ be the edge length of the smallest hyper-cube that contains all k -nearest neighbor of a test point.

Then $\ell^d \approx \frac{k}{n}$ and $\ell \approx \left(\frac{k}{n}\right)^{1/d}$. If $n = 1000$, how big is ℓ ?

Annoy

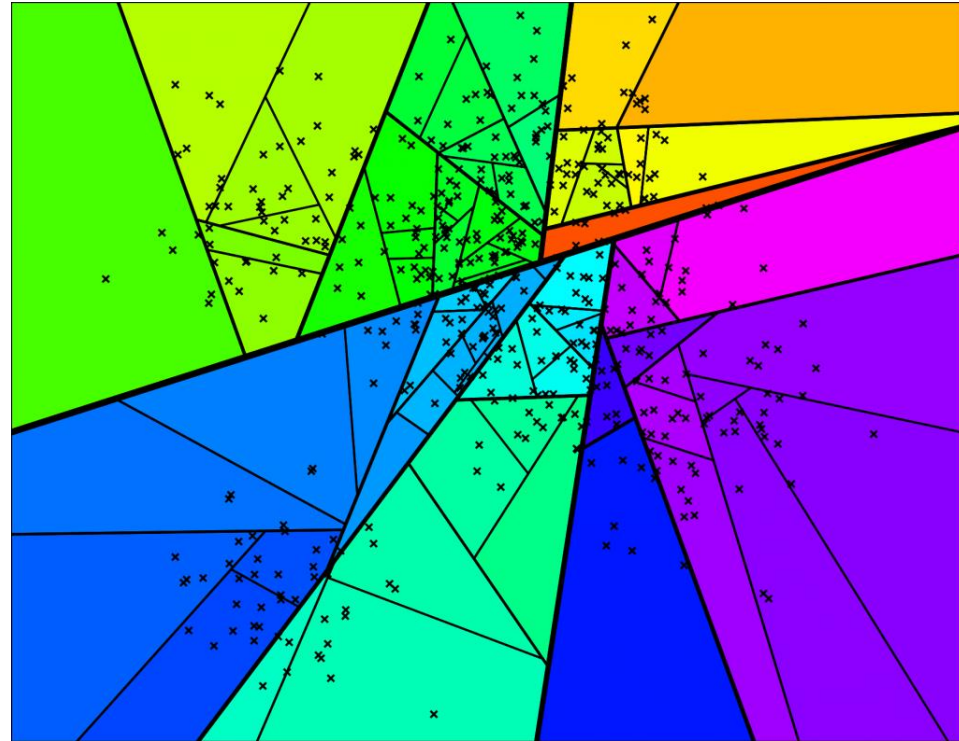
Construction

While cell contains
'too many' items:

- Pick two random points in cell
- Split along plane equidistant to the points

Create a Forest !

Great performances !



Search

Branch and bound
with additional tricks.

Search on all trees at
the same time using
same priority queue.

Approximate: don't
explore the other
side if it's *too* far.
Threshold can be
changed during the
search.

Hash based

Locality-sensitive hashing

Here we want to **maximize** collisions.

1. $\Pr(h(a) == h(b))$ is high if a and b are near
2. $\Pr(h(a) == h(b))$ is low if a and b are far
3. Time complexity to identify close objects is sub-linear.

Locality-sensitive hashing

Example : bit sampling

$a = (0 \ 0 \ 0 \ 1 \ 1 \ 1 \ 1)$ et $b = (1 \ 1 \ 0 \ 1 \ 0 \ 1 \ 1)$

$$d = 1 + 1 + 0 + 0 + 1 + 0 + 0 = 3$$

Hashing functions : $h_i : x \rightarrow x[i]$ for i in $[0, n-1]$

$$P(h(a) = h(b)) = 1 - d/n$$

Then we can combine hashing functions to achieve better accuracy.

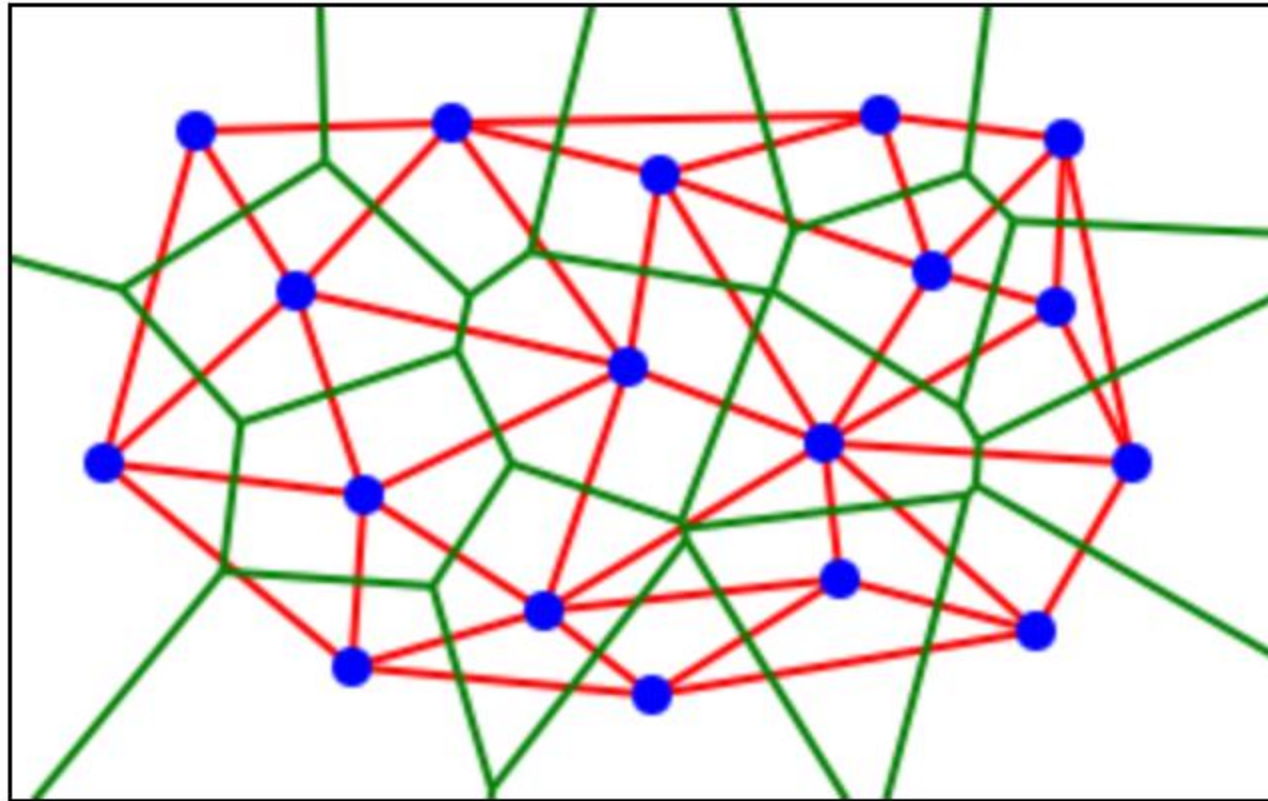
Locality-sensitive hashing

Other technique : random projection

Graph



Delaunay Graph (<1975)



item

Voronoi diagram

Delaunay graph

All points within a Voronoi tile have contained blue point as their nearest neighbor.
Don't expect to have $\text{Log}(N)$ by navigating the graph naïvely ! Needs post processing
Does not work in high dimension space ! (too many connections, too slow)

NSW (Navigable **Small World**)

« All people are six or fewer social connections from each other »

First manuscript written in ~1950 in Paris Sorbonne and circulated a lots. Milgram took up experiments 20 years later : take two random persons and see if/how the origin can manage to contact the destination.

➔ **Small world property exists in the nature !**



NSW (Navigable **Small World**)

How do we create small world network ?

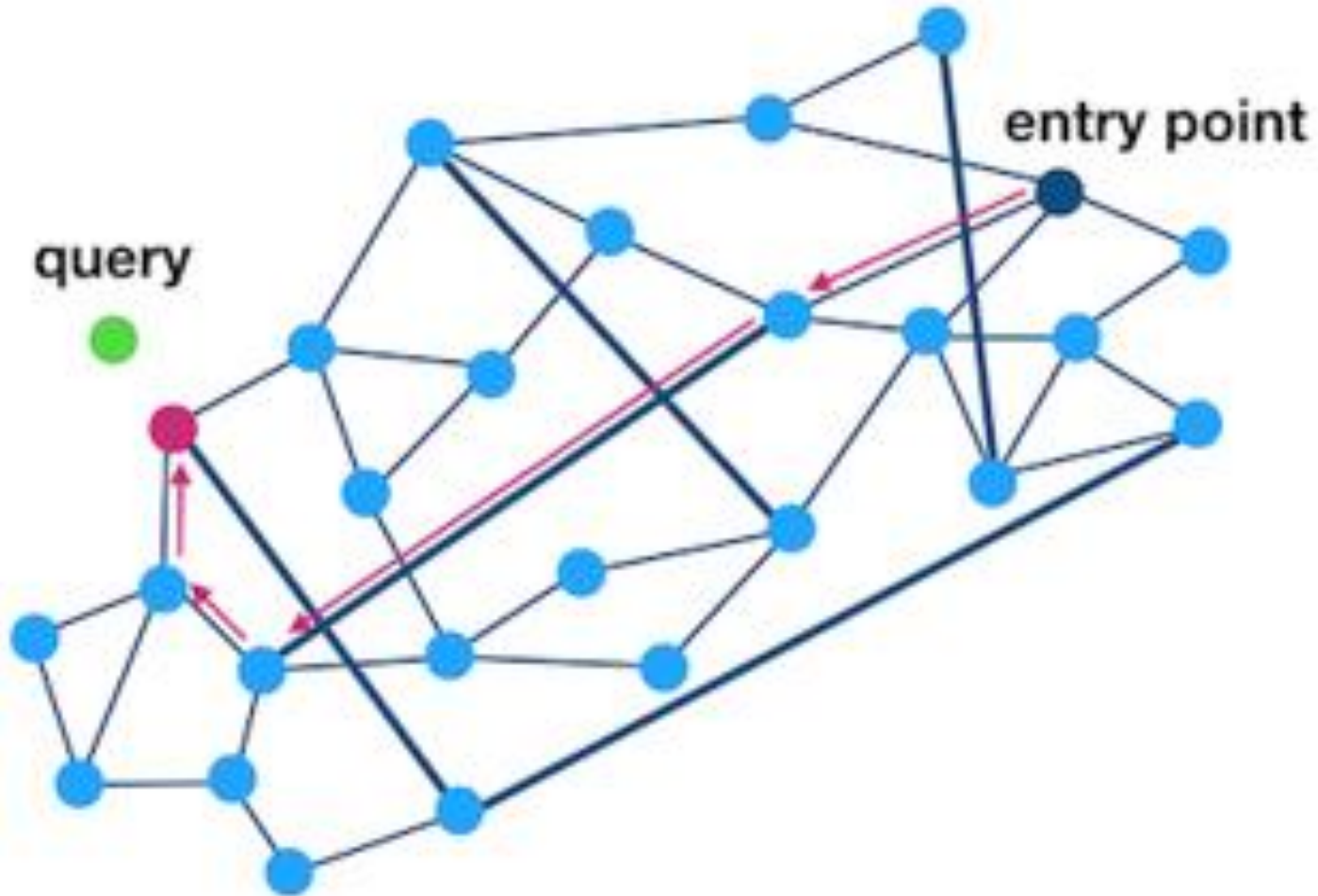
In social networks : Dual Phase Evolution

Local phase : people spend most of their time with people they know

Global phase : party/event/holidays, where people interact with people they don't know

Links are created during global phases, refined/destroyed in local phase.

Small world network : is it enough ?



NSW (Navigable Small World)

Navigable property guarantees that greedy algorithm is likely to stay on the shortest path.

Construction : try to enforce every node is connected to its M nearest neighbors.

Search : BFS (priorize neighbors nearest to the query), with stop condition : Worst neighbor found is better than best candidate to visit.

Works not too bad but performance degradations in low dimension / clustered graph.

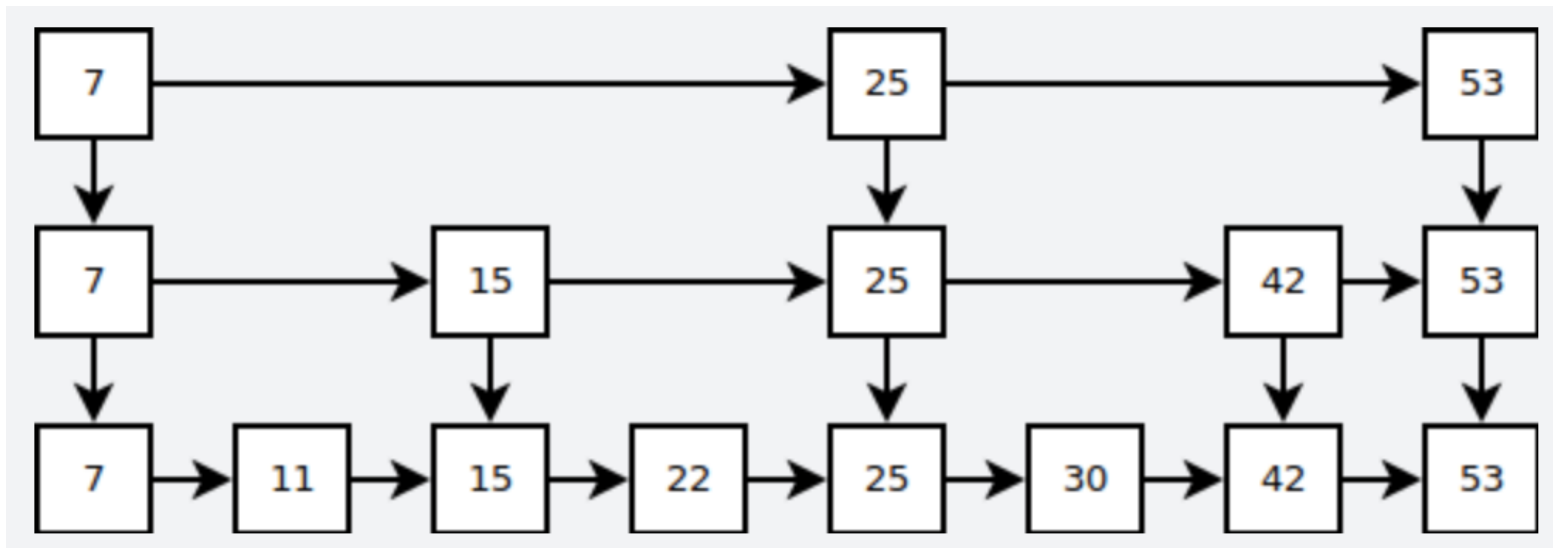
HNSW (Hierarchical NSW)

Idea

<https://arxiv.org/ftp/arxiv/papers/1603/1603.09320.pdf>

Add 'highways' to NSW

Same idea than skip-list and post-processing on Delaunay Graph !



HNSW - Construction

InsertNode(node, start)

Pick it's level L randomly (exponential decay law)

current \leftarrow start

search in upper layers

for layer in MaxL..L+1:

current \leftarrow KNNSearch(node, current, 1, layer)

fill lower layers

for layer in L..0:

neighbors \leftarrow KNNSearch(node, current, M2, layer)

toConnect \leftarrow select(node, neighbors, M)

for n in toConnect:

connect node and n

shrink connections of n

current \leftarrow neighbors

KNNSearch(query, start, k, layer)

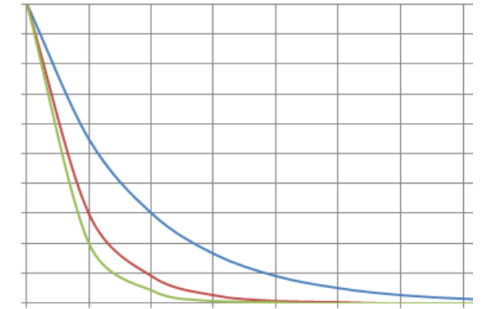
//returns k closest neighbor to query

BFS from starting point, explore points nearest to query first.

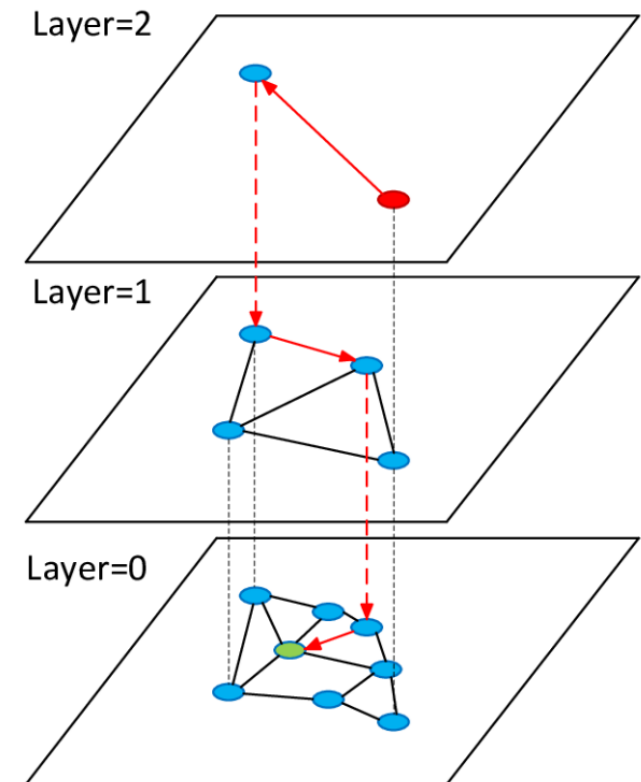
Stop condition

Worst neighbor found is better than best candidate to visit.

probability



layer



<https://arxiv.org/ftp/arxiv/papers/1603/1603.09320.pdf>











HNSW – Toy Example

ef_construction : 3
Mmax : 3

Current Layer : 0
Inserted Point Layer : 0











Legend	 Point inside layer	 Point being inserted	 Search waypoint	 Connection in layer
	 Point outside layer	 Point being queried	 Root	 Search path

Current Layer : 0
Inserted Point Layer : 0











Legend

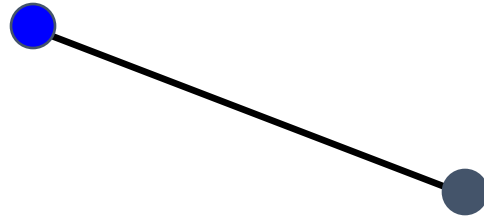
 Point inside layer	 Point being inserted	 Search waypoint	 Connection in layer
 Point outside layer	 Point being queried	 Root	 Search path









Current Layer : 0
Inserted Point Layer : 0

c

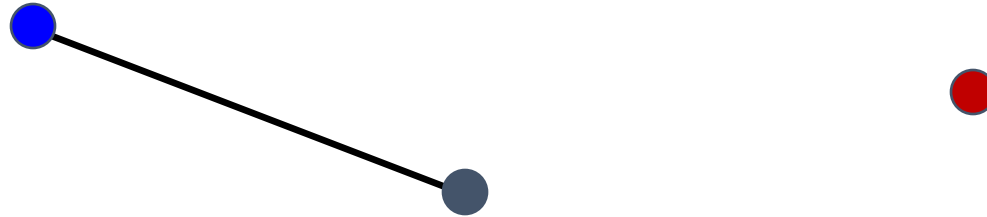
Legend	 Point inside layer	 Point being inserted	 Search waypoint	 Connection in layer
	 Point outside layer	 Point being queried	 Root	 Search path









Current Layer : 0
Inserted Point Layer : 0



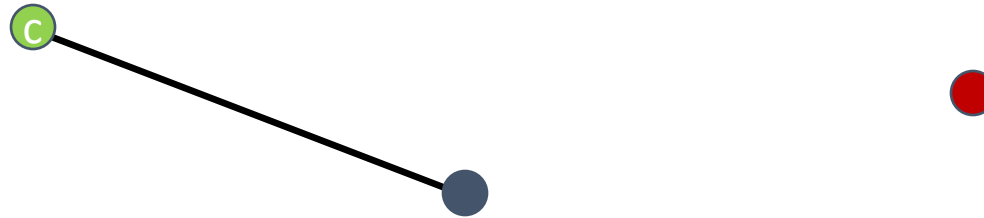
Legend	 Point inside layer	 Point being inserted	 Search waypoint	 Connection in layer
	 Point outside layer	 Point being queried	 Root	 Search path









Current Layer : 0
Inserted Point Layer : 0



Legend	 Point inside layer	 Point being inserted	 Search waypoint	 Connection in layer
	 Point outside layer	 Point being queried	 Root	 Search path









Current Layer : 0
Inserted Point Layer : 0



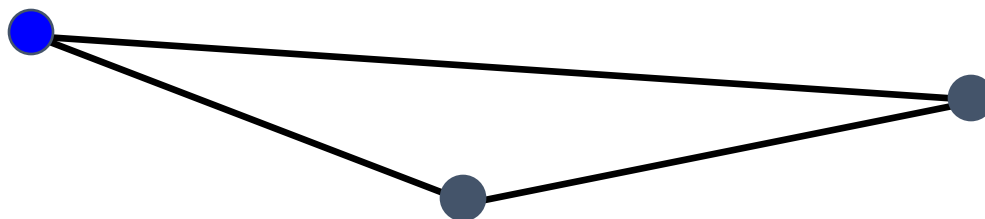
Legend	 Point inside layer	 Point being inserted	 Search waypoint	 Connection in layer
	 Point outside layer	 Point being queried	 Root	 Search path









Current Layer : 0
Inserted Point Layer : 0



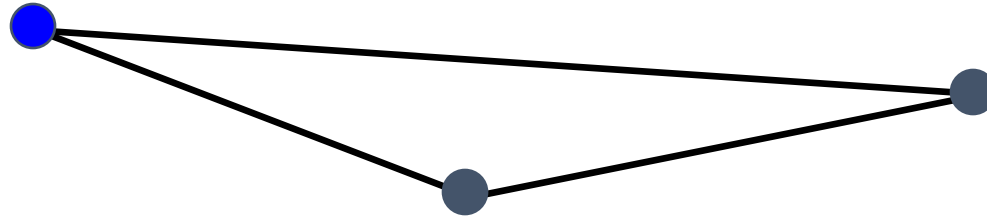
- Legend
- | | | | |
|---|--|---|---|
|  Point inside layer |  Point being inserted |  Search waypoint |  Connection in layer |
|  Point outside layer |  Point being queried |  Root |  Search path |









Current Layer : 0
Inserted Point Layer : 0



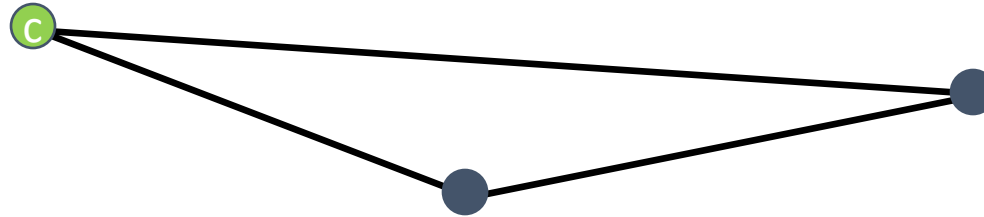
Legend	 Point inside layer	 Point being inserted	 Search waypoint	 Connection in layer
	 Point outside layer	 Point being queried	 Root	 Search path








Current Layer : 0
Inserted Point Layer : 0



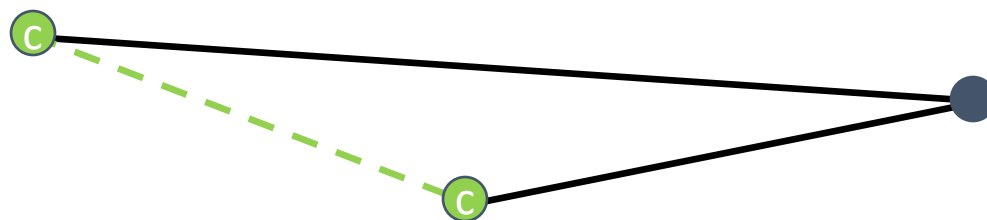
- Legend
- | | | | |
|---|--|---|---|
|  Point inside layer |  Point being inserted |  Search waypoint |  Connection in layer |
|  Point outside layer |  Point being queried |  Root |  Search path |

Current Layer : 0
Inserted Point Layer : 0



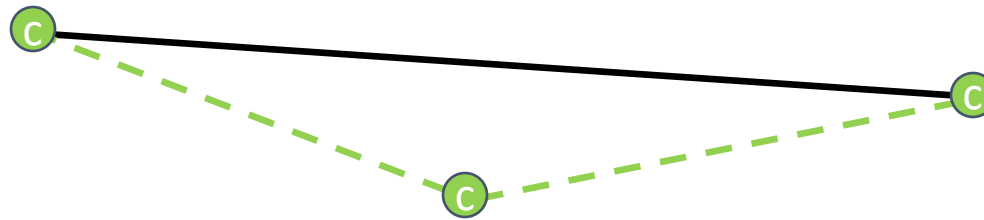
- | | | | | |
|--------|---|--|---|---|
| Legend |  Point inside layer |  Point being inserted |  Search waypoint |  Connection in layer |
| |  Point outside layer |  Point being queried |  Root | |









Current Layer : 0
Inserted Point Layer : 0



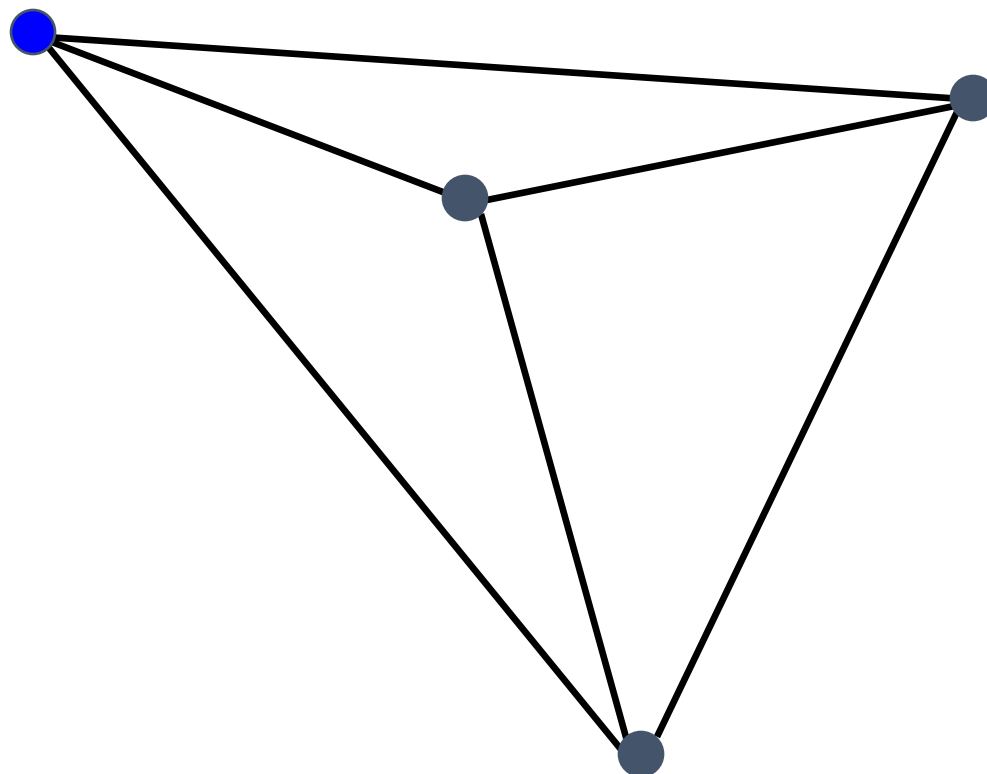
Legend	Point inside layer	Point being inserted	Search waypoint	Connection in layer
	Point outside layer	Point being queried	Root	Search path

Current Layer : 0
Inserted Point Layer : 0



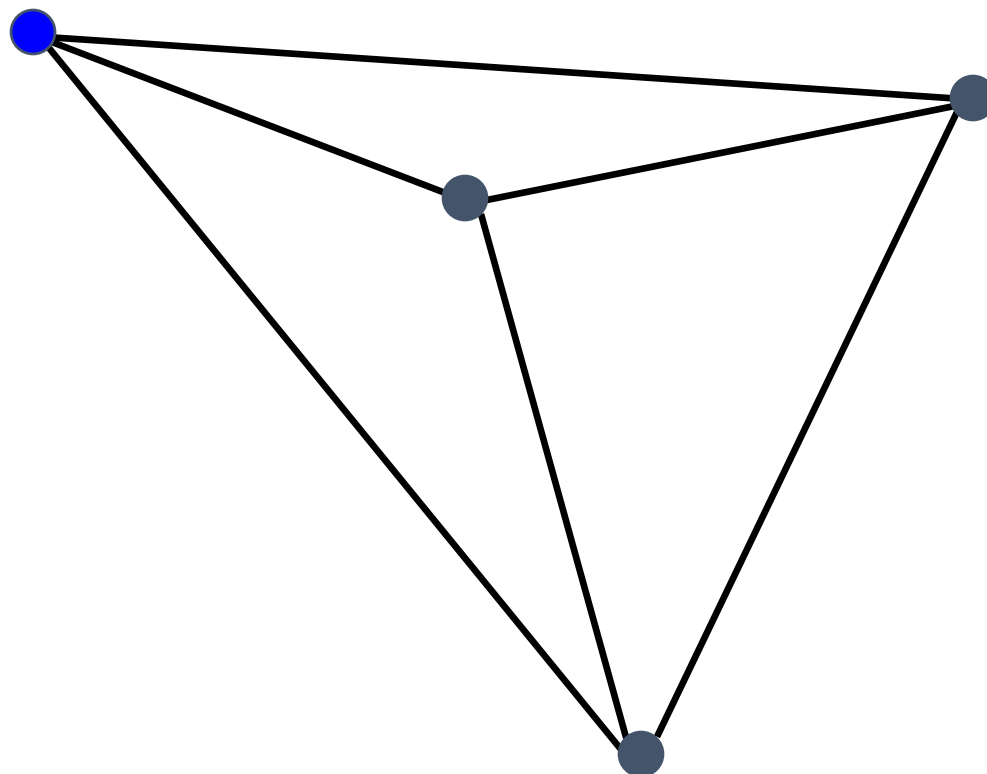
Legend	 Point inside layer	 Point being inserted	 Search waypoint	 Connection in layer
	 Point outside layer	 Point being queried	 Root	 Search path

Current Layer : 0
Inserted Point Layer : 0



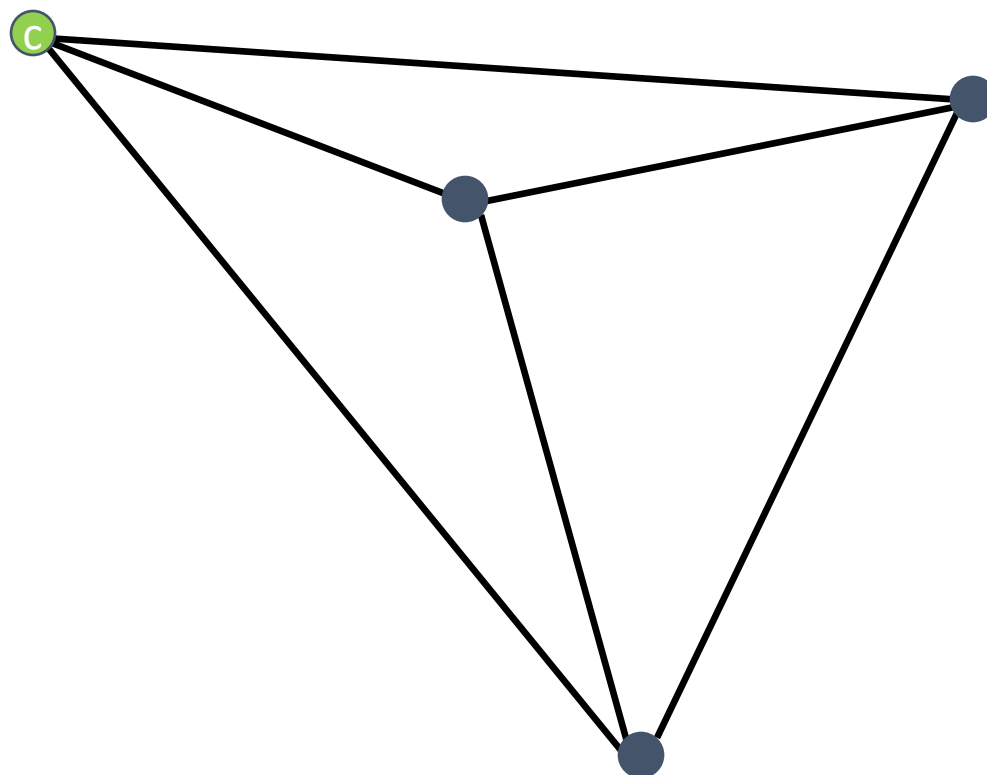
Legend	Point inside layer	Point being inserted	Search waypoint	Connection in layer
	Point outside layer	Point being queried	Root	Search path

Current Layer : 0
Inserted Point Layer : 0



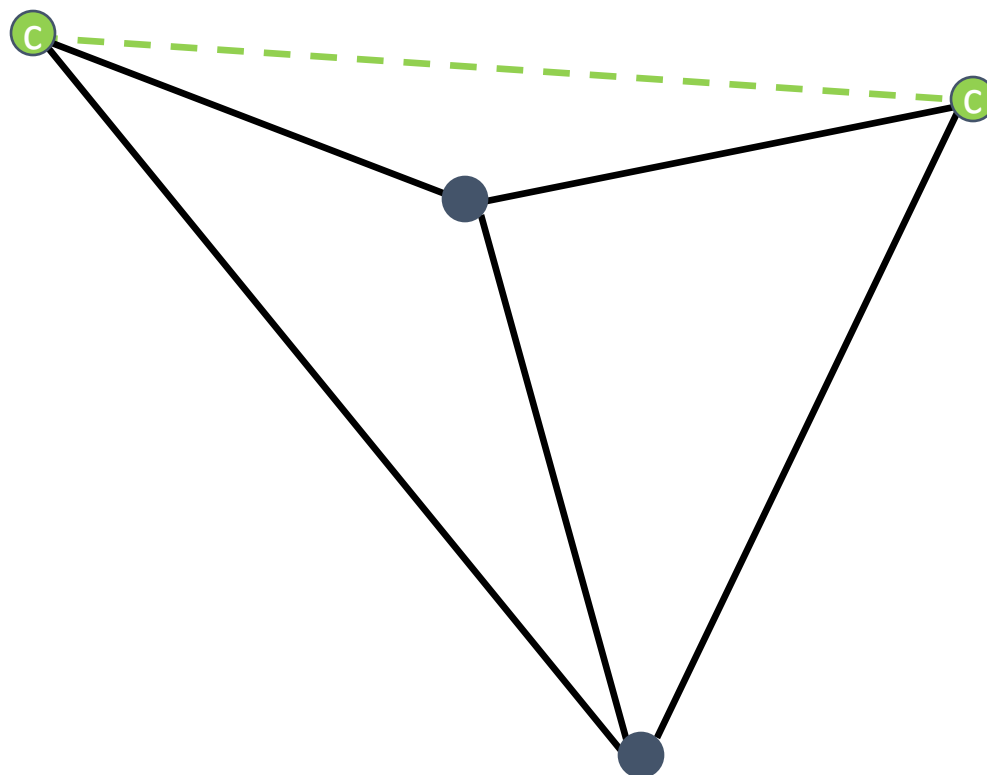
Legend	Point inside layer	Point being inserted	Search waypoint	Connection in layer
	Point outside layer	Point being queried	Root	Search path

Current Layer : 0
Inserted Point Layer : 0



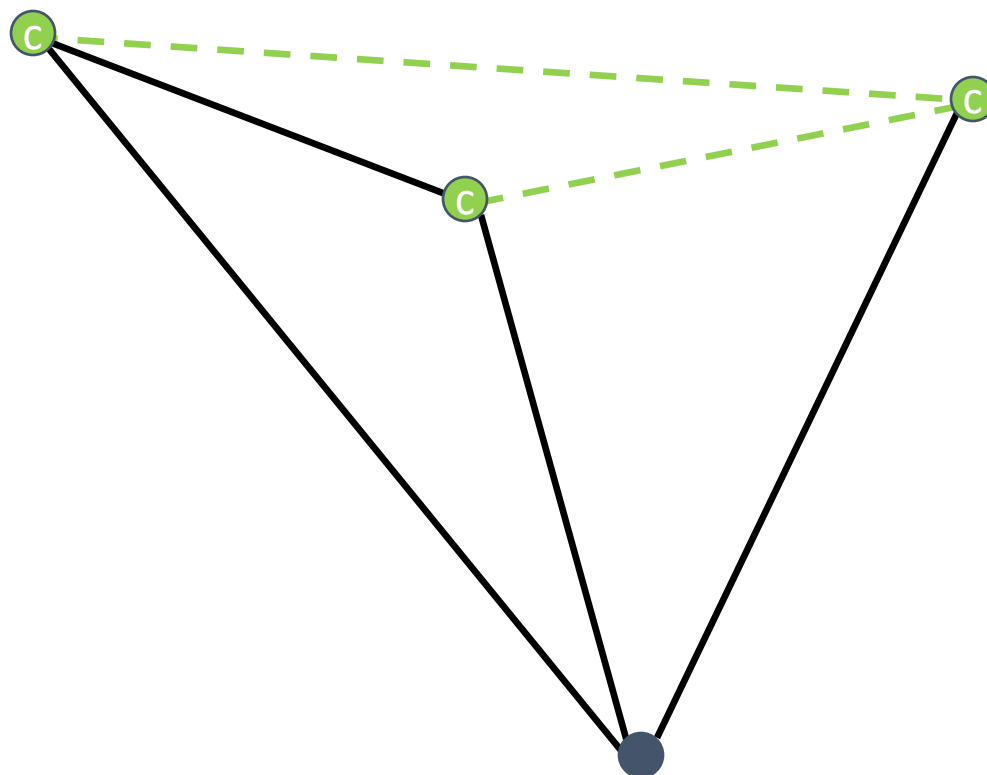
Legend	● Point inside layer	● Point being inserted	● Search waypoint	— Connection in layer
	○ Point outside layer	● Point being queried	● Root	- - - Search path

Current Layer : 0
Inserted Point Layer : 0



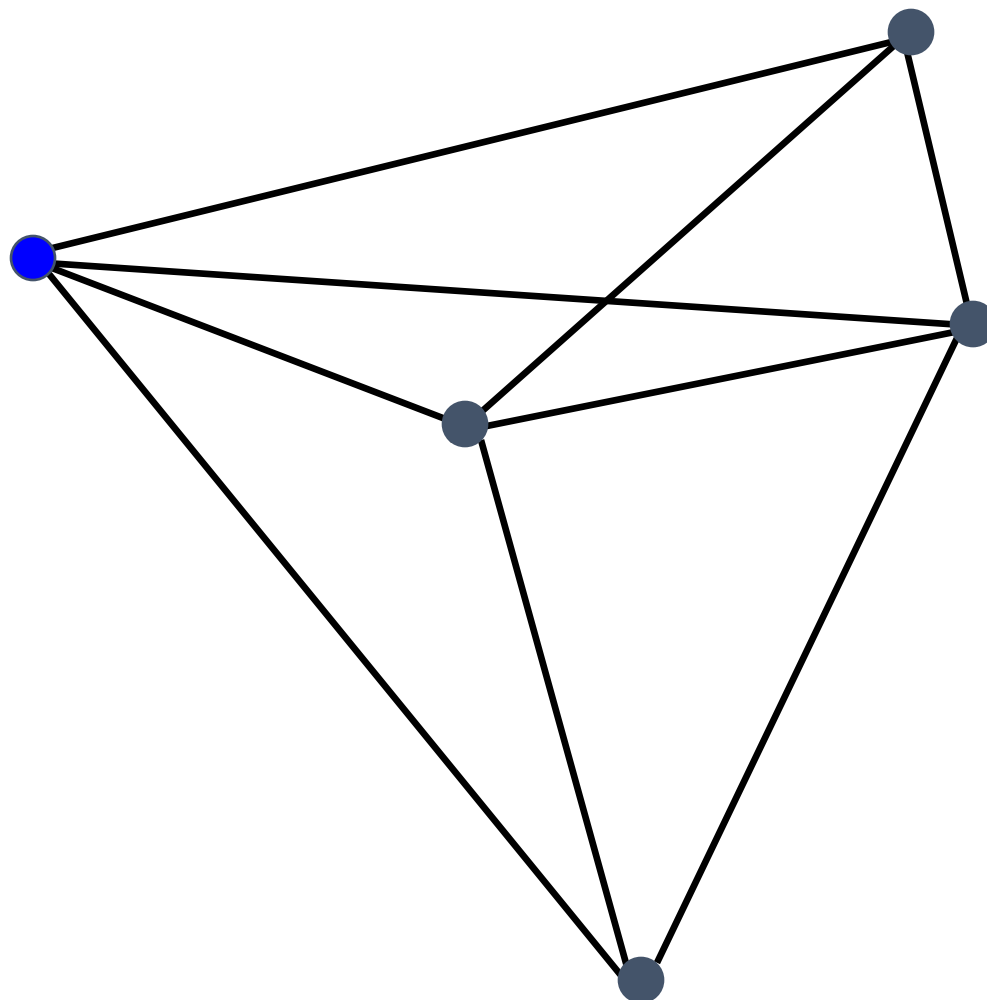
Legend	Point inside layer	Point being inserted	Search waypoint	Connection in layer
	Point outside layer	Point being queried	Root	Search path

Current Layer : 0
Inserted Point Layer : 0



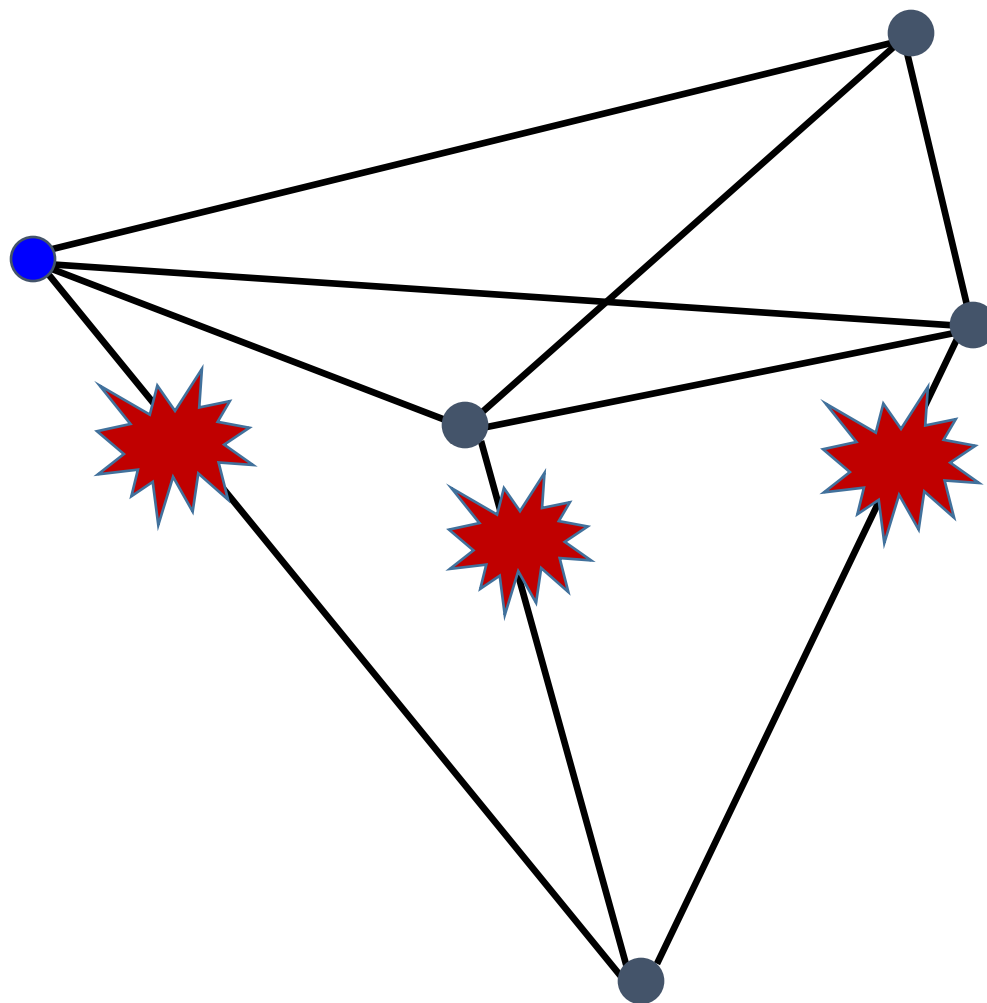
Legend	Point inside layer	Point being inserted	Search waypoint	Connection in layer
	Point outside layer	Point being queried	Root	Search path

Current Layer : 0
Inserted Point Layer : 0



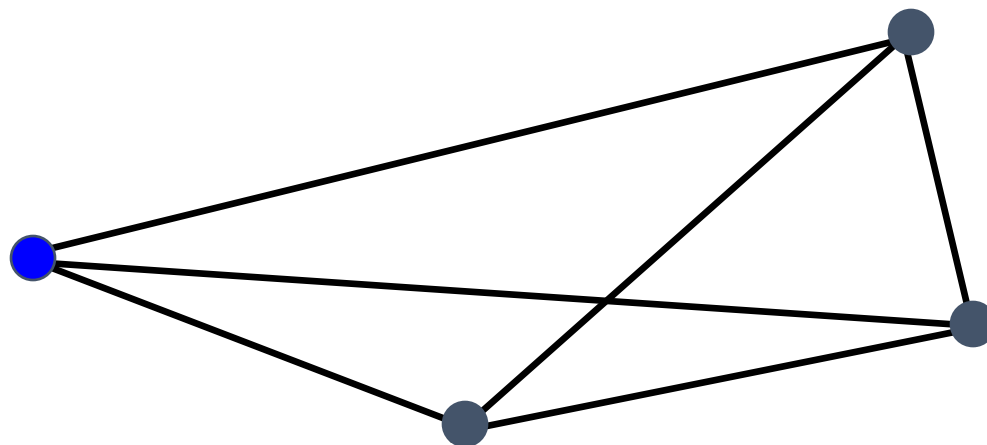
Legend	Point inside layer	Point being inserted	Search waypoint	Connection in layer
	Point outside layer	Point being queried	Root	Search path

Current Layer : 0
Inserted Point Layer : 0



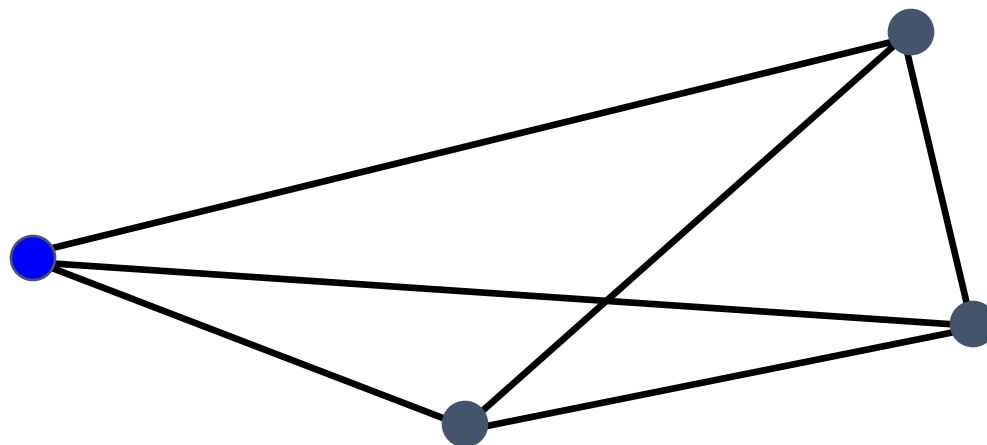
Legend	Point inside layer	Point being inserted	Search waypoint	Connection in layer
	Point outside layer	Point being queried	Root	Search path









Current Layer : 0
Inserted Point Layer : 0



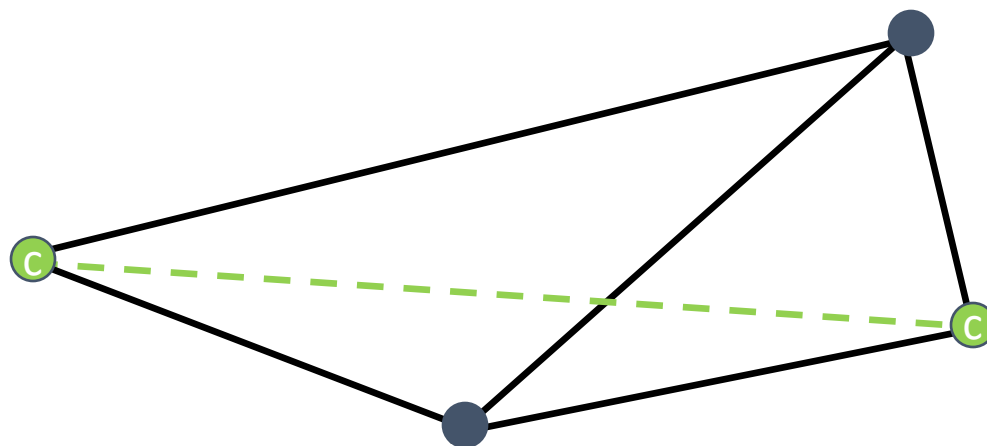
Legend	Point inside layer	Point being inserted	Search waypoint	Connection in layer
	Point outside layer	Point being queried	Root	Search path









Current Layer : 0
Inserted Point Layer : 1



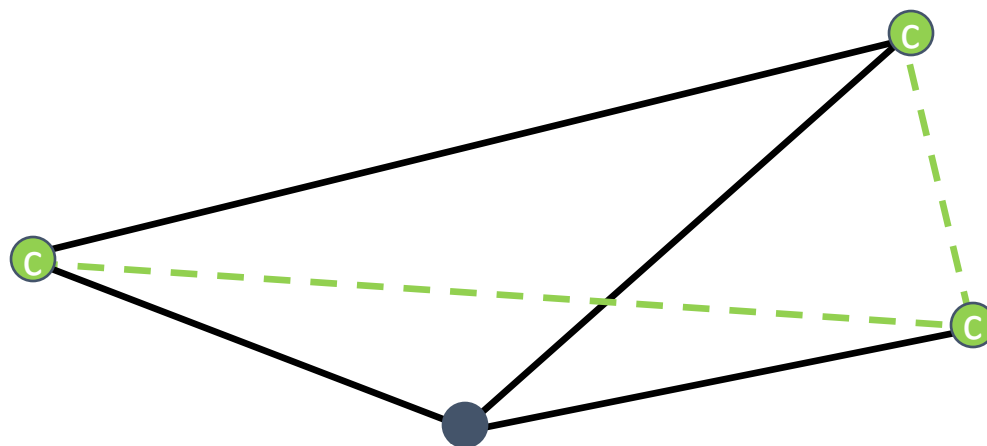
Legend	 Point inside layer	 Point being inserted	 Search waypoint	 Connection in layer
	 Point outside layer	 Point being queried	 Root	 Search path

Current Layer : 0
Inserted Point Layer : 1



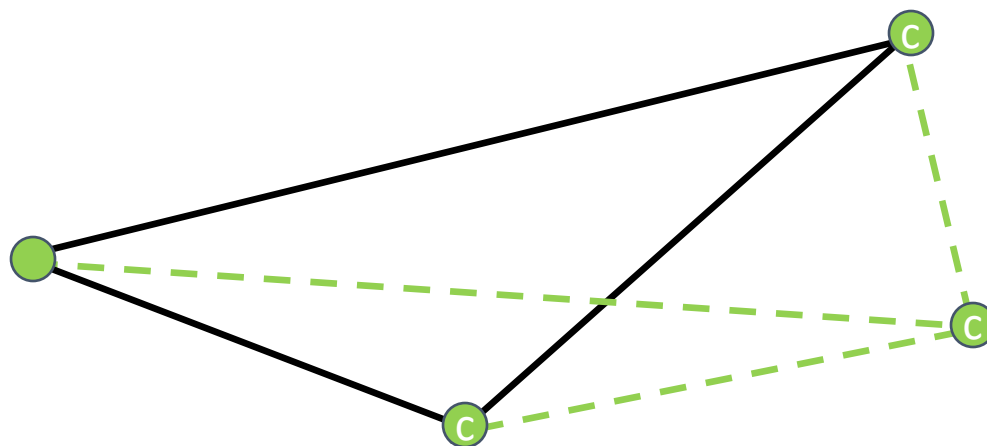
Legend	 Point inside layer	 Point being inserted	 Search waypoint	 Connection in layer
	 Point outside layer	 Point being queried	 Root	 Search path

Current Layer : 0
Inserted Point Layer : 1



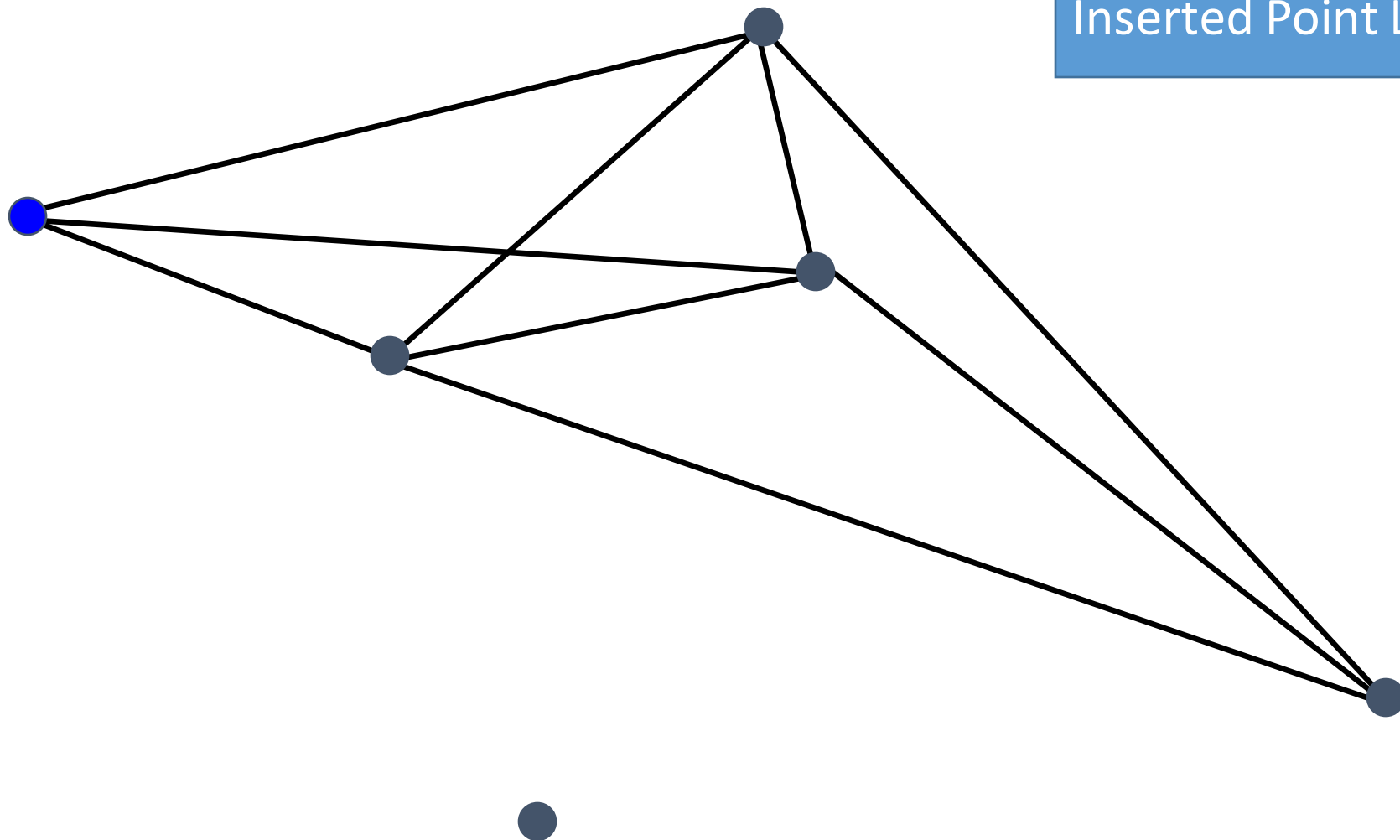
Legend	Point inside layer	Point being inserted	Search waypoint	Connection in layer
	Point outside layer	Point being queried	Root	Search path

Current Layer : 0
Inserted Point Layer : 1



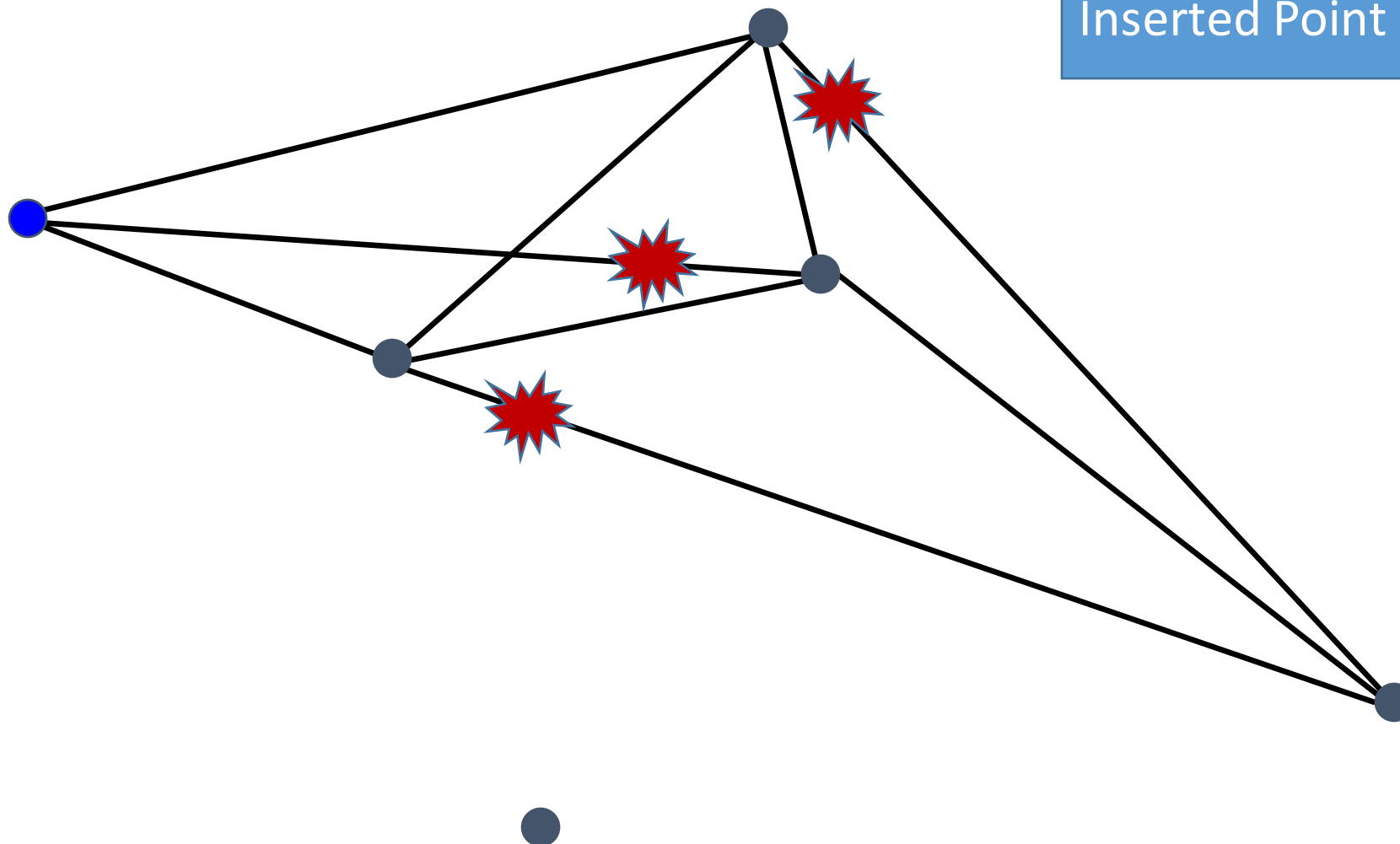
Legend	Point inside layer	Point being inserted	Search waypoint	Connection in layer
	Point outside layer	Point being queried	Root	Search path

Current Layer : 0
Inserted Point Layer : 1



Legend	Point inside layer	Point being inserted	Search waypoint	Connection in layer
	Point outside layer	Point being queried	Root	Search path

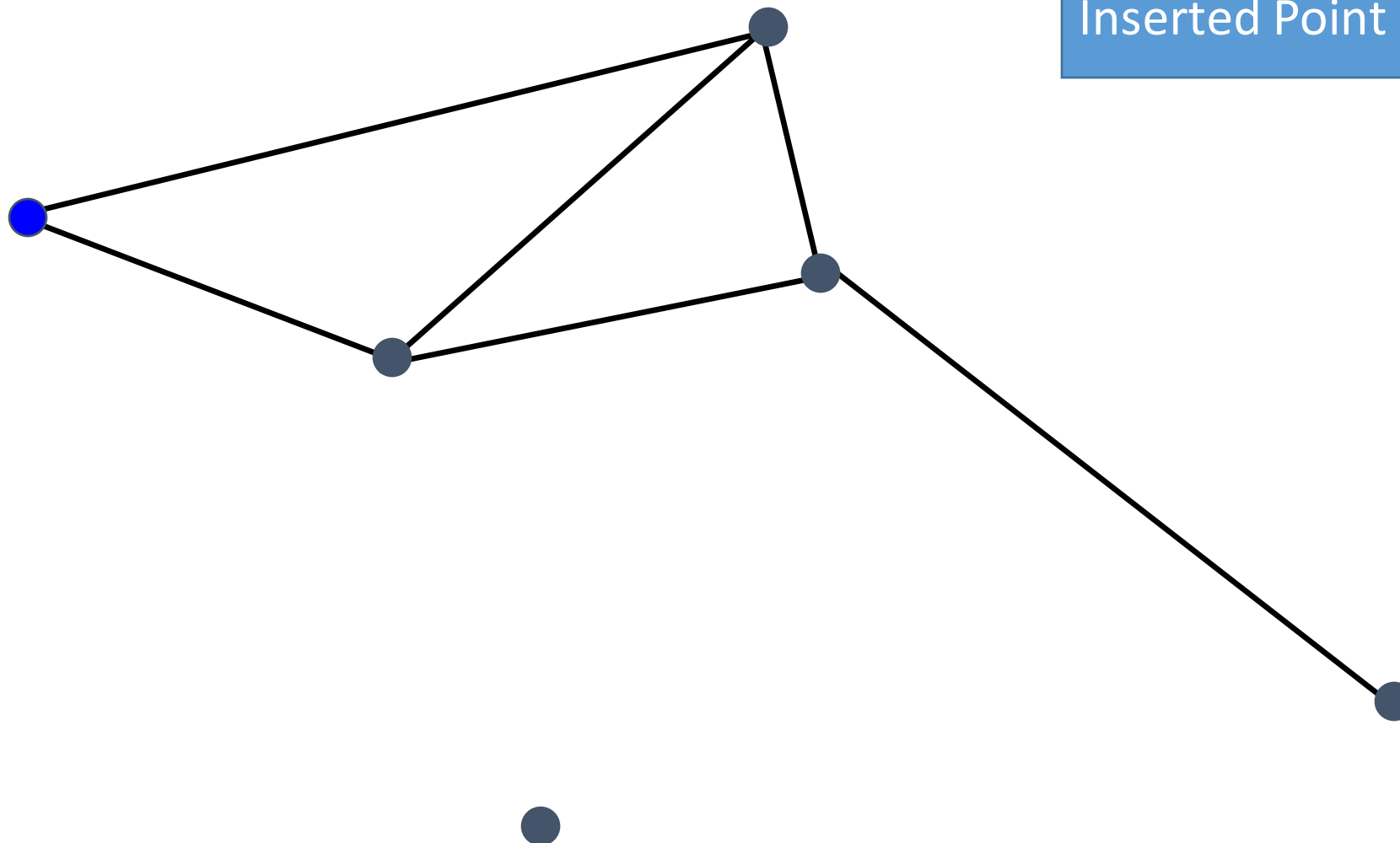
Current Layer : 0
Inserted Point Layer : 1



Legend

● Point inside layer	● Point being inserted	● Search waypoint	— Connection in layer
○ Point outside layer	● Point being queried	● Root	- - - Search path









Current Layer : 0
Inserted Point Layer : 1

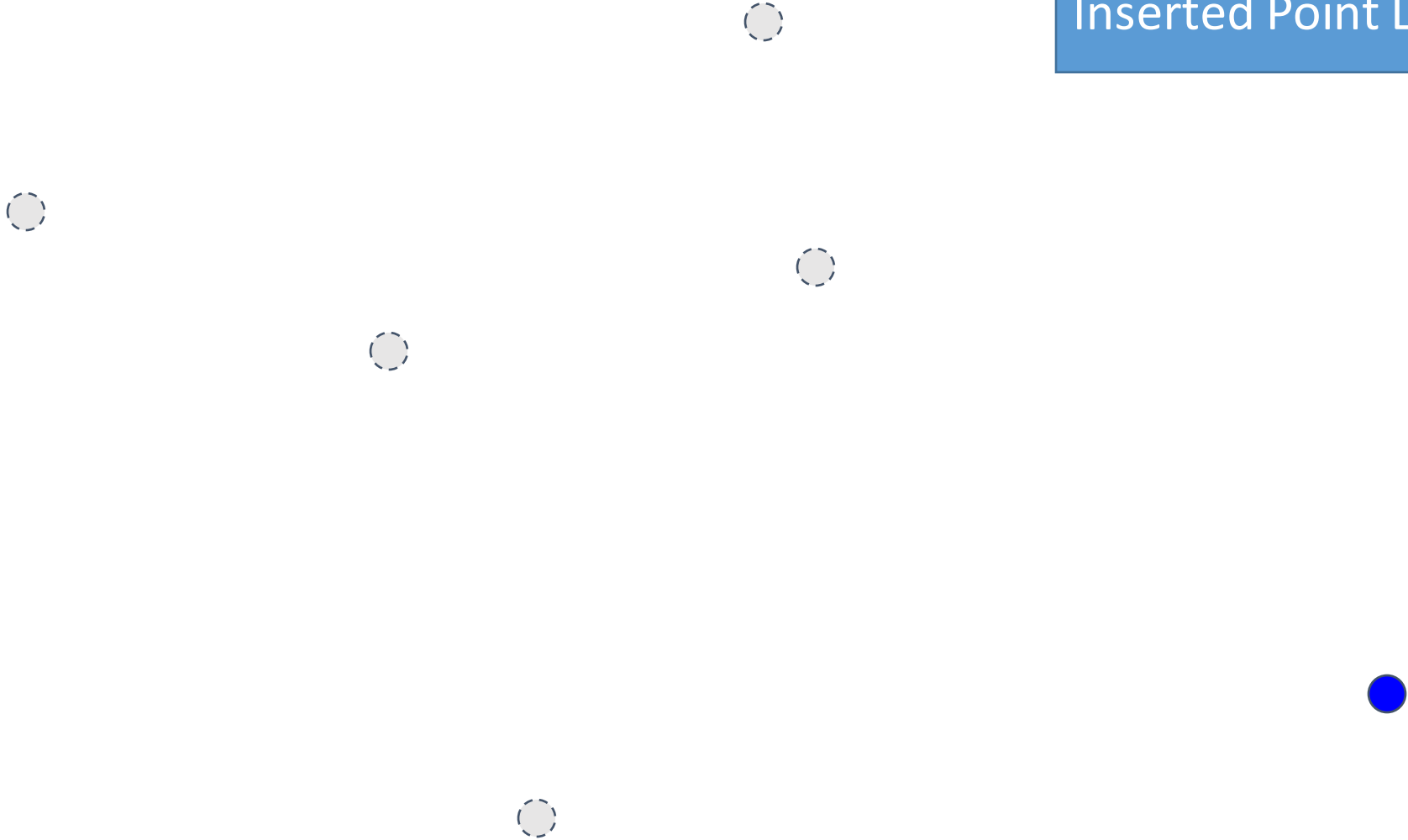


Legend	Point inside layer	Point being inserted	Search waypoint	Connection in layer
	Point outside layer	Point being queried	Root	Search path

Current Layer : 1
Inserted Point Layer : 1









Legend

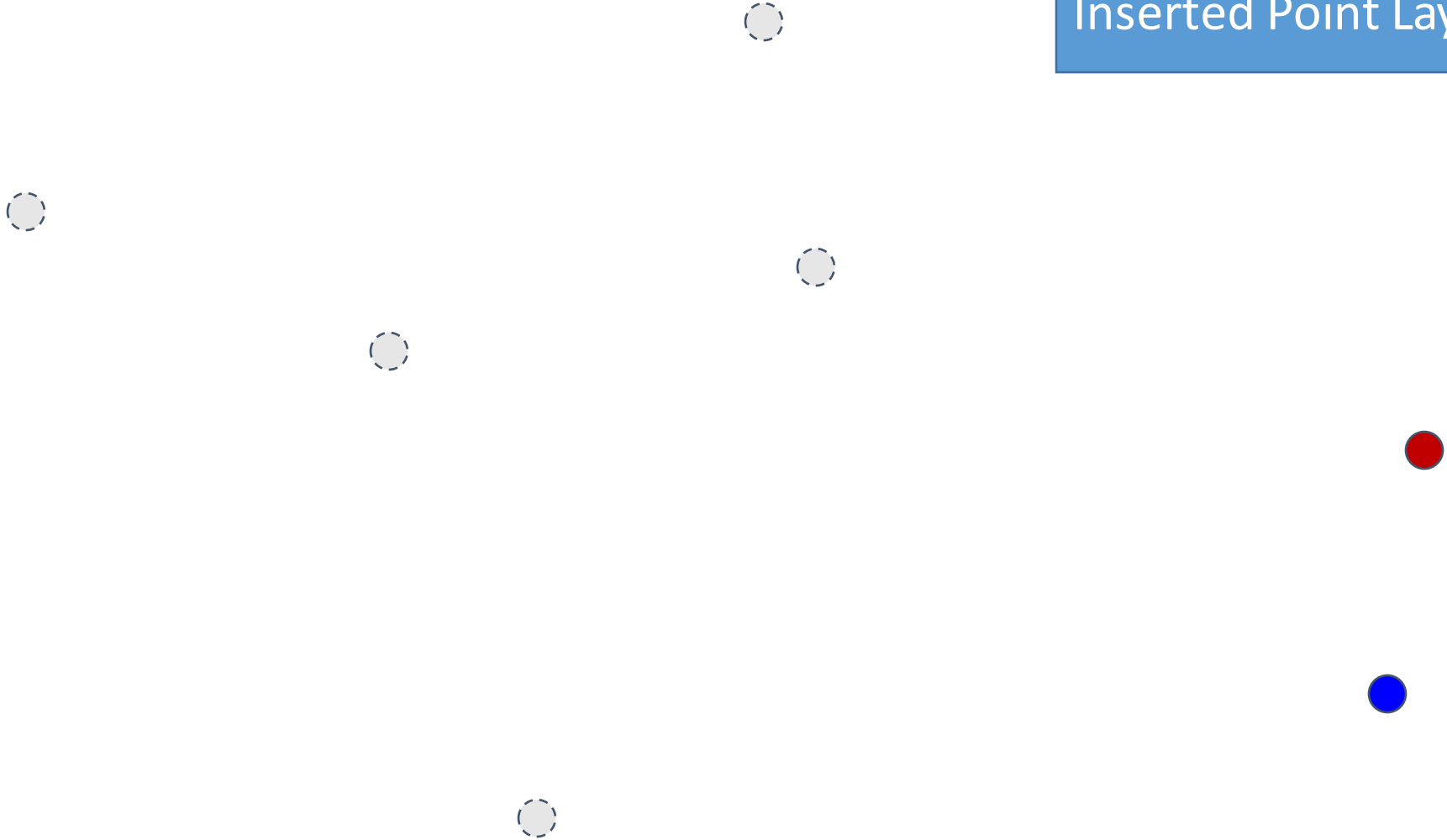
	Point inside layer		Point being inserted		Search waypoint		Connection in layer
	Point outside layer		Point being queried		Root		Search path



Current Layer : 1
Inserted Point Layer : 0









Legend

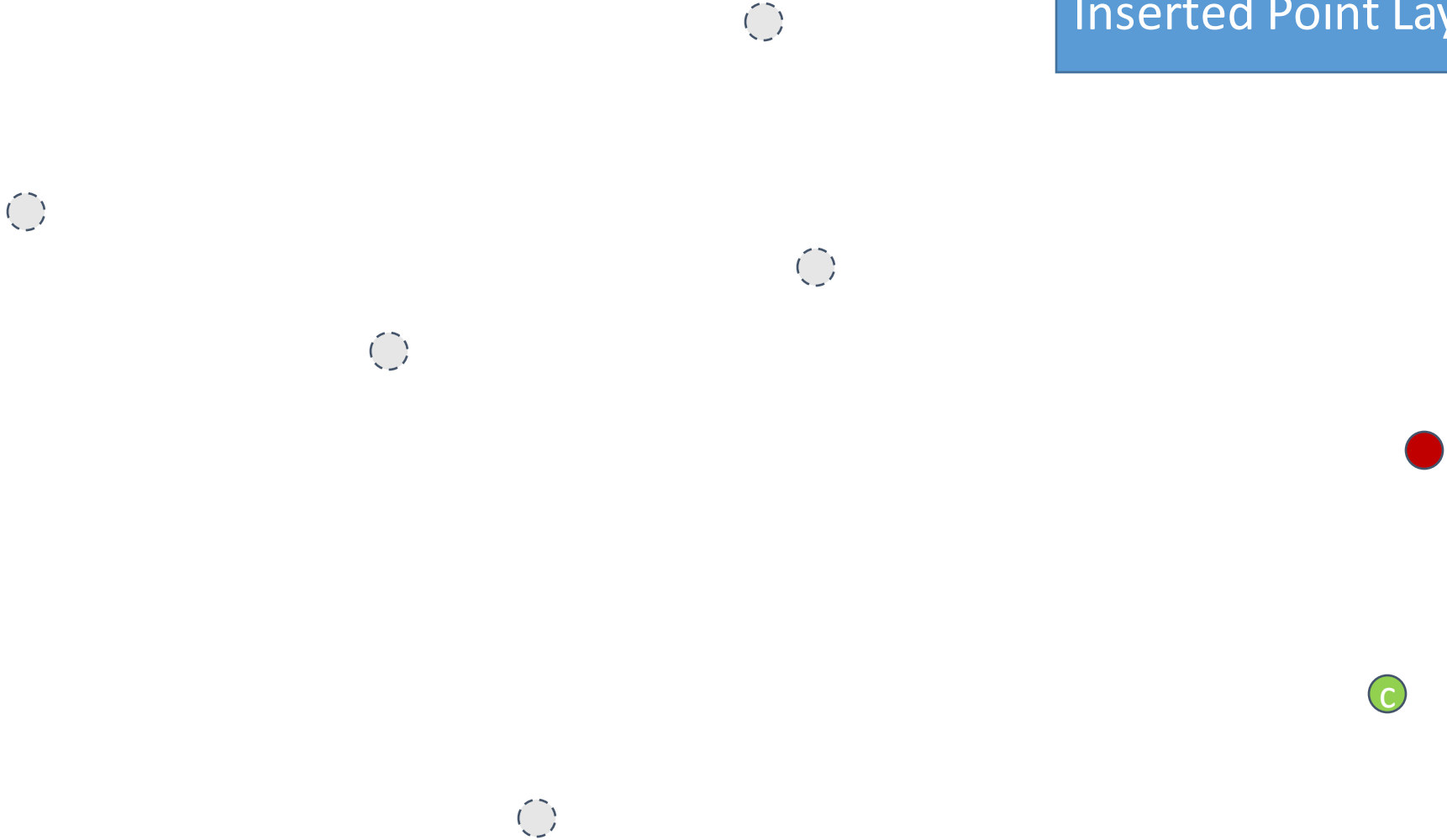
	Point inside layer		Point being inserted		Search waypoint		Connection in layer
	Point outside layer		Point being queried		Root		Search path



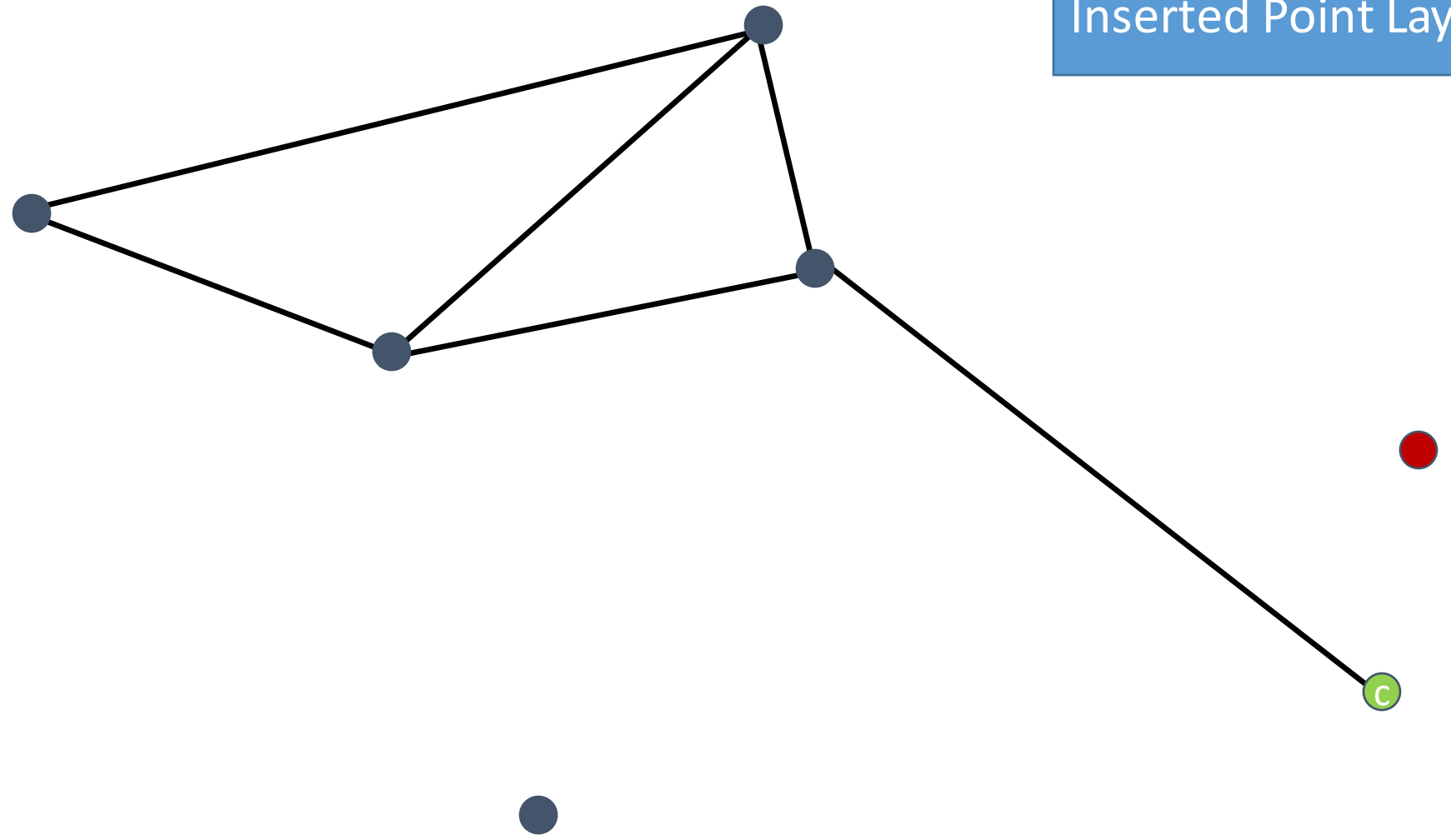
Current Layer : 1
Inserted Point Layer : 0

Legend

	Point inside layer		Point being inserted		Search waypoint		Connection in layer
	Point outside layer		Point being queried		Root		Search path

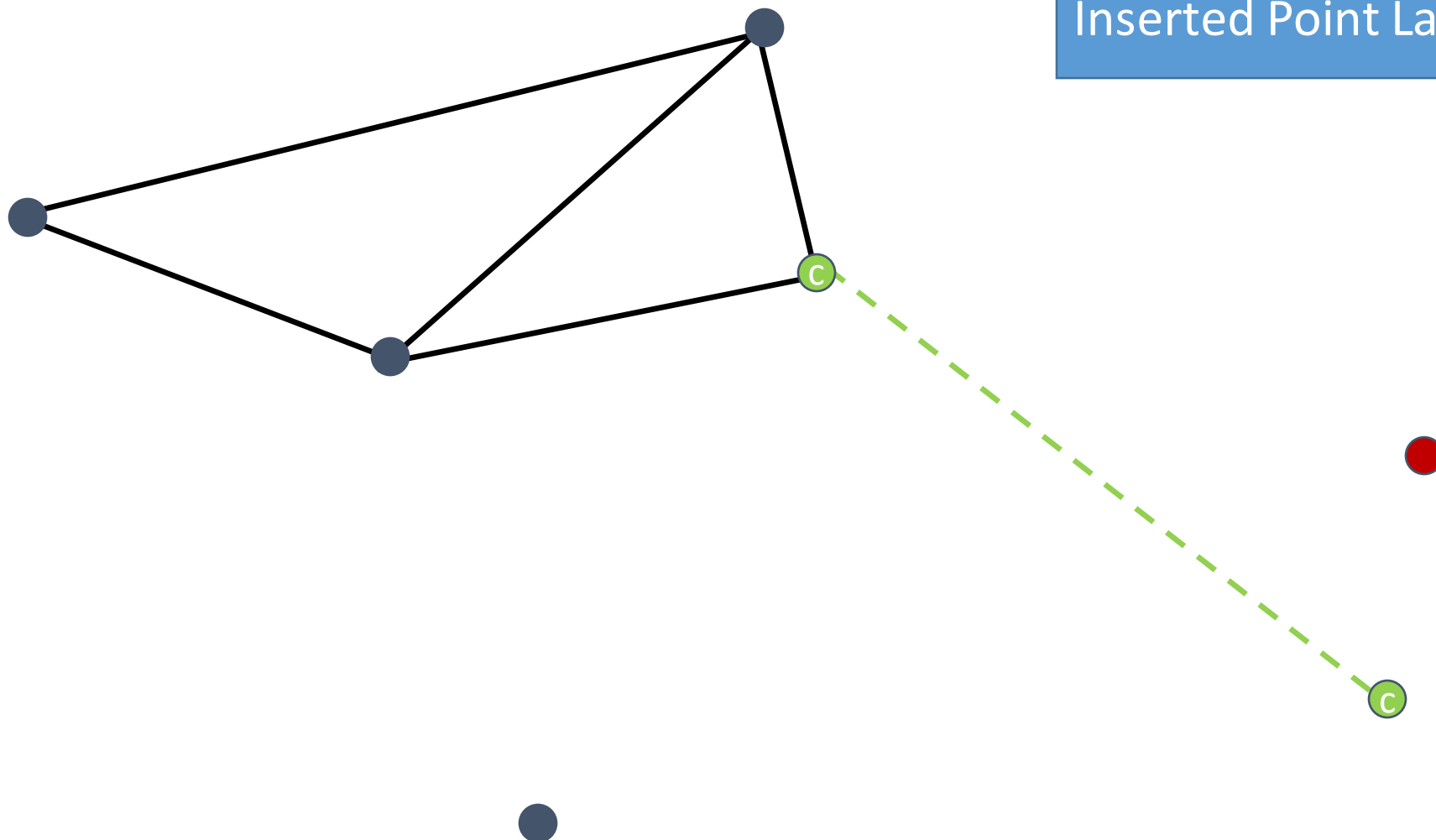


Current Layer : 0
Inserted Point Layer : 0



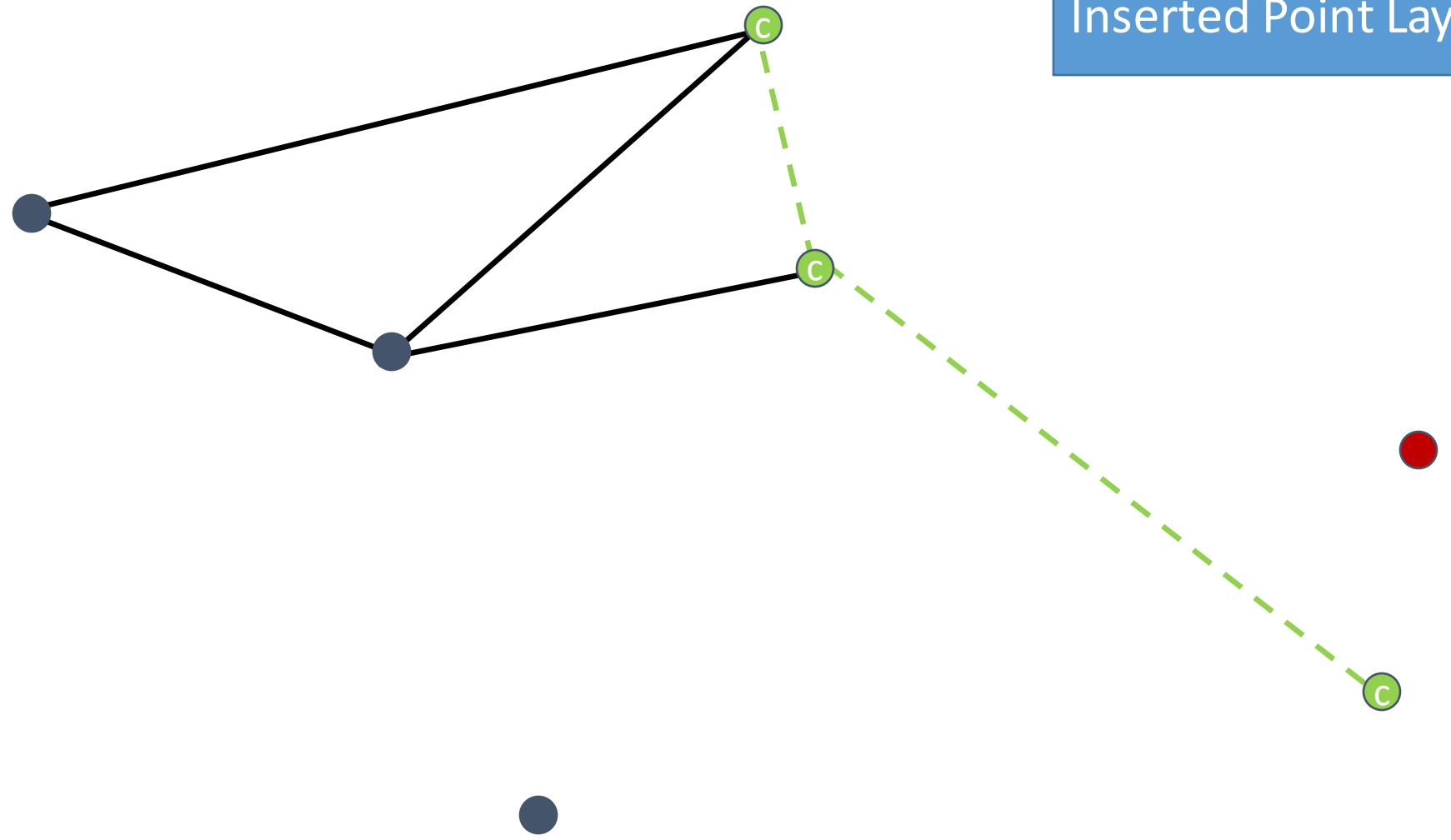
- | | | | | |
|--------|---------------------|----------------------|-----------------|---------------------|
| Legend | Point inside layer | Point being inserted | Search waypoint | Connection in layer |
| | Point outside layer | Point being queried | Root | Search path |

Current Layer : 0
Inserted Point Layer : 0



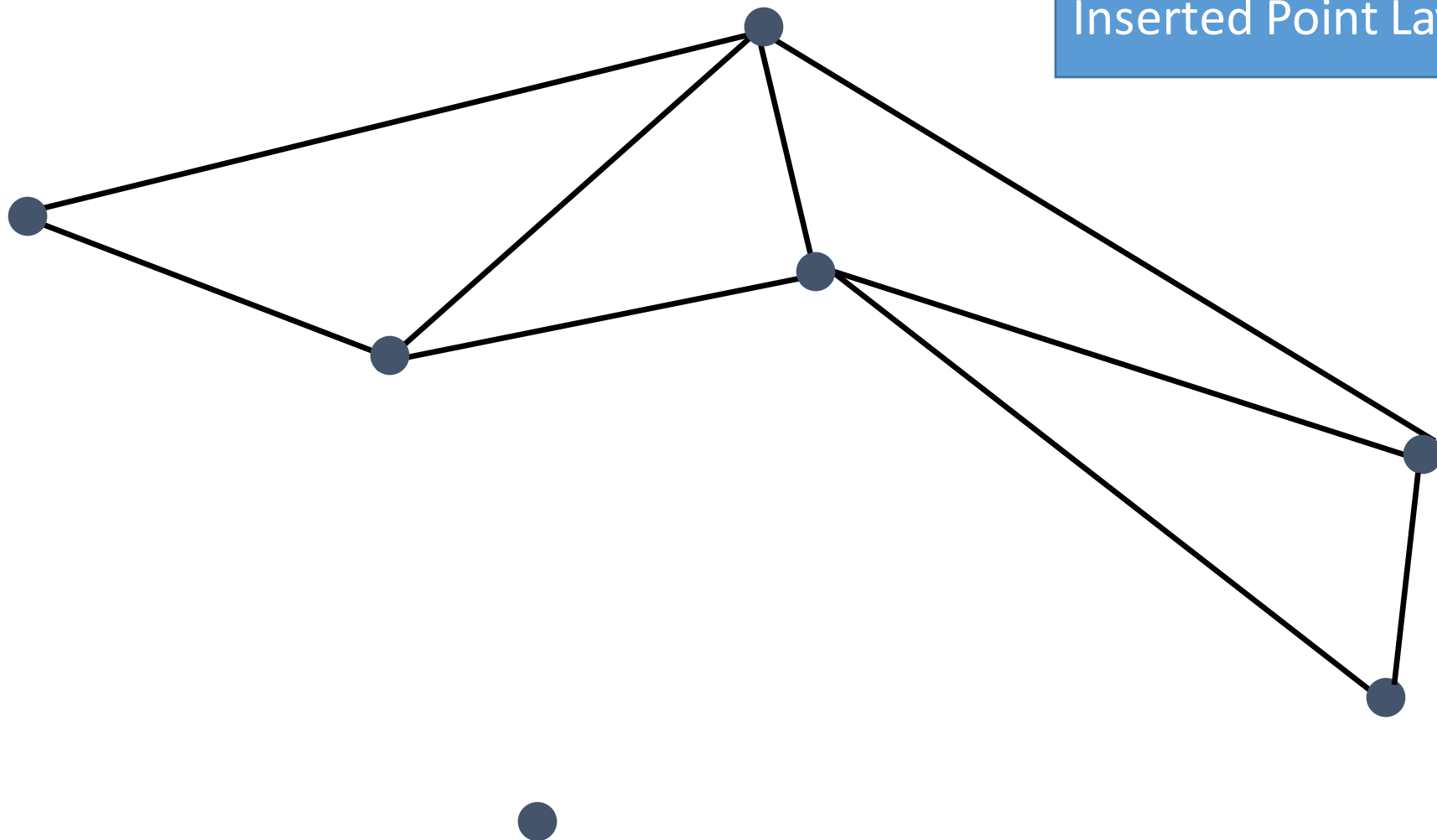
Legend	Point inside layer	Point being inserted	Search waypoint	Connection in layer
	Point outside layer	Point being queried	Root	Search path

Current Layer : 0
Inserted Point Layer : 0



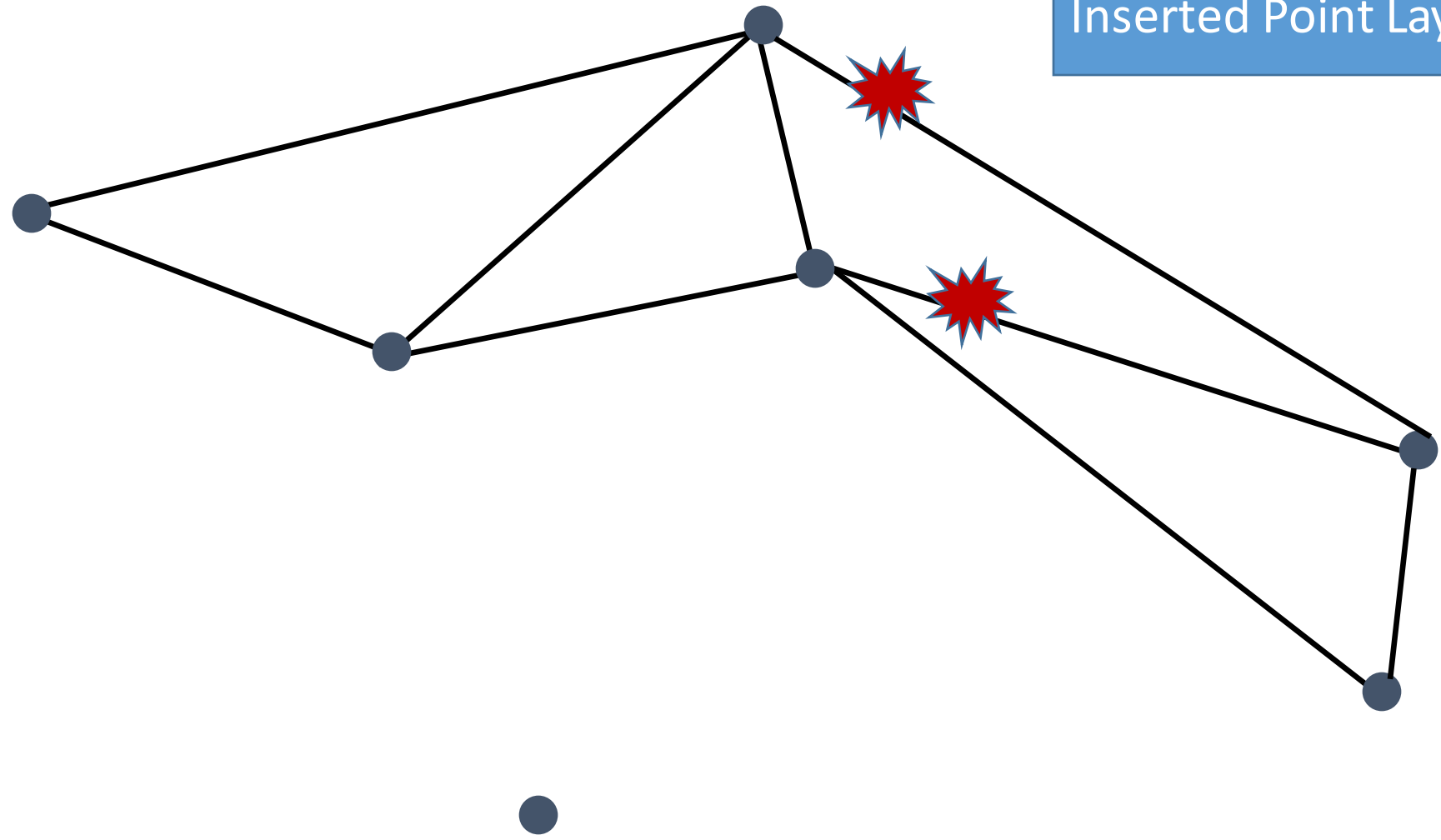
- | | | | | |
|--------|---------------------|----------------------|-----------------|---------------------|
| Legend | Point inside layer | Point being inserted | Search waypoint | Connection in layer |
| | Point outside layer | Point being queried | Root | Search path |









Current Layer : 0
Inserted Point Layer : 0



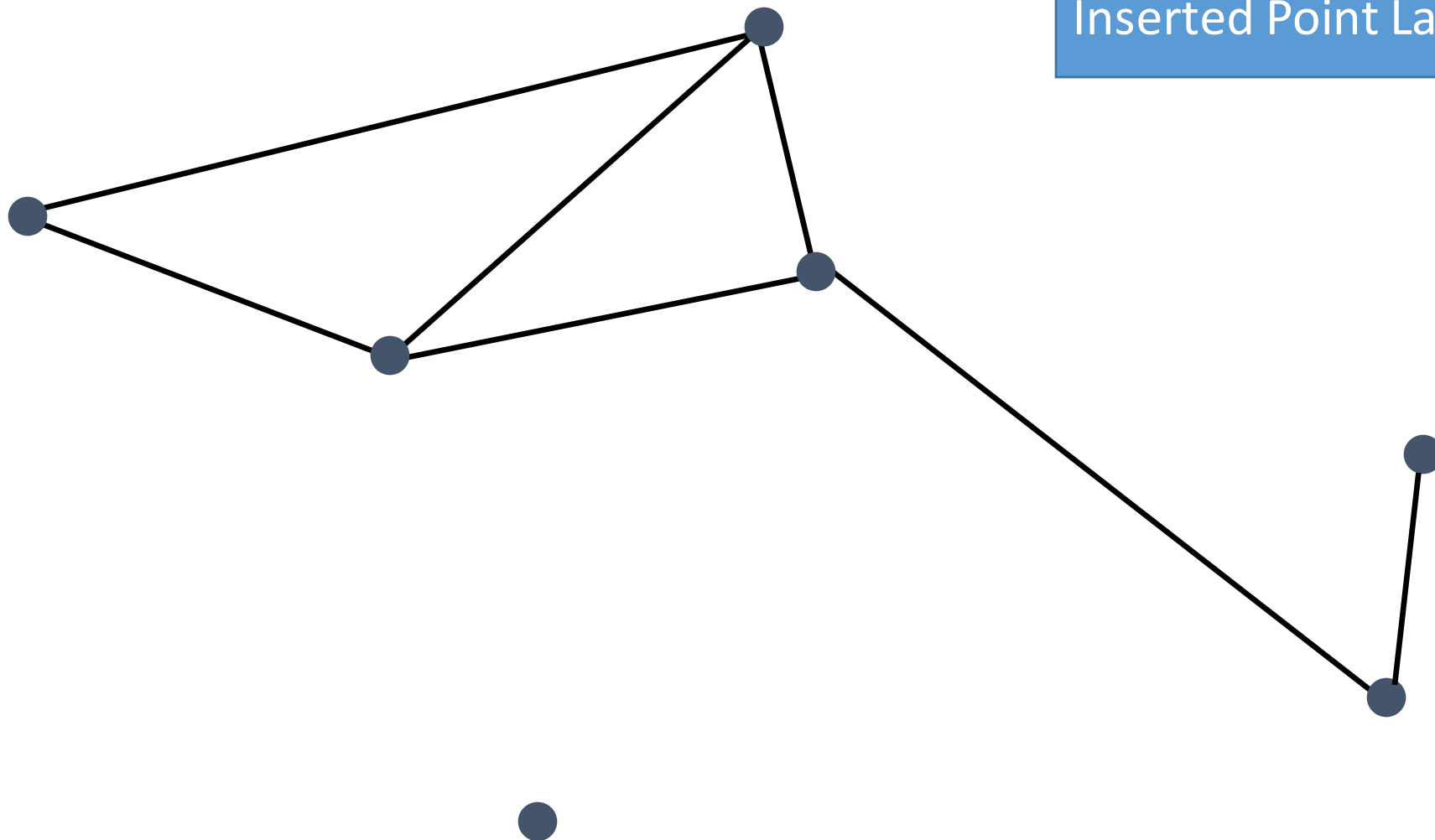
Legend	Point inside layer	Point being inserted	Search waypoint	Connection in layer
	Point outside layer	Point being queried	Root	Search path

Current Layer : 0
Inserted Point Layer : 0



- | | | | | |
|--------|---|--|---|---|
| Legend |  Point inside layer |  Point being inserted |  Search waypoint |  Connection in layer |
| |  Point outside layer |  Point being queried |  Root |  Search path |









Current Layer : 0
Inserted Point Layer : 0

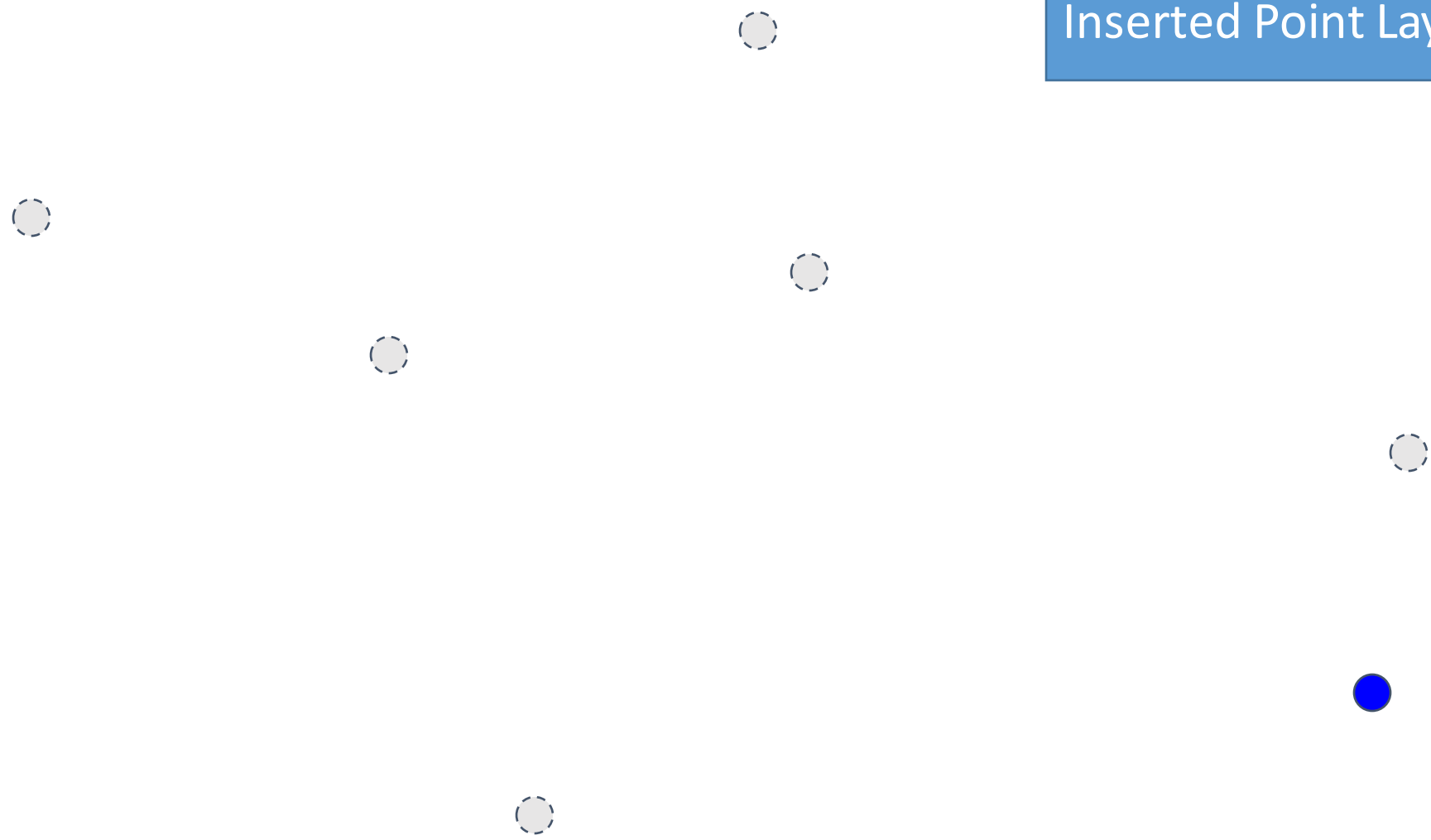


Legend	Point inside layer	Point being inserted	Search waypoint	Connection in layer
	Point outside layer	Point being queried	Root	Search path

Current Layer : 1
Inserted Point Layer : 0









Legend

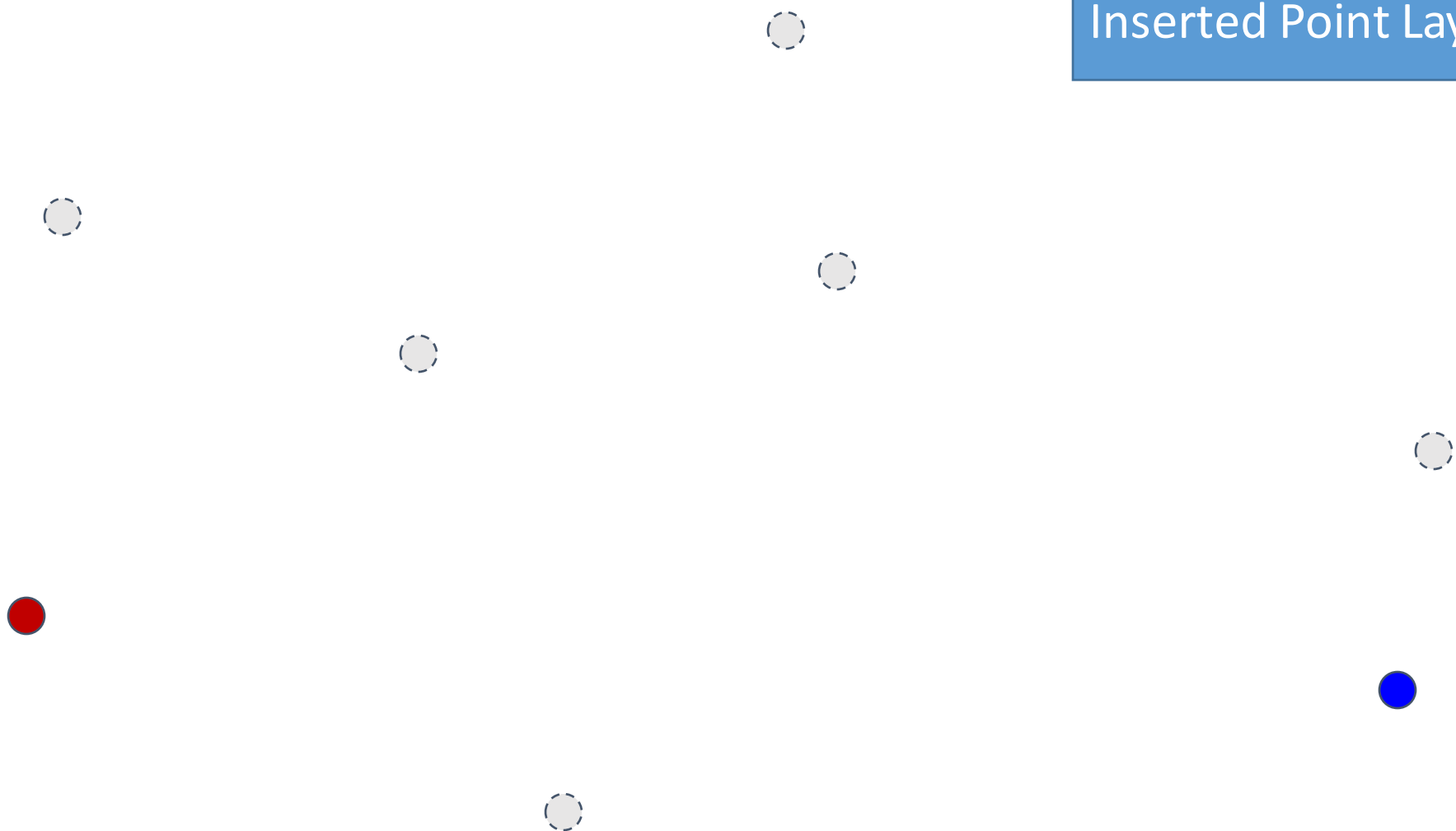
	Point inside layer		Point being inserted		Search waypoint		Connection in layer
	Point outside layer		Point being queried		Root		Search path



Current Layer : 1
Inserted Point Layer : 1









Legend

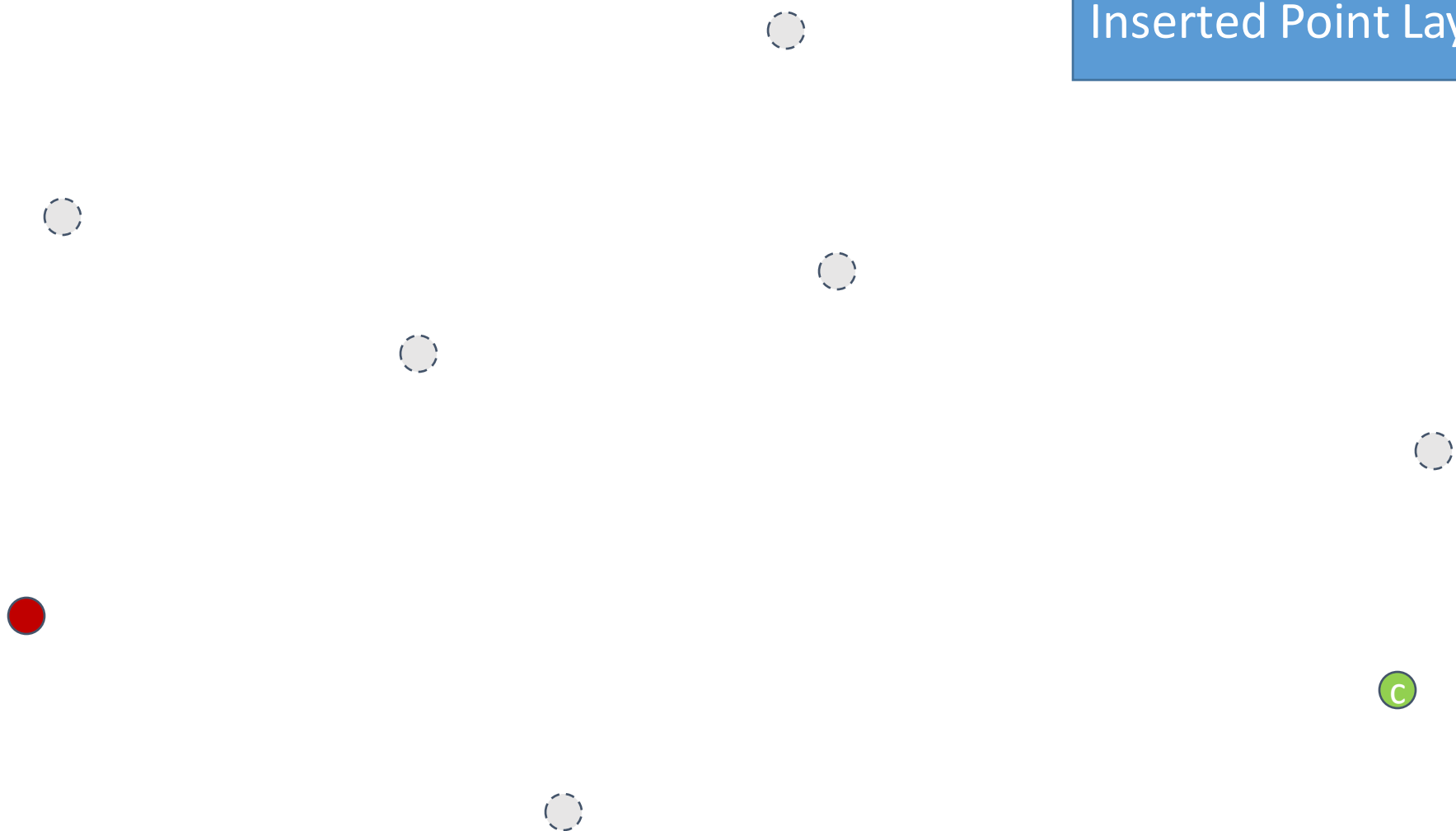
	Point inside layer		Point being inserted		Search waypoint		Connection in layer
	Point outside layer		Point being queried		Root		Search path



Current Layer : 1
Inserted Point Layer : 1

Legend

	Point inside layer		Point being inserted		Search waypoint		Connection in layer
	Point outside layer		Point being queried		Root		Search path

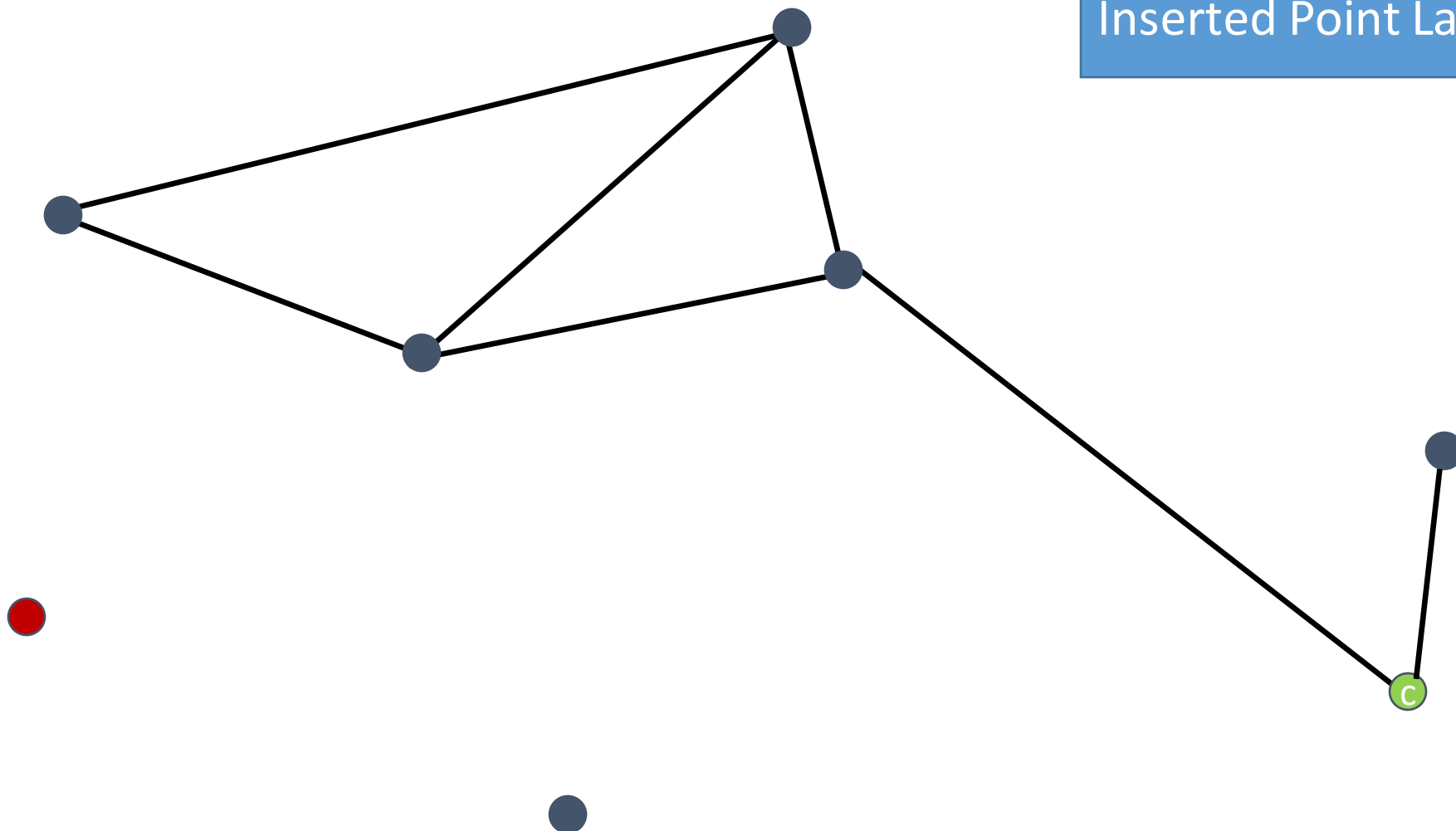


Current Layer : 1
Inserted Point Layer : 1



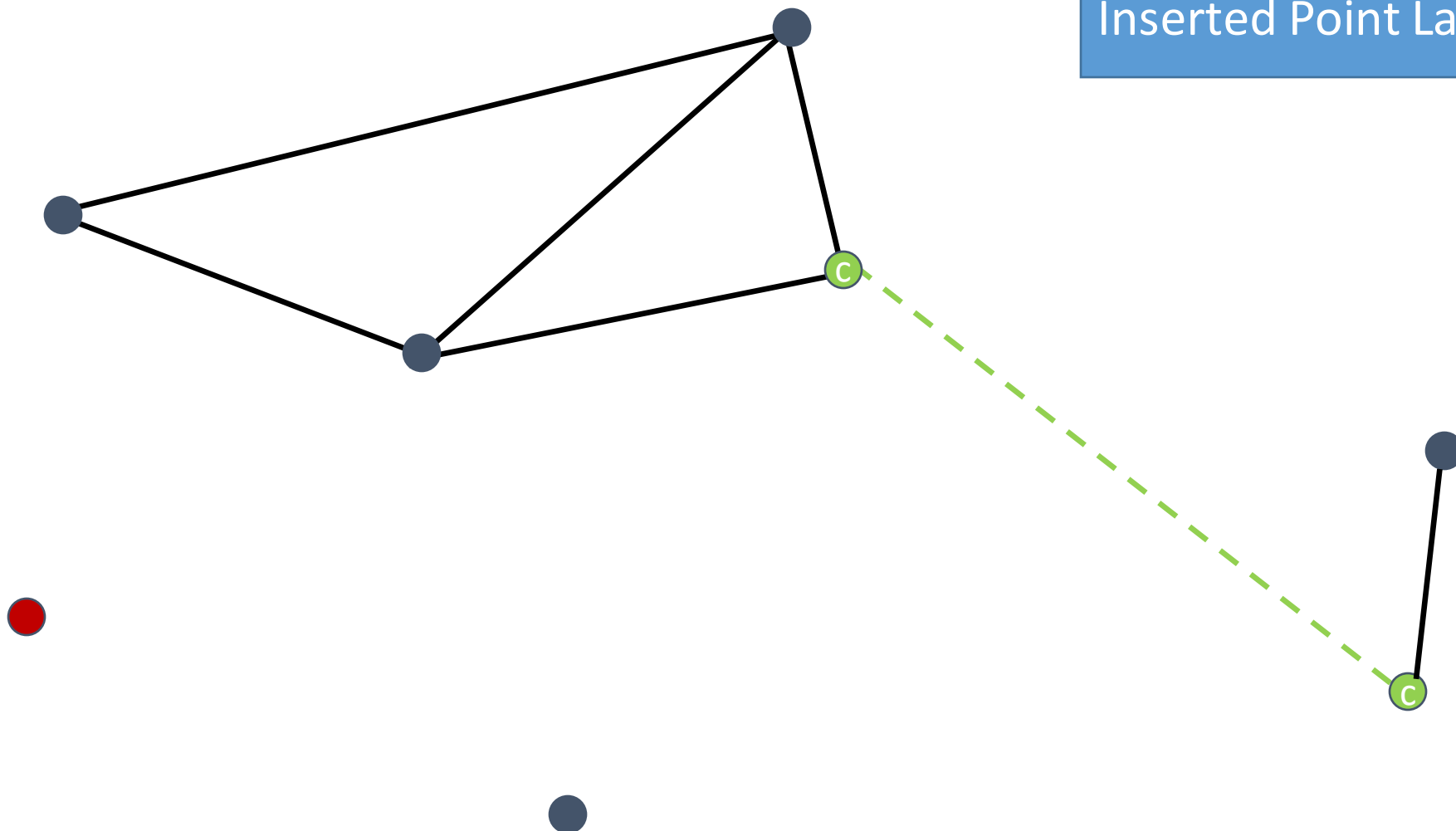
- | | | | | |
|--------|---------------------|----------------------|-----------------|---------------------|
| Legend | Point inside layer | Point being inserted | Search waypoint | Connection in layer |
| | Point outside layer | Point being queried | Root | Search path |

Current Layer : 0
Inserted Point Layer : 0

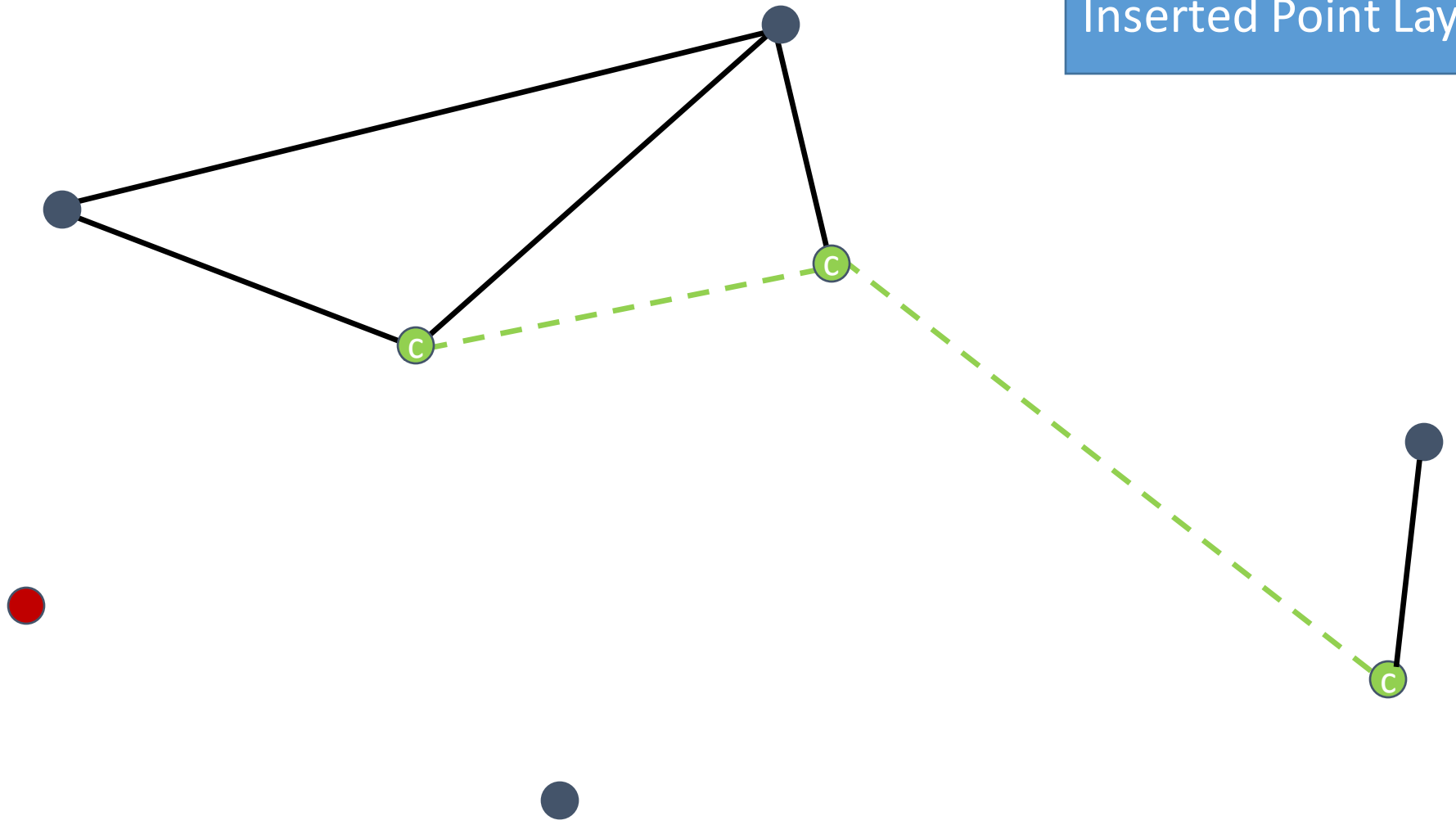


Legend	Point inside layer	Point being inserted	Search waypoint	Connection in layer
	Point outside layer	Point being queried	Root	Search path

Current Layer : 0
Inserted Point Layer : 0

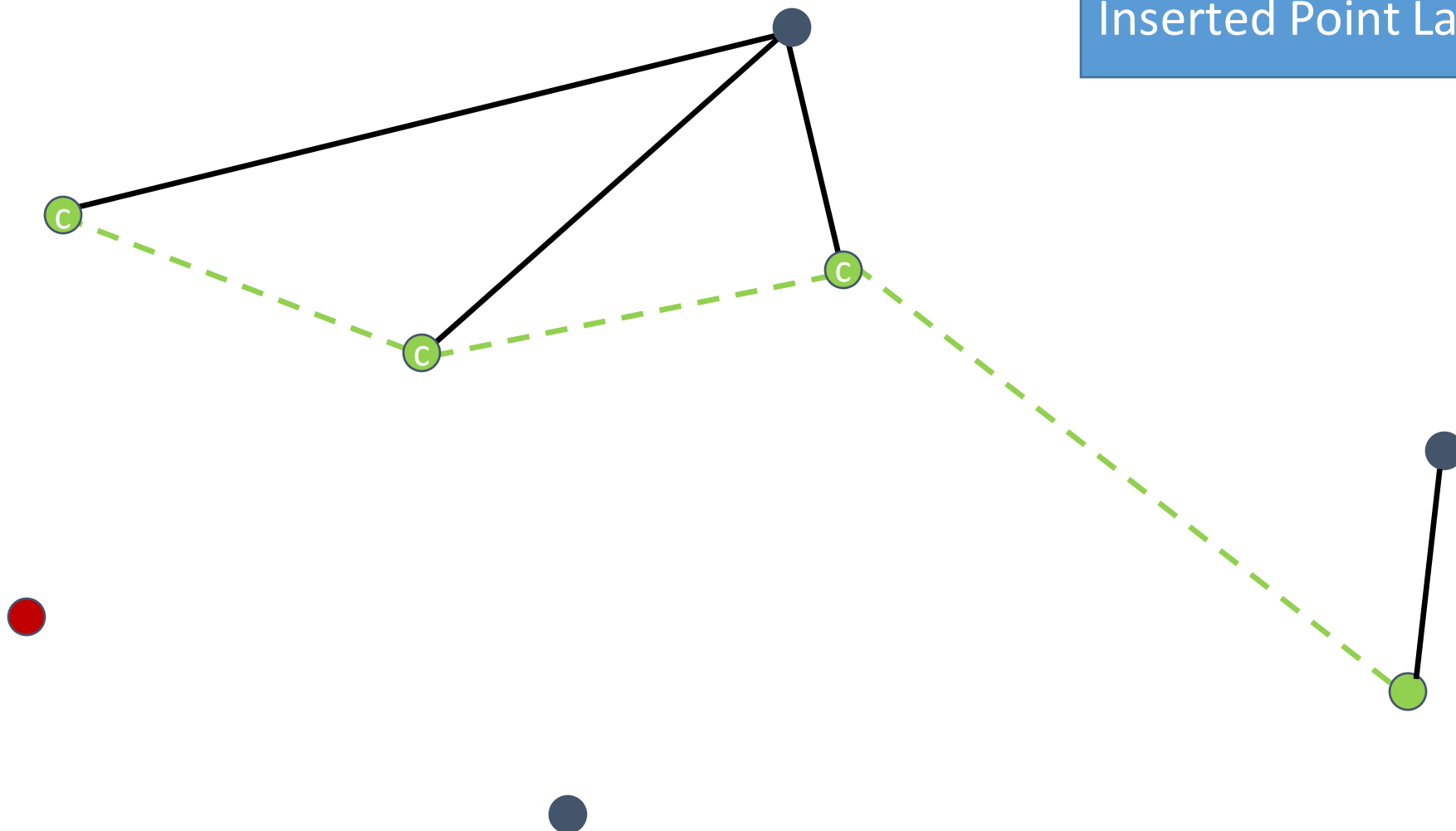


Current Layer : 0
Inserted Point Layer : 0



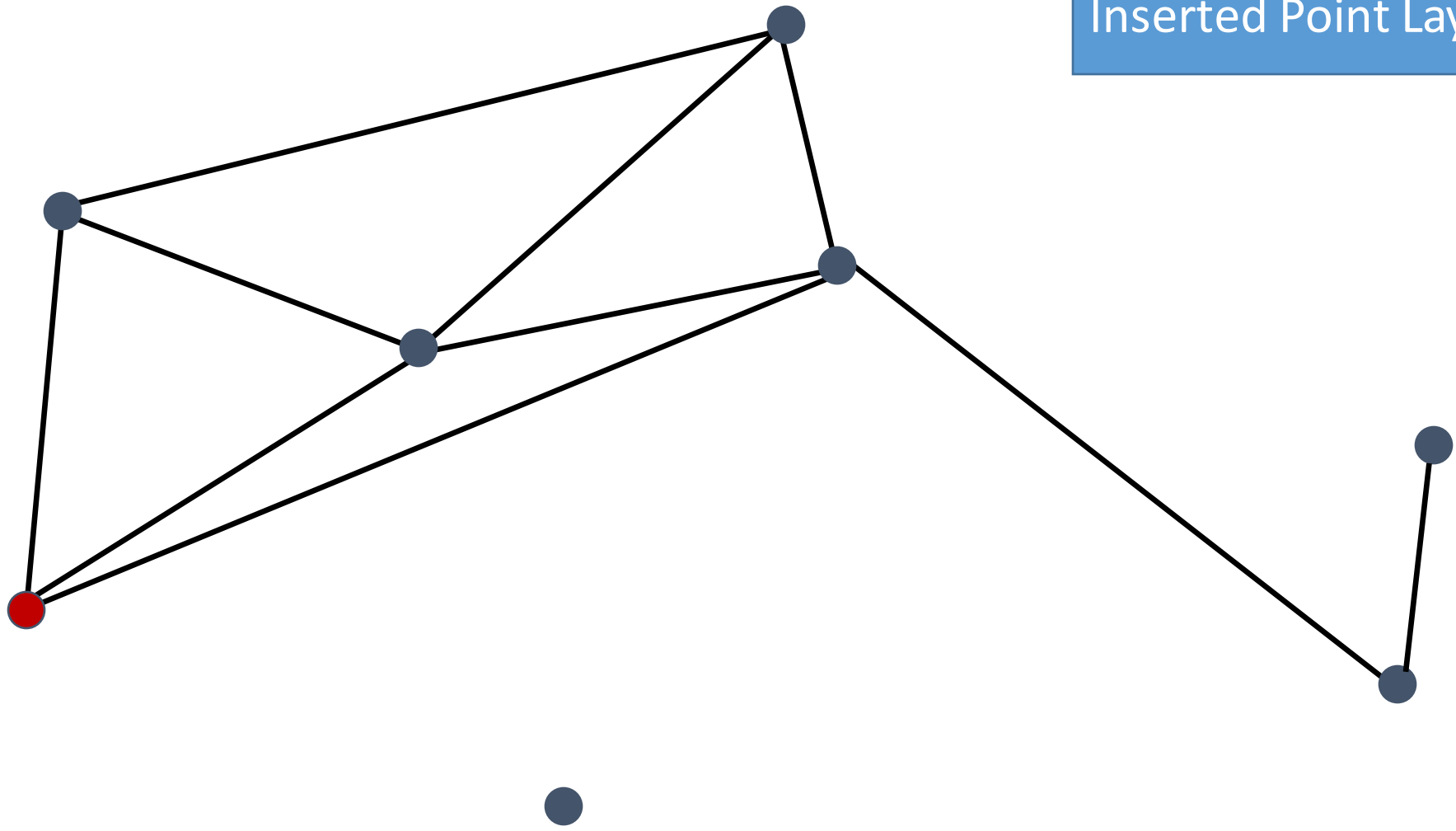
- | | | | | |
|--------|---------------------|----------------------|-----------------|---------------------|
| Legend | Point inside layer | Point being inserted | Search waypoint | Connection in layer |
| | Point outside layer | Point being queried | Root | Search path |

Current Layer : 0
Inserted Point Layer : 0



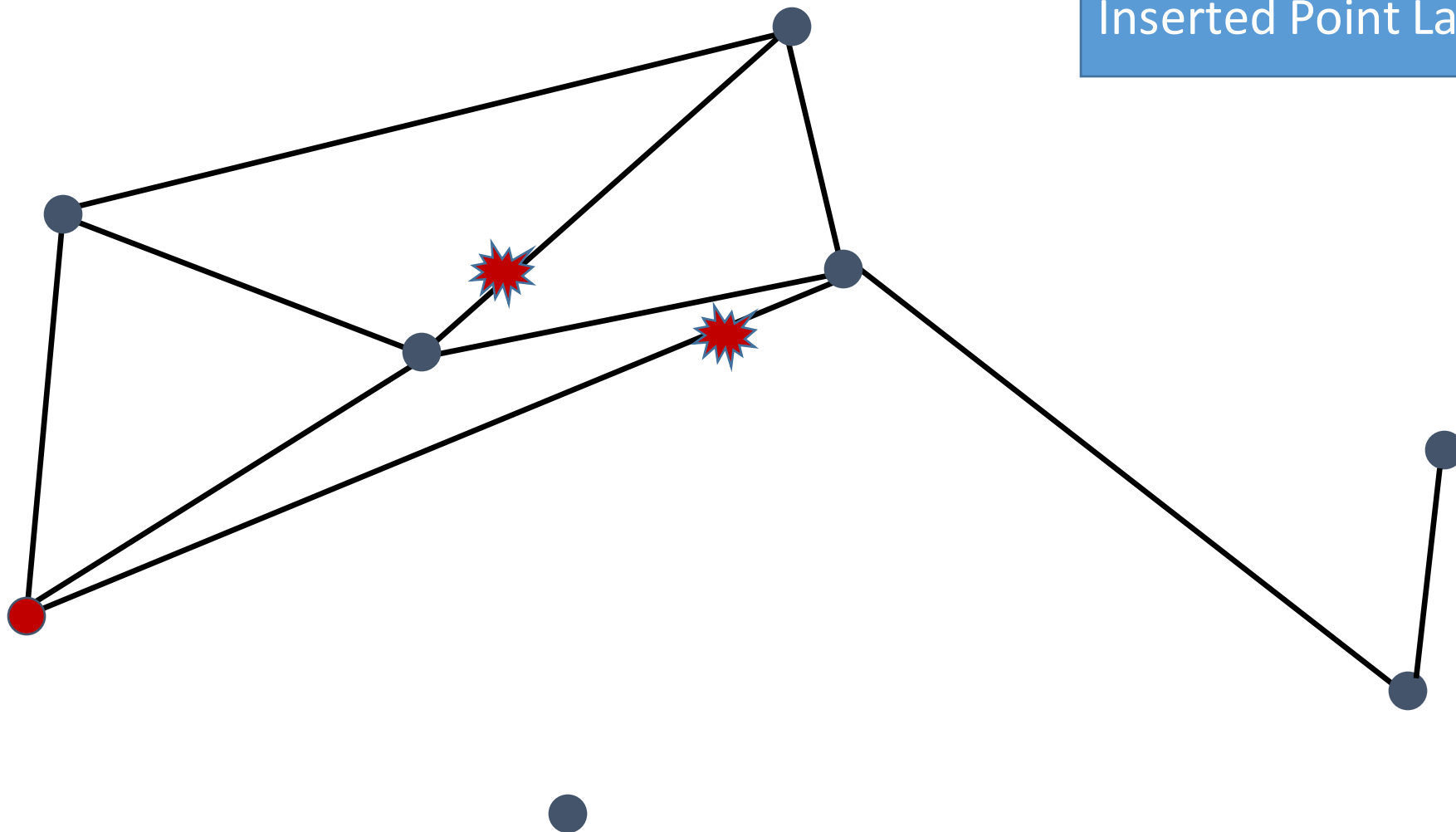
Legend	Point inside layer	Point being inserted	Search waypoint	Connection in layer
	Point outside layer	Point being queried	Root	Search path

Current Layer : 0
Inserted Point Layer : 0



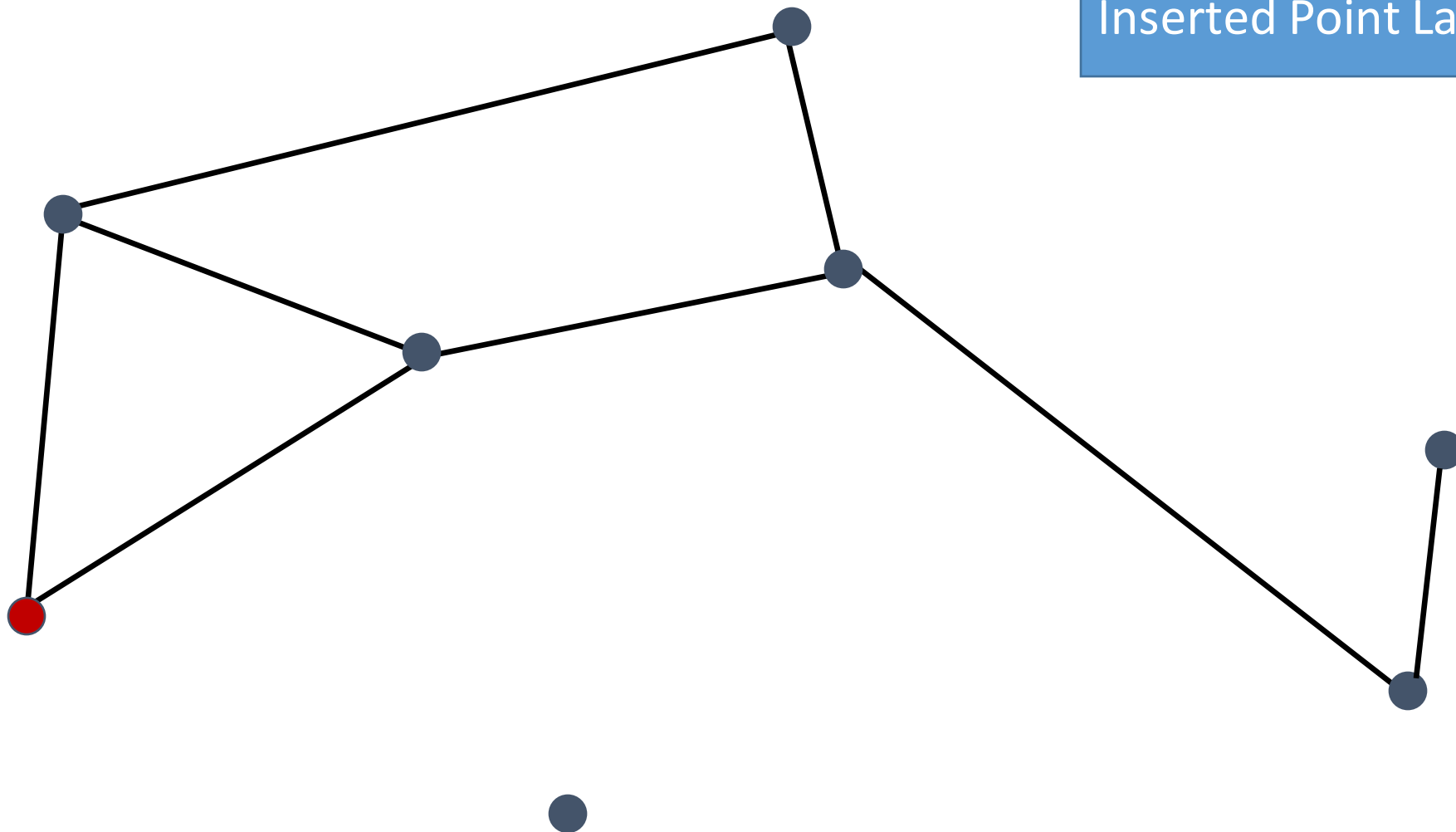
- | | | | | |
|--------|---------------------|----------------------|-----------------|---------------------|
| Legend | Point inside layer | Point being inserted | Search waypoint | Connection in layer |
| | Point outside layer | Point being queried | Root | Search path |

Current Layer : 0
Inserted Point Layer : 0



Legend	● Point inside layer	● Point being inserted	● Search waypoint	— Connection in layer
	○ Point outside layer	● Point being queried	● Root	- - - Search path









Current Layer : 0
Inserted Point Layer : 0

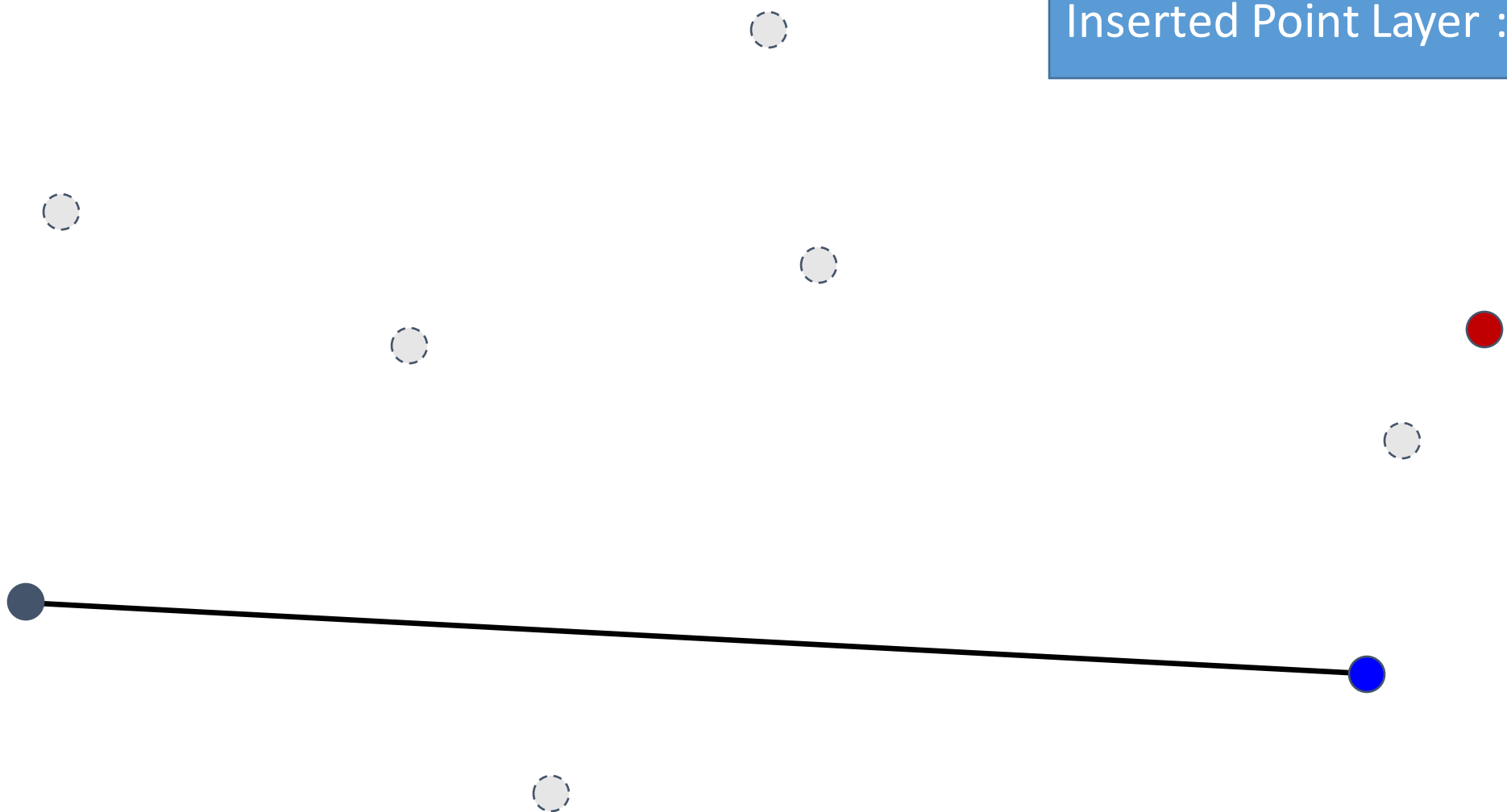


Legend	Point inside layer	Point being inserted	Search waypoint	Connection in layer
	Point outside layer	Point being queried	Root	Search path

Current Layer : 1
Inserted Point Layer : 0









Legend

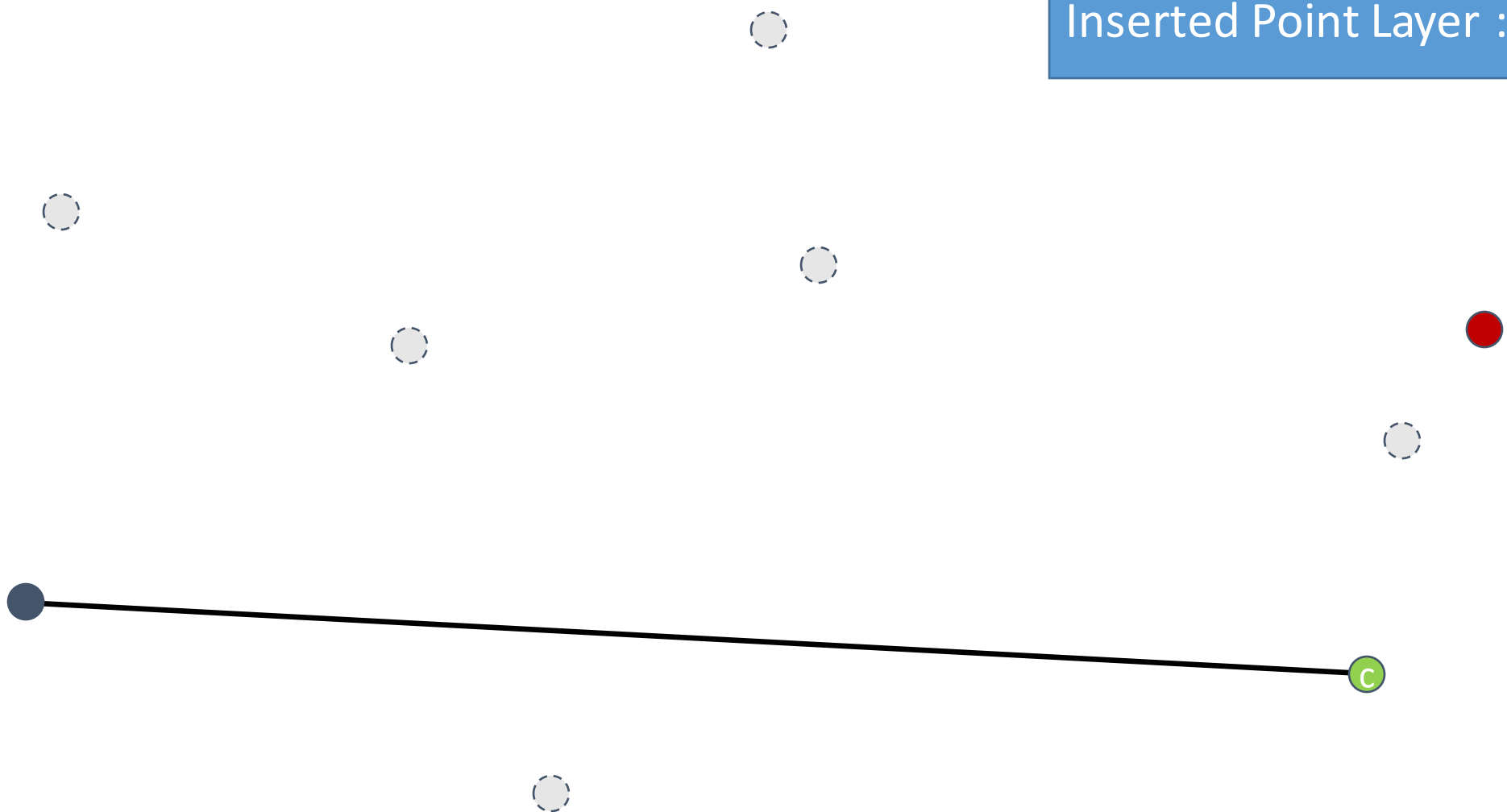
	Point inside layer		Point being inserted		Search waypoint		Connection in layer
	Point outside layer		Point being queried		Root		Search path



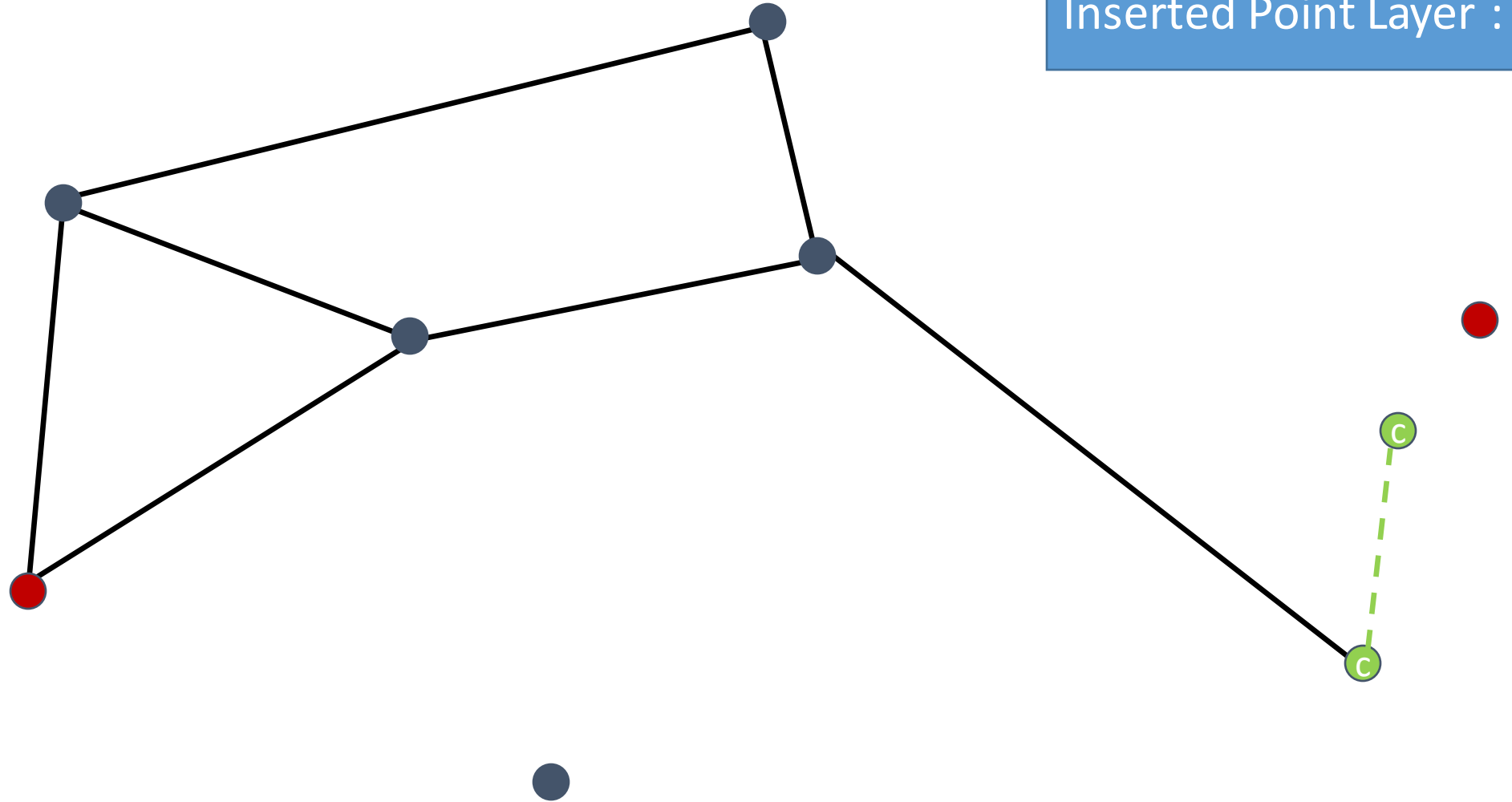
Current Layer : 1
Inserted Point Layer : 0

Legend

	Point inside layer		Point being inserted		Search waypoint		Connection in layer
	Point outside layer		Point being queried		Root		Search path

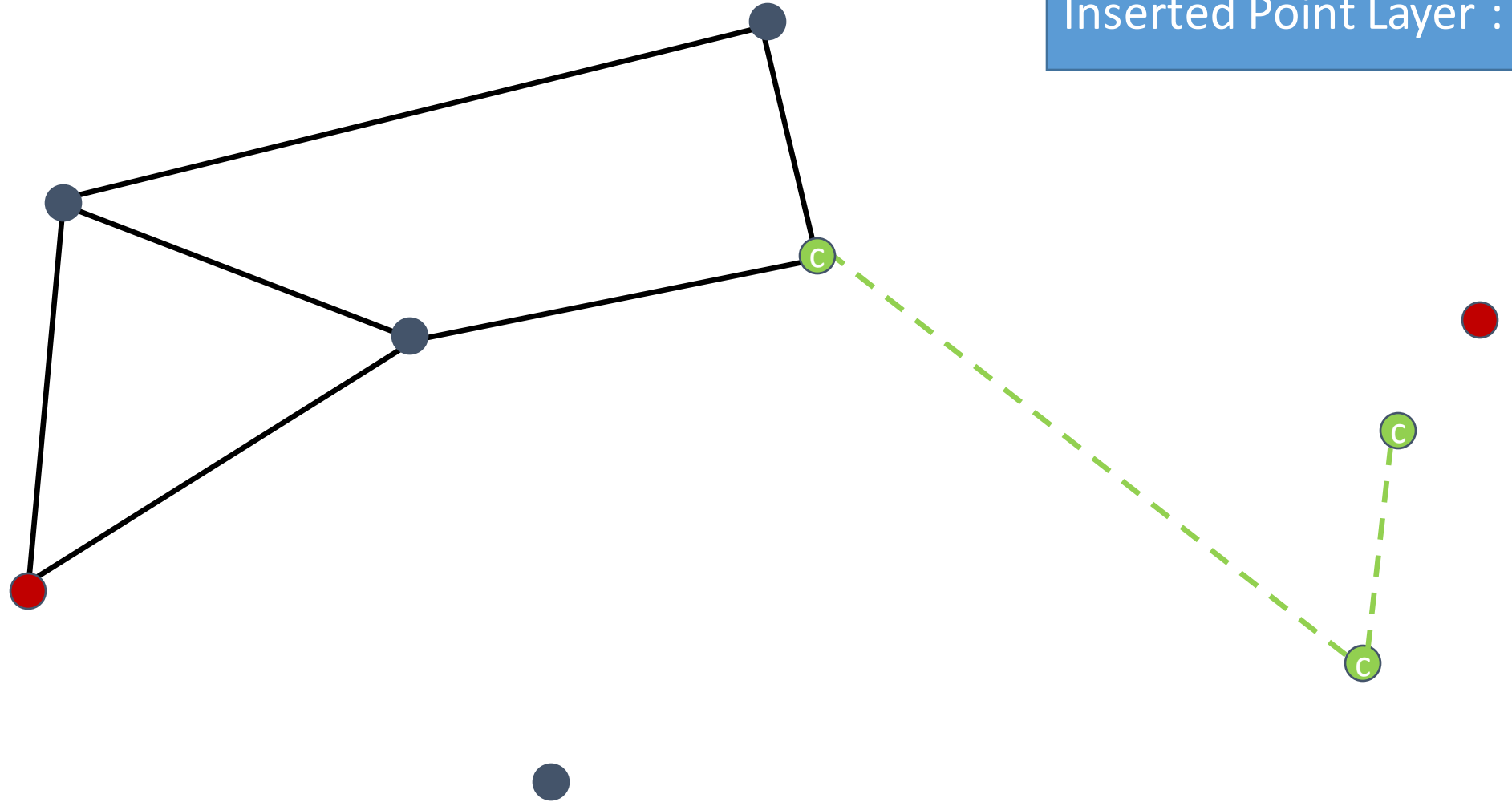


Current Layer : 0
Inserted Point Layer : 0

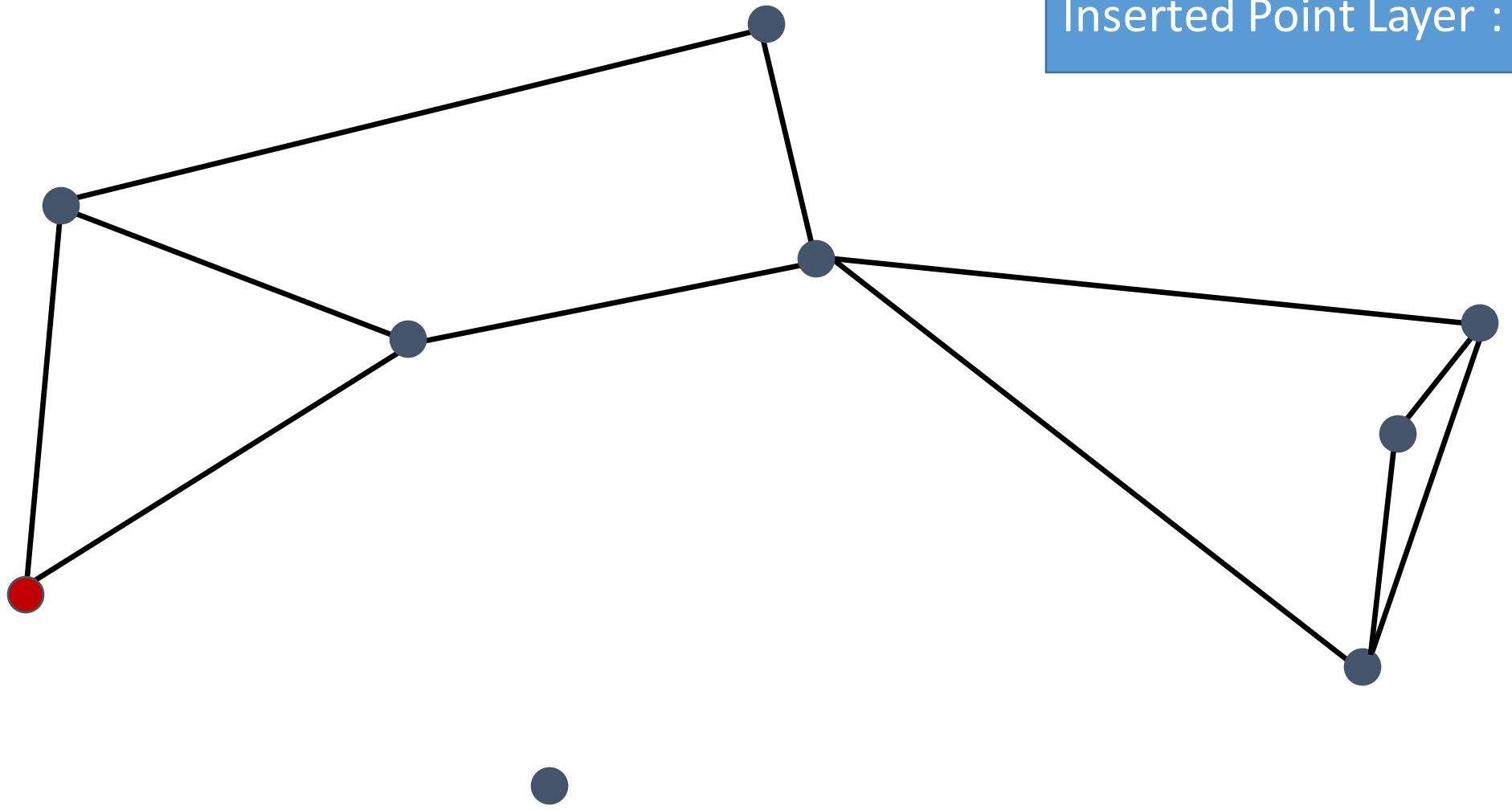


Legend	Point inside layer	Point being inserted	Search waypoint	Connection in layer
	Point outside layer	Point being queried	Root	Search path

Current Layer : 0
Inserted Point Layer : 0

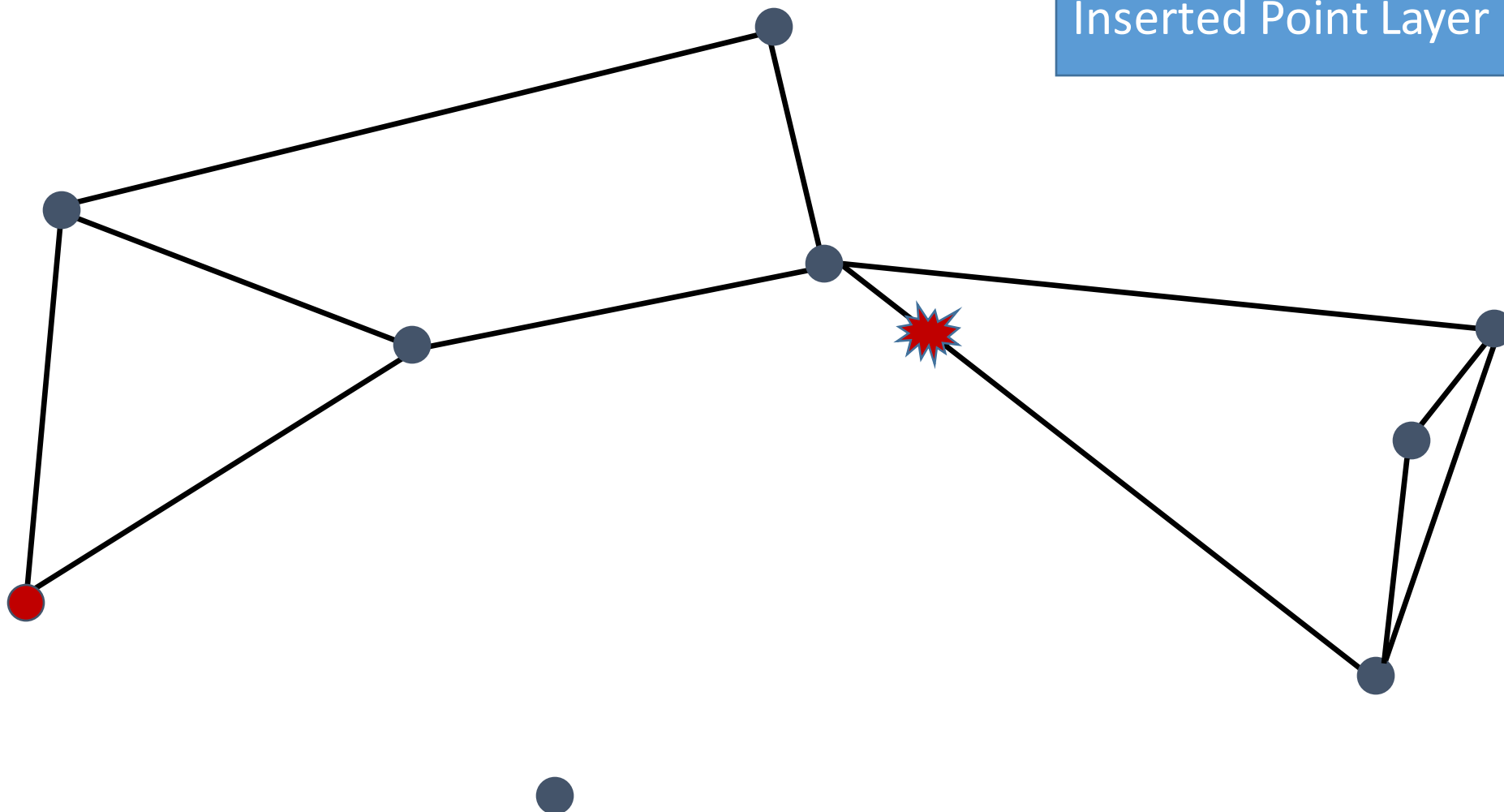


Current Layer : 0
Inserted Point Layer : 0

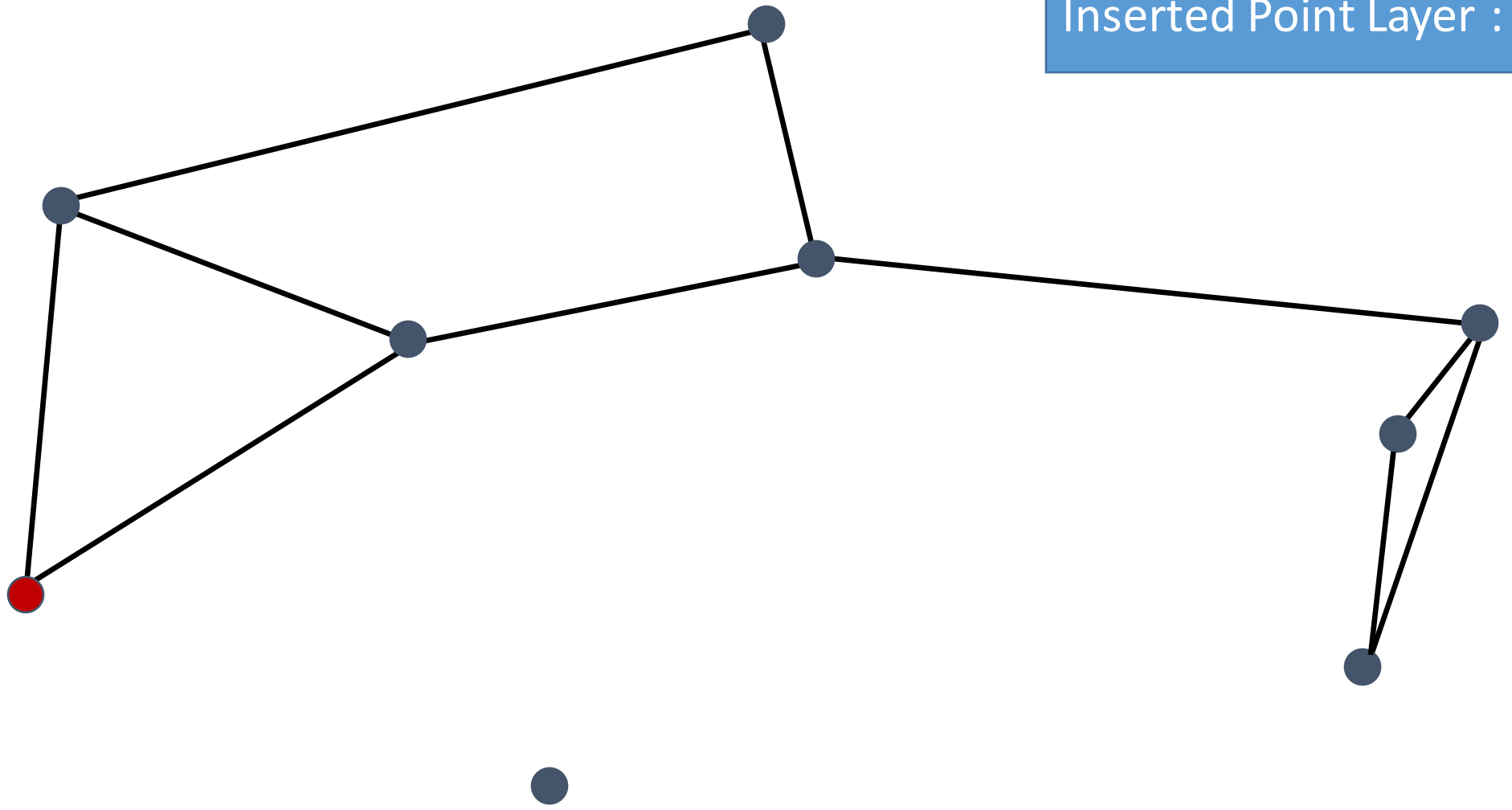


Legend	Point inside layer	Point being inserted	Search waypoint	Connection in layer
	Point outside layer	Point being queried	Root	Search path

Current Layer : 0
Inserted Point Layer : 0











Current Layer : 0
Inserted Point Layer : 0



Legend	Point inside layer	Point being inserted	Search waypoint	Connection in layer
	Point outside layer	Point being queried	Root	Search path

Current Layer : 1
Inserted Point Layer : 1









Legend

	Point inside layer		Point being inserted		Search waypoint		Connection in layer
	Point outside layer		Point being queried		Root		Search path



Current Layer : 1
Inserted Point Layer : 1









Legend

	Point inside layer		Point being inserted		Search waypoint		Connection in layer
	Point outside layer		Point being queried		Root		Search path



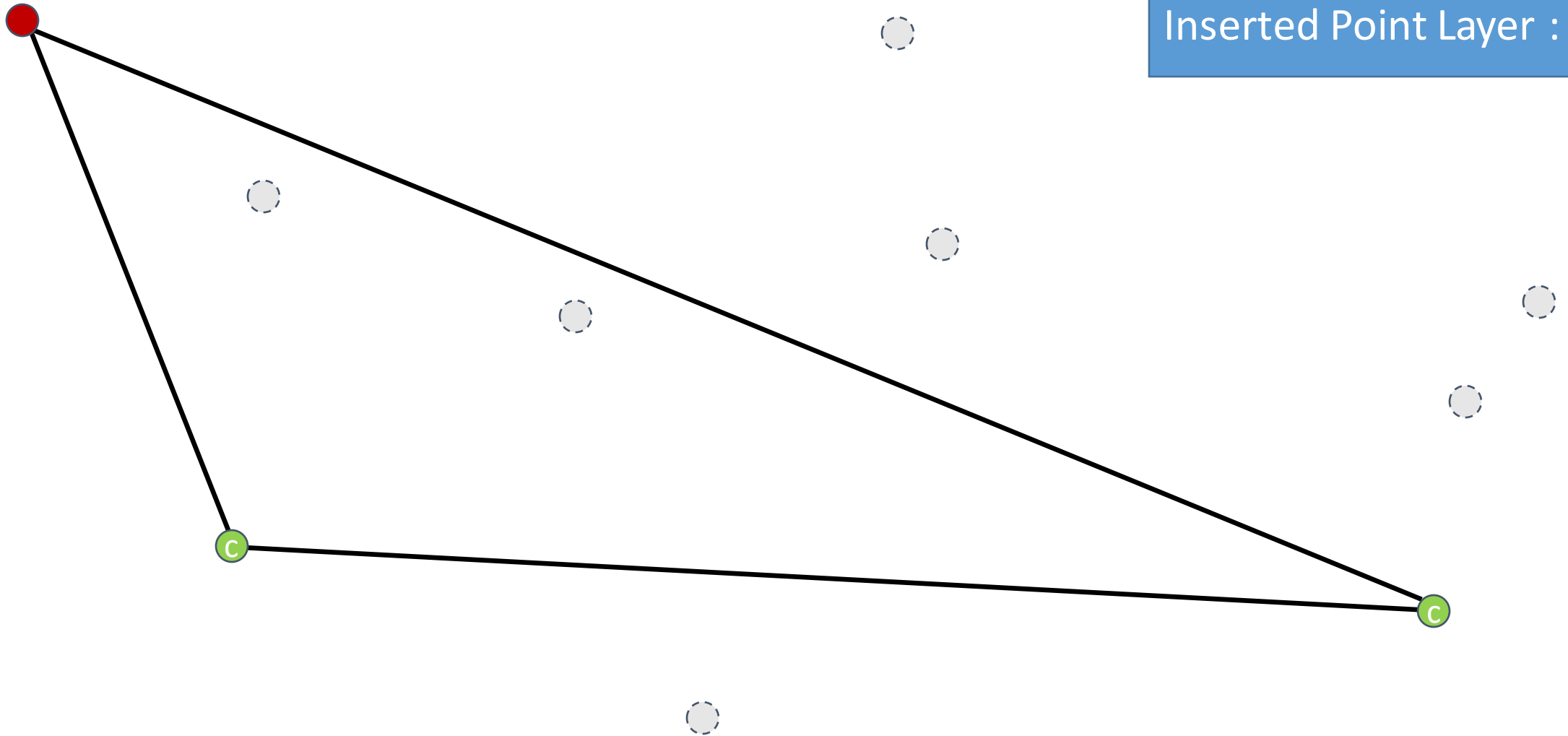
Current Layer : 1
Inserted Point Layer : 1

Legend

	Point inside layer		Point being inserted		Search waypoint		Connection in layer
	Point outside layer		Point being queried		Root		Search path

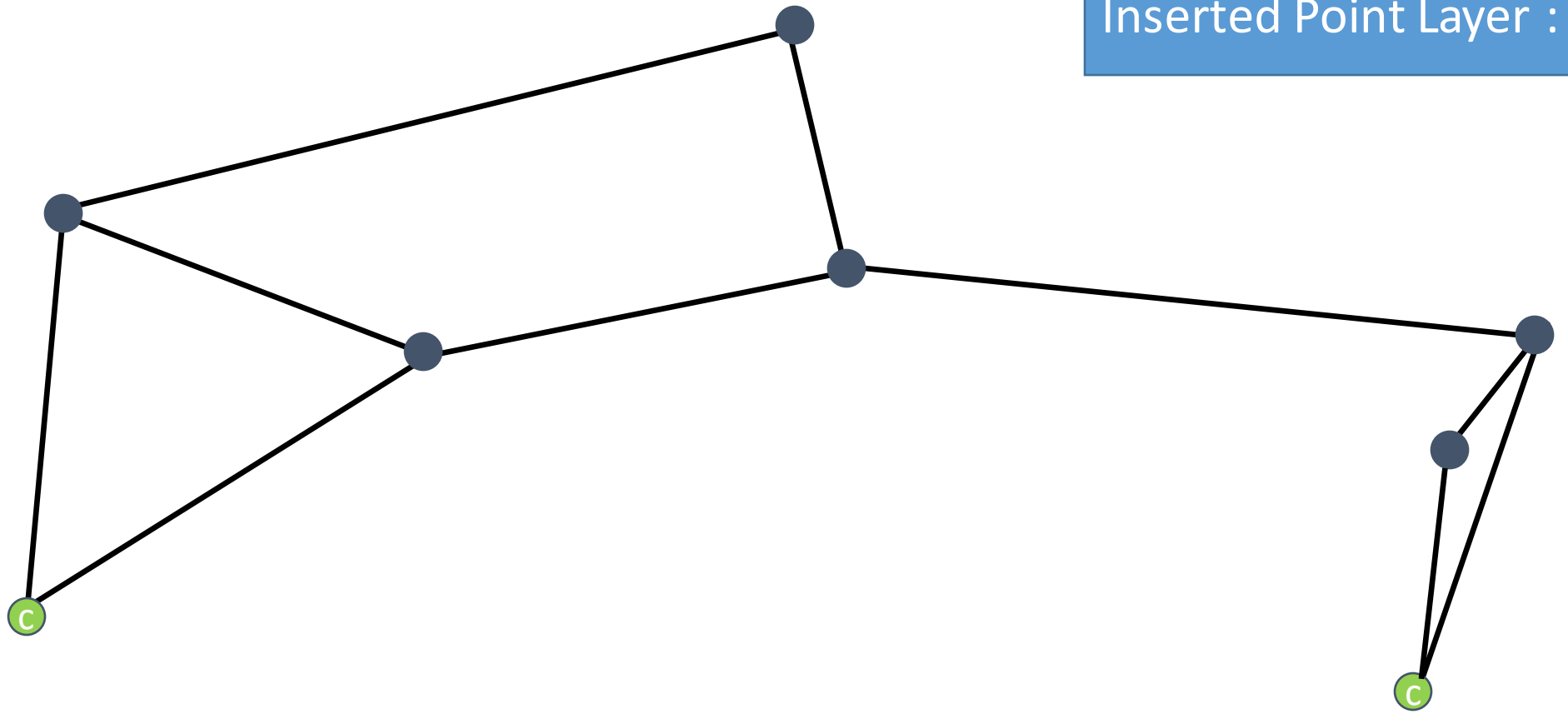


Current Layer : 1
Inserted Point Layer : 1



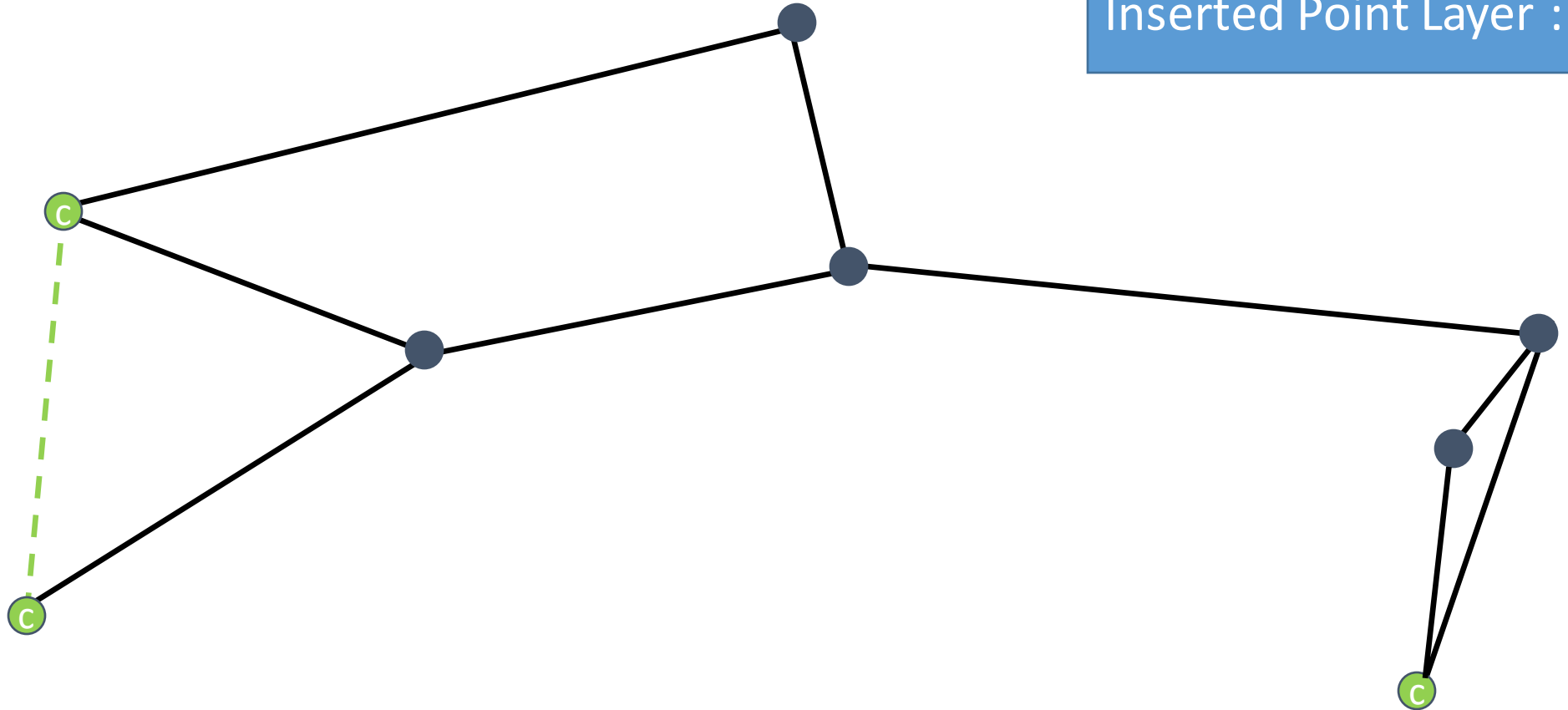
Legend	Point inside layer	Point being inserted	Search waypoint	Connection in layer
	Point outside layer	Point being queried	Root	Search path

Current Layer : 0
Inserted Point Layer : 1



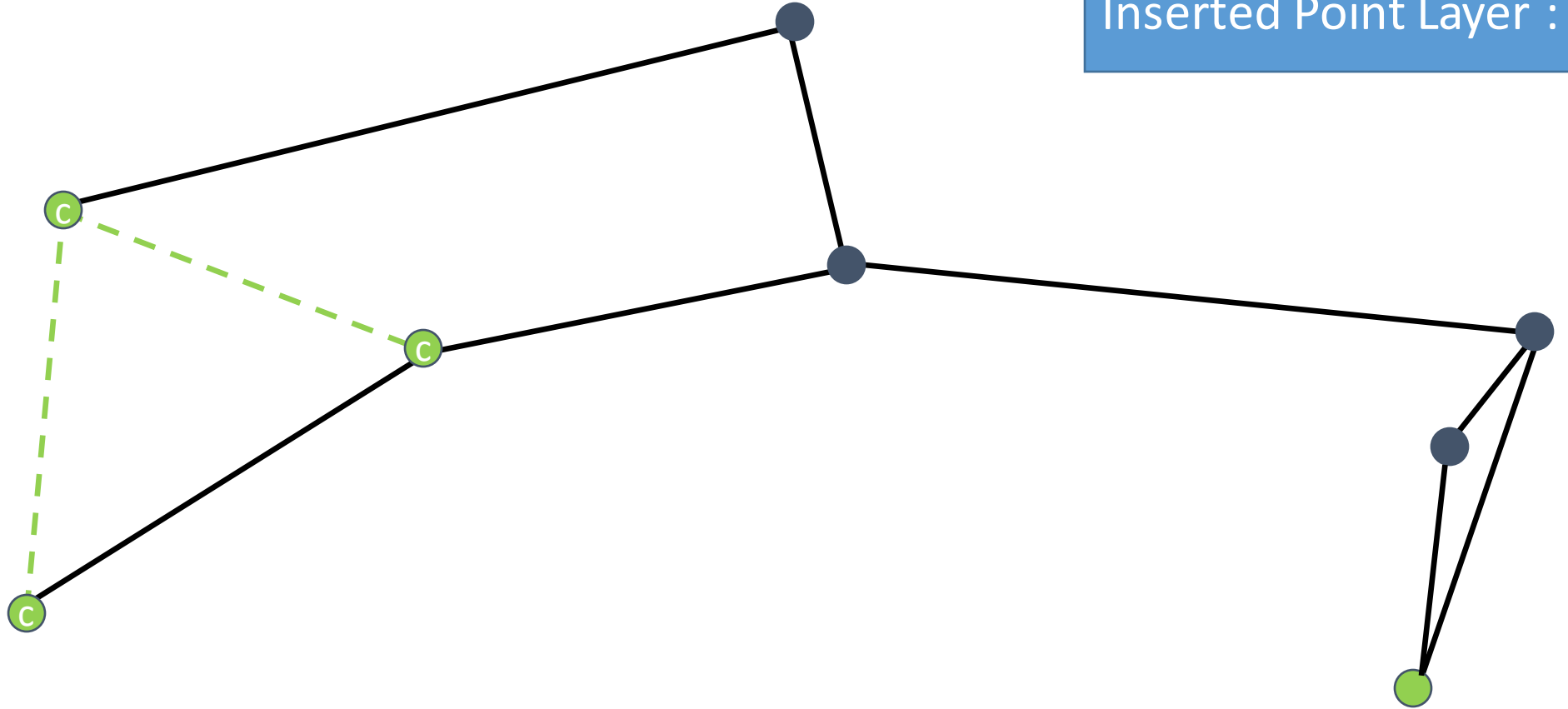
- | | | | | |
|--------|---------------------|----------------------|-----------------|---------------------|
| Legend | Point inside layer | Point being inserted | Search waypoint | Connection in layer |
| | Point outside layer | Point being queried | Root | Search path |

Current Layer : 0
Inserted Point Layer : 1



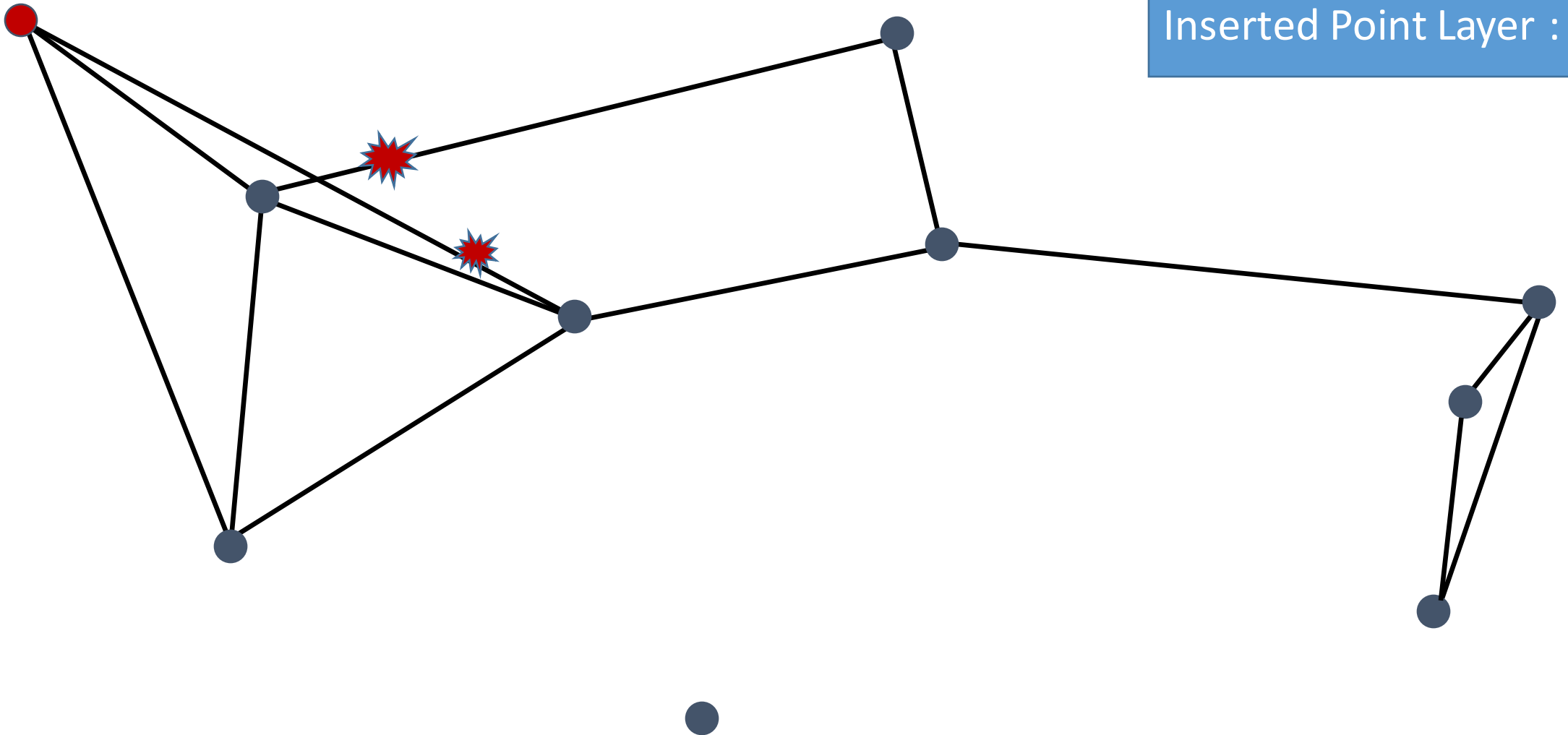
- | | | | | |
|--------|---------------------|----------------------|-----------------|---------------------|
| Legend | Point inside layer | Point being inserted | Search waypoint | Connection in layer |
| | Point outside layer | Point being queried | Root | Search path |

Current Layer : 0
Inserted Point Layer : 1



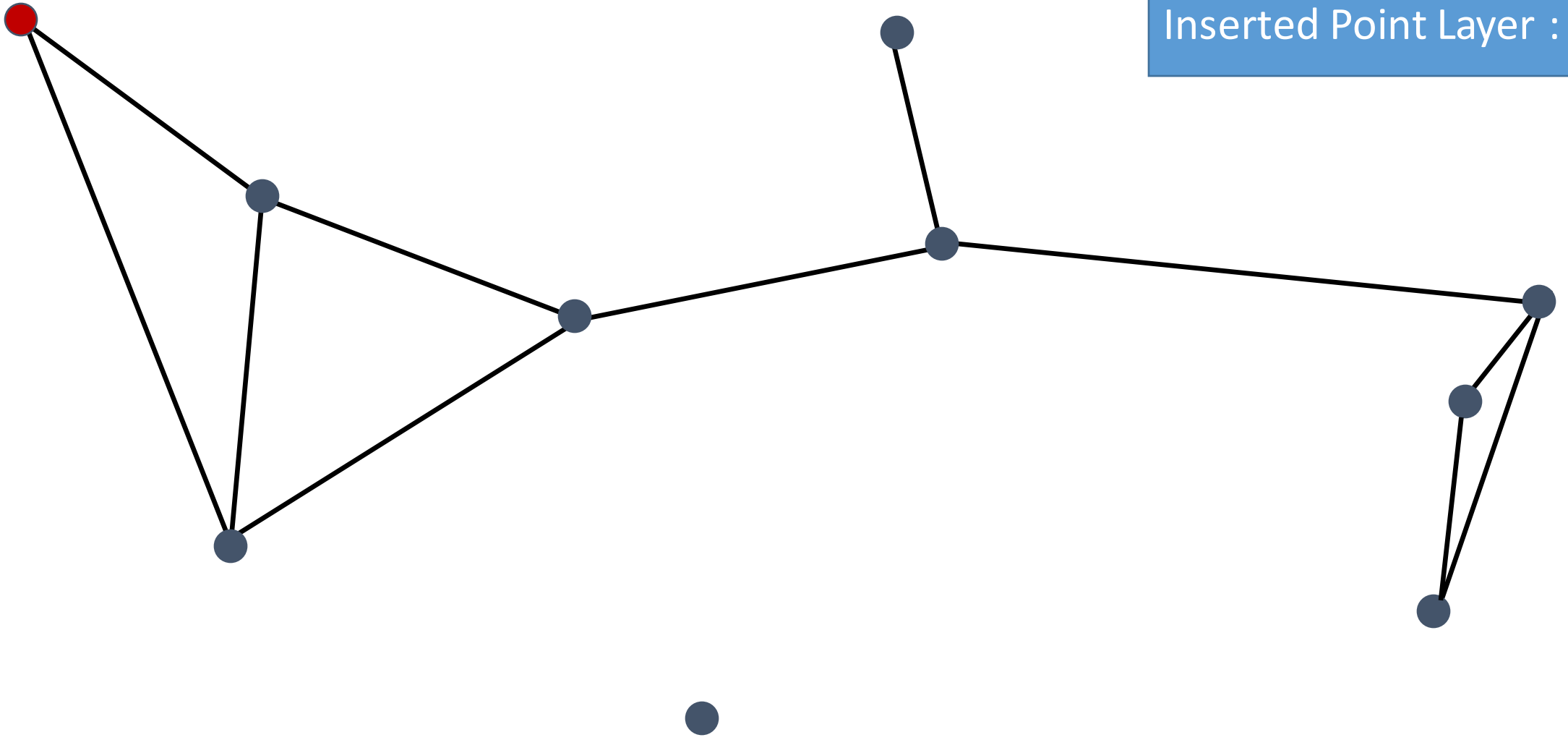
- | | | | | |
|--------|---------------------|----------------------|-----------------|---------------------|
| Legend | Point inside layer | Point being inserted | Search waypoint | Connection in layer |
| | Point outside layer | Point being queried | Root | Search path |

Current Layer : 0
Inserted Point Layer : 1



- | | | | | |
|--------|---------------------|----------------------|-----------------|---------------------|
| Legend | Point inside layer | Point being inserted | Search waypoint | Connection in layer |
| | Point outside layer | Point being queried | Root | Search path |

Current Layer : 0
Inserted Point Layer : 1

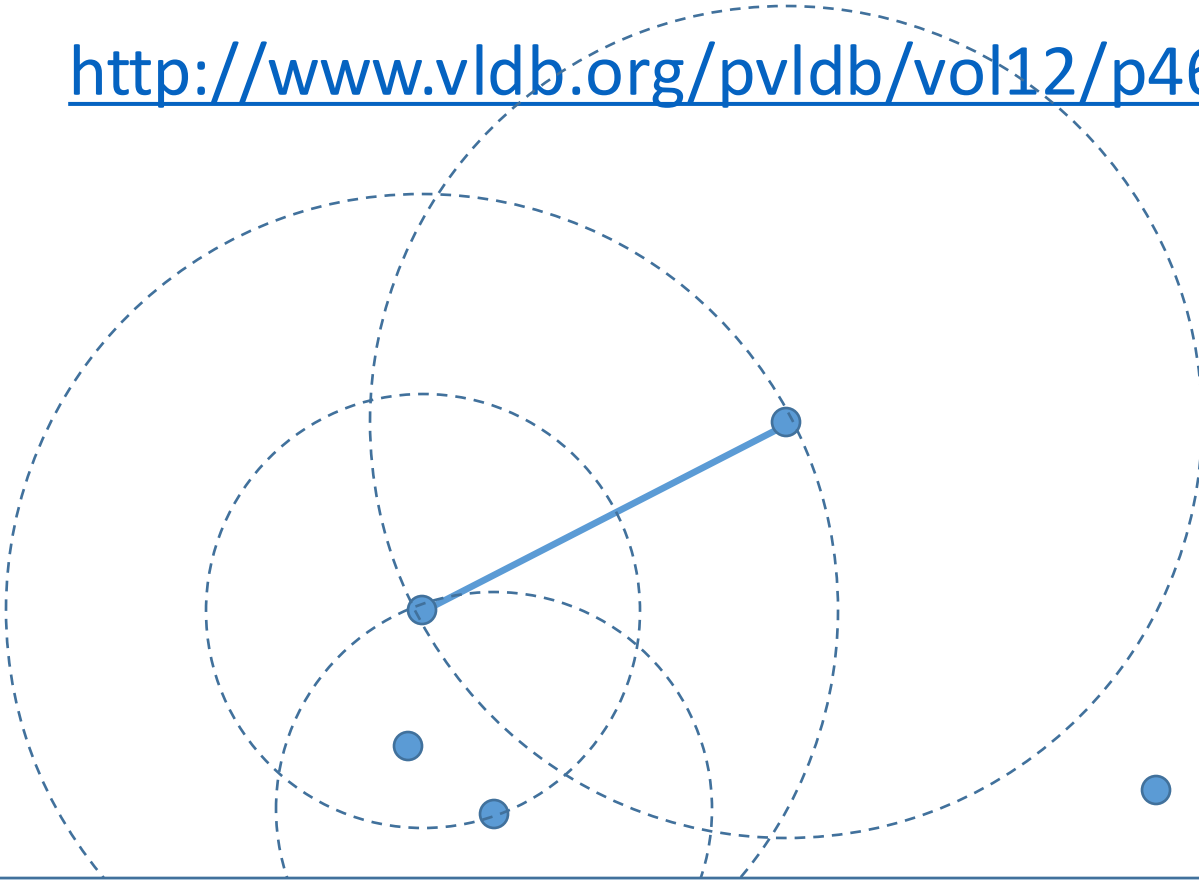


- | | | | | |
|--------|---------------------|----------------------|-----------------|---------------------|
| Legend | Point inside layer | Point being inserted | Search waypoint | Connection in layer |
| | Point outside layer | Point being queried | Root | Search path |

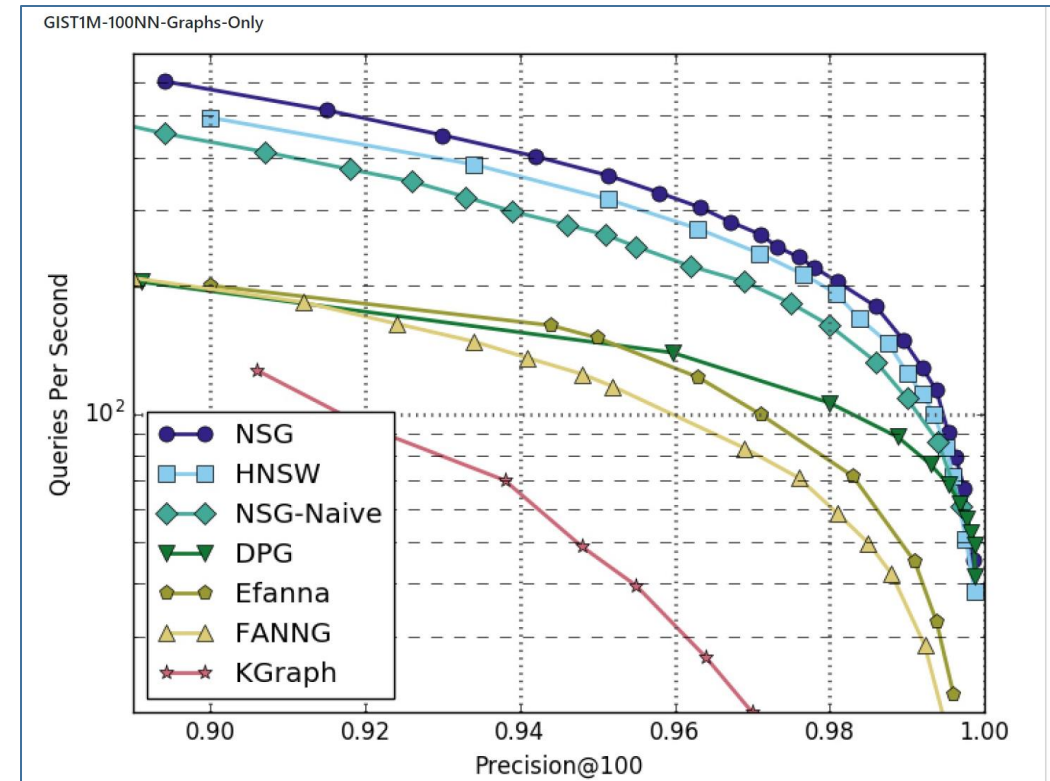
Still state of the art ?

EFANNA NSG (Alibaba)

<http://www.vldb.org/pvldb/vol12/p461-fu.pdf>



Connect p and q iif : $B(p, \delta(p, q)) \cap B(q, \delta(p, q)) \neq \emptyset$



Still state of the art ?

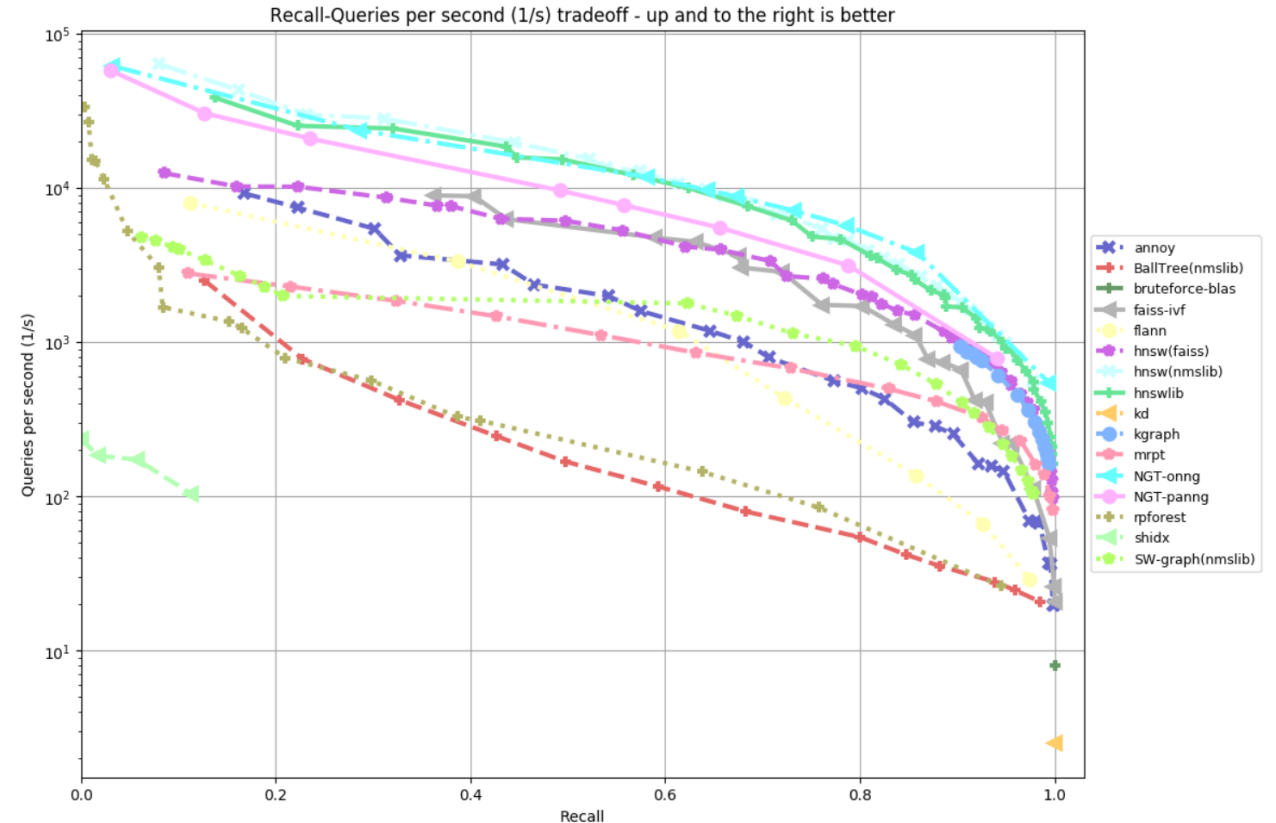
NGT ONNG (yahoo Japan)

Same principle than hnsu with extra heuristics to decide which connections to cut when a node has too many connections.

<https://arxiv.org/pdf/1810.07355.pdf>

Benchmark Results

glove-100-angular



HnswLib : hnsw implementation, very simple C library, Cpu based, uses vectorization heavily, python bindings

Faiss : supports hnsw, quantization ; python bindings (also see <https://github.com/criteo/autofaiss>)

Quantization

New bottleneck : embeddings need to be stored in **memory**.

1 billion products just can't be stored the naive way.

First solution : decrease float precision : float16, float8

Second solution : compression with centroids

