Before designing the proposed algorithm, it is a good habit to read the existing papers of relevant topic and do some analysis of existing methodology. Therefore the researcher has started an analysis on existing papers relevant to graph and subgraph isomorphism. The literature papers first are categorized into Algorithm, Application, Survey, Comparative experimental results, etc. As per the category, reviews are given below.

**Alan Chia, et. al (2015)**

In this paper [29], Authors have designed new algorithm. A brief introduction of maximum common subgraph isomorphic problem is given here. Graph technique is used for the study of mathematical structure and it contains nodes and links between the other nodes. Graph technique is used in many research areas like social networking, networking of computers and data analysis. As per this paper, graph is a mathematical structural model and this mathematical structure model contains nodes set and links set and links connect to the node pairs.

The problem is described in the form of: Input: Two graphs G1 and G2 are given as an input. Output: G1 graph is isomorphic to graph G2.

The modular product graph solution is presented by authors. In this solution, Authors have compared the combinations of nodes and edges of input two graphs i.e. G1 and G2. For solving the problem mentioned in the paper, authors have calculated the maximal clique in one graph. The conclusion of this paper is to find out the approximate maximum common subgraph isomorphic problem. Weighted graphs are generated here, which gives the probability of the link is available or not in the maximum common subgraph. The proposed method discussed here to determine two input graphs G1 and G2 are isomorphic or not. For finding the isomorphism, minimum time should be required by the system and this solution of the proposed method is used to generate more accurate results.
B. Sobota, et. al (2008)

In this paper [30], comparison of Blind search and A* algorithms is presented. The author’s goal is to develop land map representation by using graph when there is a direction searching from origin point to the end point. A graph is a set of vertices and edges and Path is made of vertices and edges. Many tools and technologies used for finding the path are also discussed here.

There are various real time examples for implementing such type of technologies. Euler, a great mathematician in eighteen century invented a graph. Several algorithms that are used for finding the shortest path from start point to end point as well as algorithms for route finding are also discussed. After performing the comparison, Authors concluded that Blind-search algorithm is faster, but it is not accurate, whereas A* algorithm is 100% reliable and provides better results. Hence A* algorithm is used in most of the technologies, but not faster.

In this paper, Authors have presented their views on a graph theory from virtual reality. They have shown in the applications of modified Dijkstra’s, A* algorithm of path-finding in the graph with a geographical information system.

Blankstein, et. al(2010)

In [31], authors have performed a simple parallelization of a subgraph isomorphism library, VFLib, which has near-linear speedup when working with random graphs. They have examined the advantages and drawbacks of several data structures when used during the backtracking search. They have also examined the several heuristics for spawning threads. Finally, they have used conditional copying to achieve near-linear speedup with the number of CPU cores on random graphs and reasonable performance on parasitic inputs.

The most important feature for this work is the cilk_spawn keyword. This keyword marks a function called as a candidate for parallel work. The first line of additional work for this paper involves removing the need for a snapshot copy in conditional copy implementation. They could do this by keeping one clean copy at the initial state and then maintaining a log of operations as they perform the search. Whenever a clone is required, they would replay the log on a clone of the initial copy. This would increase the cost of a clone but remove the need for snapshot at each
level. Authors have suggested some language features for Cilk++. It would be extremely useful to have a hyper object that would abstract the conditional copying code. This could be a splitter object. This object would magically determine when it was appropriate to clone the underlying data structure and abstract away the actual operation. This newly designed vnmFlib algorithm is an extension of Vflip graphs matching algorithms. vnmFlib algorithm allows links mapping to the path that are smaller than predetermined distance value, but Vflip is bounded to link mappings. At each step of mapping, network restraint is analyzed by vnmFlib algorithm. In vnmFlib algorithm, two important functions are described are as follows:

1. genneigh() Function
2. valid() Function

This vnmFlib algorithm is a sub graph isomorphism detection based on the capability of managing numerous restraints and online requests that are arriving dynamically. They have implemented a pattern and then comparison of its performance is done with the two stage approach.

**Boominathan, et. al (2014)**

In [32], authors have discussed the problems are used for analyzing graph theory and their applications. There are two problem areas:

1. Classical problem-defined on the basis of graph drawing, used in database design, circuit design etc.
2. Problems from applications-defined on the basis of experimental research and graph theory algorithm implementation


In this paper [33], Authors have implemented an algorithm: it gives a brief understanding about the problem that arises in subgraph isomorphism and also suggested an advance request of identifying the difficulties with it. They have mentioned the method which is represented for any systems that can alter between graphs called as a model graph or know graph. Unknown graphs are called input graphs. That advanced method is found in a concentrated prepossessing, for generation of decision trees. The algorithm of isomorphism is used to input vertices of
the given graph in which authors have used a quadratic number of vertices. Recently in preprocessing step the new algorithm is constructed and hopefully extended rapidly with vertices of model graphs.

The new algorithm is nothing but the decision tree algorithm works with the idea of decreasing the given area of decision trees. The new method is given to a computational complexity analysis. Therefore the new algorithm has advantages and disadvantages to generate unspecified graphs with practical experiments. In this paper, a graphic symbol recognition system of the algorithm is also discussed.

Two basic approaches towards graph isomorphism problems are: The first approach related to study of permutation groups and work on group theory concepts, these approaches practically design at growing and progressing procedure for isomorphism detection.


In this paper [34], authors have introduced new and advance problem detection of graph and subgraph isomorphism. It is achieved on some step in where graphs are used for creating a decision tree. Subgraph isomorphism is identified by tree traversal at run time.

The computational complication of any graph algorithm which is newly formed is a polynomial of graph vertex. The new algorithm is able to stand on one's own in any of the graphs for the model graphs and edges. Still the size of decision trees is increasing with large size. Numerous trimming techniques are developed that aims to decrease the size of decision trees.

**Bunke, et. al(2002)**

Graph isomorphism and subgraph isomorphism are used to identify if two entities are similar or if an entity is available in the dataset of various entities [35]. But, graph and subgraph isomorphism are restricted to their applications since realistic objects generally get influenced by distortion or noise and as a result, graph of similar object may not match perfectly. Therefore, it is mandatory to add some noise in the match process. One of the ways to deal with distortion is to make use of Maximal Common Subgraph as a measure of similarity. Maximum Common Subgraph (MCS)
isomorphic problem solution is applicable in different areas like Bioinformatics, social network analysis, biology, image recognition, etc. The problem that lies with subgraph isomorphism algorithm is that it takes more time for the execution on larger graphs. Hence there are more challenges in Science and technology.

Following are the proposed methods based on Maximal Common Subgraph:

- Basic idea of proposed system
- Mapping of graph nodes
- Weighted graph creation

Performances of Maximal Common Subgraph are

- Implementation and testing of method in Java
- JUNG library is used for graph modelling and calculation of measures
- The results of this MCS are compared with existing subgraph isomorphism matching algorithm
- MCS is able to generate solution instantly and immediately

Bunke, et. al (1998)

In this work [36], Authors have shown various graph applications like social networking and analysis of chemical compound. Some examples related to graph application in real time environment are given below:

- Graph technique to represent chemical compound structure in their reaction process
- Graph technique for representation of traffic network

The goal of this paper is to present light-weight structure for filtering and encoding method novel for the approximate subgraph search problem. Subgraph search is one of the best tools for finding sub-structure in the graph databases.

Some of the points are given below for matching continuous subgraph search:

- Graph, labeled graph is a set of vertices, edges and labels
- Sub-graph, its vertices and edges from subgraph
- Subgraph Isomorphism
- Graph change operation
- Sub-graph search over the graph
Random graphs are selected from 10,000 datasets. Average number of nodes are selected here is 24.8 and the average number of edges are 26.8. Authors have used six sets of queries and each set of the query contains 1,000 queries. Comparison of technique is performed with the help of existing algorithms: GraphGrep and gIndex. Experimental results of this paper are given below:

- The proposed method is more effective than gIndex method and GraphGrep method on a different dataset
- Experimental results on all types of dataset was also done extensively


A faster algorithm is described in the graph as well as subgraph isomorphism that is based on the intensive pre-processing step in which graph database is converted into decision trees. Nowadays decision tree approach is extended from an exact graph and subgraph isomorphism for error tolerance graph matching [37].

There are various approaches: in the first approach, error correction is taken into consideration at the time of creating a decision tree, whereas in the second approach, error correction is taken into consideration at the run time. With the current developing enthusiasm on graph databases, it has turned out to be more critical to have a quick answer to the issue of subgraph isomorphism, or finding related patterns in large graphs. The subgraph isomorphism is the problem of detection of the input graph in the target graph.

Nowadays, many complex objects, biological structures, chemical compounds and social networks are modeled as graphs. Many existing applications in chemistry, software engineering and Bioinformatics require effective and efficient management of graph structured data. The first practical algorithm for subgraph isomorphism for graph search was proposed by Ullman.

Some of the serious problems that encountered in the current practice of experiments are listed below:

- The algorithm which has not been described in a common framework, it is difficult to compare
- Comparison of research work has not been compared logically
Changliang Wang, et. al (2009)

In this paper [38], various applications of graphs are discussed. Detection of subgraph isomorphism problems are arising in many applications. These applications deal with data modeled as graphs.

Some existing algorithms came forward to solve this problem. This is a sensible dataset which uses several trimming rules, join orders and additional neighborhood information. There are more problems reported where comparison is often done using the original author’s binaries which are written in the different environments. Some serious problem by implementing or re-implementing algorithms in a common code base and their query loads is also discussed here.

Contribution of paper:

- Differences of other existing algorithms using a common framework
- Implementation or re-implementation of five state-of-the-art algorithms in the common code base
- Fairly and logically compare these algorithms using many actual and synthetic datasets
- Some experiments are explored in depth in order to figure out why other algorithms outperform for specific queries and data graphs
- Generic subgraph isomorphism algorithm

Authors have used four real-world datasets which were referred to as NASA, Yeast, Human and AIDS. The datasets used here possess their own characteristics.

- The AIDS datasets contain a set of sparse graphs where the unique label number is not so big and the average graph size is small
- The NASA datasets contain a set of trees where the unique label number is greater than AIDS whose average tree size is 32.2
- The Yeast datasets contain one large graph with 12519 edges and 3112 vertices

Chen, et. al(1996)

In this paper [39], Authors have implemented the multilevel parallel concept for identification of pattern in graphs. Basically graphs are used for representation of diverse data and the linking between the data.
Nowadays usage of graphs is rising higher and higher since many research analyzers are trying to figure out diversity of issues in various areas such as data mining, classifying chemical compound, biological computation, link analysis.

The graph size goes on increasing corresponds to data volume growth received from the sensors. Techniques required for computation of such larger graphs should be fast enough for typical usage and therefore, parallel approach is implemented in the existing system.

A new approach of parallelism is designed on the basis of Kuramochi and Karypis’ VsiGraM algorithm. A huge amount of inherent parallelism is accomplished in the algorithm. But two parallel levels and iterative nature do not plan slightly to the existing approach.

By using OpenMP along with the algorithm, effective results are achieved in terms of time.

**Coen Bron, et. al (1971)**

In [40], authors have described a number of techniques to find the complete subgraph from given undirected graphs. Two backtracking algorithms are introduced here, which make use of bound and branch technique.

- Implemented the basic algorithm where small groups in alphabetic order are generated
- Second version is extended from first it generates small group in a changeable order that makes an effort to reduce the number of branches to traverse

Version 1 algorithm: This algorithm contains 5 different steps

- Selection of the candidate
- Adding the candidate to compsub
- Create new sets candidates by dropping other points which are not connected to the particular candidate
- Call extension operator to work on sets
- Remove the selected candidate from compsub

compsub behaves like a memory and can be enabled to continue and updated in the form of a standard array. The operator then announces an array in which the new sets
are built up that are handled to the inner call.

Version 2 algorithm: This algorithm is based on the given factor:

- Algorithm of Bierstone

Cook, et. al(2006)

P class problems can be solved in polynomial amount of time by Deterministic machine [41]. P is set of all decision problems which can be solved in polynomial time by a deterministic. Since it can be solved in polynomial time, it can be verified in polynomial time. Therefore, P ⊆ NP.

This class includes those languages which are recognised by some deterministic Turing machine in polynomial time. Deterministic problems are those problems which have only one possible solution from initial state to final state.

Any decision problem which can be solved in polynomial time is called as P class problems. Yes or No problems are called as a decision problem. P Problems are easy to solve.

Example: Find whether a given number is even or odd.
If it is divisible by 2 then the number is even else the number is odd.

NP means one can solve problems in polynomial time if we can break the normal rules of Step-by-step computing”.

Decision problems are solvable in Non-Deterministic polynomial time. It consists of those problems that are verifiable in polynomial time.

Deng, et. al(2012)

This paper [42] focuses on various applications. Fuzzy technique is presented in this paper used for finding the shortest path in the routing algorithm. Fuzzy Routing algorithm (FRA) is used to reduce path request blocking and also used to increase the overall utilization.

Fuzzy technique has been used since this technique generates approximate fixed and exact result.

Two important points necessary in shortest path algorithm are:
1. To find addition of two edges
2. To compare distance between two different paths and their respective edge
Graph theory is used in many applications such as transportation, communication and social media. The path finding algorithm is also used for handling special characteristics of the transit network. Proposed method presented the process which finds the minimum distance between initial to final node. The final result of shortest path is generated with respect to cost, time and distance. This method is widely used in many applications; one of the examples is efficient network.

**Dennis Shasha, et.al(2002)**

In this survey paper [43], Authors described algorithms and applications that are related to tree and graph searching. There are various approaches for tree and graph queries. Two main approaches are discussed here:

1. Approximate Embedding Queries
2. Selectivity Estimation

The proposed algorithms are:

1. ATreeGrep algorithm: Key tree searching algorithm
2. GraphGrep algorithm: Key graph searching algorithm

Comparison of Ullman and Nilsson Algorithms are discussed by authors. Key graph searching refers to a graph and matching in the data graph. The graph and subgraph matching algorithms have used key-graph searching method. For minimizing the search space, indexing technique is used in large databases. Authors have evaluated the performance of GraphGrep algorithm. Mathematical operations are organized by authors who has NCI database containing 16,000 molecules and used Linux workstation with the 1GHz Pentium III processor.

The NCI database has 20 nodes and several graphs up to 270 nodes. Query time for query sizes 13 to 189 nodes, database sizes 1,000 to 16,000 graphs and lp values 4 and 10. As per the conclusion, objective of this paper is to match and search the trees and graphs.

Future work was suggested by Authors are:

- Development of framework for selectivity estimation for queries
- Turning searching to pattern discovery in trees, graph searching,
- Meaningful distance measures
- Algorithm for approximate query processing.
Deshpande, et. al(1986)

A binary search tree is a binary tree [44]. A node’s left child must have a value less than its parent value and the node’s right child must have a value greater than its parent value in binary search tree. Operation on a Binary Search Tree:

Initialize operation: Initially the tree is empty and hence the referring pointer root node should be set to null. Insert operation: Very first insertion creates a tree. After that, whenever next node to be inserted, first locate its proper position. Start searching from the root node than if the data in less than the key value, search for an empty location in the left sub tree and insert the data. Otherwise, search for the empty location in the right sub tree and insert the data. Search Operation: Whenever an element is to be searched, start searching from the root node, then if the data in less than the key value, search for an empty location in the left sub tree and insert the data. Otherwise, search for the empty location in the right sub tree and insert the data.

Make empty operation: This operation deletes every node of the tree. It also releases the memory occupied by the node. Delete operation: Whenever, the delete operation is needed to perform. We must find the node to be deleted can be a root node (A node with one child, A node with two children). If the node is a leaf node, it can be deleted immediately, by setting the corresponding pointer to null. Even when a node with one child has to be deleted, it can be deleted easily. For e.g. If the node has to deleted is ‘q’ and it is the right child of a parent node ‘p’ then, the only child of a will become the right child of p after deletion of a. Similarly, if in a left child than the only child of a will become the left child of p after deletion.

The height of a binary tree is the largest number of edges in a path from the root node to the leaf node. A binary tree is said to be a full binary tree ,if each of its nodes has two children or no child at all every level is completely filled. A binary tree is said to be a complete binary tree if all leaf nodes are at level n or n-1 and levels are filled from left to right. A skewed binary tree is skewed either to the left or to the right. In left skewed binary tree almost nodes have left child without corresponding to right and similarly. In case of right skewed root nodes will have a right child without corresponding to left. The preorder transversal of a non-empty binary tree is: First, visit the root node. Next, transverse the left sub tree. At last, transverse the right sub tree. Post order Transversal: The post order transversal of a non-empty binary tree is:
First, visit the left sub tree. Next, transverse the right sub tree. At last, transverse the root node of a tree.

**Emden Gansner, et. al(2006)**

dot is used to draw the directed graphs in hierarchical form and is divided into four phases: This helps in cause to comprehend what sort of formats does the dot makes and how can we control them [45].

The format technique utilized by dot depends on the graph being non-cyclic. In this way, the initial step is to break any cycles which happen in the input graph by turning around the interior bearing of certain cyclic links.

The second stage relegates vertices to discrete positions or levels. In a top-down, positions decide Y values. The next step orders nodes inside positions to keep away from intersections.

The final phase is to decide X values to minimize the number of edges. According to Warfield this is an indistinguishable general methodology of most progressive graph drawing programs

**Evgeny B. Krissinel, et. al(2004)**

This paper [46] states that backtracking algorithms for the Computer Science and Information was developed 20 years ago is rarely used. Improved backtracking algorithms for CSI which is different from its predecessor is used for better searching strategy and hence are more efficient. Authors found that there are new algorithms that outperforms as compared to traditional techniques of backtracking algorithms. It states that knowledge is useful in the construction of an improved backtracking CSI algorithm. This algorithm is initially derived from an Ullman algorithm by making modifications of the part related to the expansion of partial solution and also elimination of improper branch of the search tree. This algorithm differs from backtracking CSI algorithm which has a control complexity of the parameter. Computational complexities of CSIA are bounded by similar range. In CSI Algorithm, there are most of the applications that are sufficiently larger common subgraph and considered for the useful results of graph matching.

As per author's approach, CSIA algorithm is one of the most efficient
algorithm for finding larger sub graph. This CSIA algorithm is useful for the purpose of serving structure query in the EBI-MSD legend database, EBI-MSD protein 3D comparison service SSM, CCP4 Molecular Graphics. The Efficient backtracking algorithm is described in this paper for using detection of common subgraph isomorphism. This algorithm is considerably most efficient result for any other algorithms which are based on the maximal clique’s approaches. This is the newest algorithm named as CSIA and is inherited by the space and time complexity.

Foggia, et. al(2001)

In the paper [47], Authors have compared Ullman algorithm, VF2 algorithm and nauty algorithm of graph and subgraph isomorphism algorithms. Algorithms are designed to enhance the complexity and better performance on large graphs. Algorithms are compared by using the same datasets of different types as given below.

1. Randomly connected graphs
2. 2D mesh graphs
3. Bounded valence graphs (BVG)

10,000 isomorphic graph database used here for testing the performance of an algorithm.

VF2 performance

- VF2 algorithm is capable to perform on 56 out of 100 combinations of the graph
- In 44 cases, nauty algorithm is better than VF2 algorithm.

Finally, this paper concludes that VF2 algorithm is more efficient for large graphs. Matching time of VF2 algorithm is good as compared to Ullman algorithm and Ullman is one of the best subgraph isomorphism algorithms for smaller graphs.

Gao, et. al(2010)

Edit operations such as insertion, deletion or substitution is performed so that errors are corrected that may have crooked the graphs [48]. Edit operation is applicable to edges and nodes as well. Transformation of graphs G into G’ by shortest sequence of edit operation is known as edit distance. Edit distance is applicable to calculate, graph
similarity. This transformation also denotes a mapping from nodes of G to nodes of G' and less effort are taken to correct distortions. In real time applications, some of the edit operations are more concerned than other operations. Cost is assigned to each definite edit operation and this is known as cost function. Basically edit operation is inversely proportional to its cost. In some papers, study is performed on the relationship between graph edit distances. Since these studies are not suitable for real time issues, but on the contrary, they are helpful in achieving better theoretical knowledge of graph matching. A huge number of algorithms possessing various characteristic are available for matching of the graph.

Garey, et. al(2002)

Decision problems are the problems where the answer is either Yes or No. Such problems are also referred as Recognition Problem [49]. A decision problem is normally formalized as the issue of choosing whether a given string belongs to some predefined set of strings, also called a formal language. The set contains precisely those answers whose answers are "Yes". If the algorithm is able to correctly decide for every possible input string whether it belongs to the language, then the problem is called decidable otherwise it is called undecidable.

Following are some important observations about:

All types of problems that exist in the world can be categorized into two groups based on the complexity of solutions. The first group of problems is the set of problems for which the polynomial time complexity solutions exist and they are categorized into a class called polynomial time complexity problems. Polynomial complexity means the time bounded solutions by polynomial of small degree. Example: $O(\log n)$, $O(n)$, $O(n^2)$, $O(n^3)$ etc.

To state a few examples of such problems are searching like: Linear Search, Binary Search, Sorting algorithms etc. which have polynomial time complexities.

The second group of problems is the set of problems which have the non-polynomial complexity solutions instead of polynomial time complexity solutions and they are categorized into a class called as Non Polynomial time complexity problems. Non-polynomial means, the exponential time complexities of the solutions that exist. Example: $O(2^n)$, $O(n!)$ etc.
To state: Hamiltonian Cycle is one of the examples of such Non-polynomial time complexity class problem.

**Gibson, et. al(2004)**

Pattern Recognition is done with the help of model forming [50]. The main advantage of these applications is to verify the pattern of one graph is presented in other graphs or not. These formal models are used to represent domain knowledge and all possible inputs and outputs.

- **Pattern**: It can be any event or object
- **Pattern Class**: It is a set of objects that shares the common features and generated from a common source
- **Pattern Recognition**: It is a classification process in which objects or patterns are classified
- **Classifier**: It is a software module that performs all the operations

**Advantages:**

- Classification and fake biometric detection problems are easily solved by the pattern recognition
- The pattern recognition system is useful for visually impaired people
- It is useful for speaker diarization

**Disadvantages:**

- Larger dataset is required for better performance
- It cannot clarify why a specific object is perceived
- Implementation of syntactic pattern recognition is a lengthy and complex process

**Han, et. al(2011)**

It is the intelligent and effective method to discover a meaningful data which is an important sub area of Computer Science [51]. The ultimate reason of the data mining is to abstract material derived from a random dataset and converting that into an understandable structure for auxiliary applicability.
Since ancient era, the scientist has collected the art facts, signs and other random source of knowledge. After the data collection they have applied different methods of mining, which was developed the meaningful pattern for their required applications.

The manual mining of the pattern has been happening for centuries. The first systematic data pattern identification was introduced in the Seventeenth century through Bayes Theory whereas the Regression Analysis was introduced in the Eighteenth century. After the introduction of computers, there has been a drastic rise in data manipulation, data gathering and storage capability. As per the growth in complexity and size of the dataset, the automated processing of data has been increased with the enhancement of Computer Science. Some new methods like cluster analysis, genetic algorithm, neural network, support vector machine, decision tree, and decision rules have been introduced in recent years/decades.

These methods have predominantly been used for finding hidden patterns in a huge data set. New Object-Relational Database Management System, has added new features and application’s speed for these algorithms in data mining using software like ORACLE, MYSQL Server, Sybase, etc. Here are some aspects which involve a modern data mining.

**Hand, et. al(2007)**

Data mining is a step by step method of knowledge discovery in a database which has a random pile of data. In contemporary times, the data mining is used in business houses, research institutions, political analysts, defence sector, agriculture sector and financial analyst for their Decision Support System where in the data miner (which could be a software or hardware) gather the data from different sources of knowledge and converts that into certain patterns for making particular decisions.

Nowadays, the real time data mining happens by using automated or semi-automated software. These software analyses large amounts of data from different known, unknown sources and find out the interesting patterns. These patterns later are processed as an input data in the DSS System to predict an accurate result.

For example, if you are a frequent shopper and regularly go to a certain store for shopping then all of a sudden you will start getting SMSs, or alerts on your mobile
on your personal mail ID for some sell or a new stock which is particularly of your interest.

**J. Brian Burns, et. al(1992)**

In this survey paper [53], the author’s goal is to search and match for 3-dimensional object from multiple object. Multiple object libraries are enhanced by developing projection definition of objects from expected view.

Number of selected 3-dimension matches for authentication purpose should be diminished to make the system efficient. A Compelling approach for matching of view, the description network for image is discussed here.

As per from experiments, the new matching system was adopted here for recognizing objects in digital and real images. For every image, the same process of initialization was proposed. Experiments have shown that the recognition system is based on view description network which is able to find the precise match to 3-dimensional object in complicated image that too with probably great performance level.

**Jens Lischka, et. al(2009)**

In [54], authors have proposed a new algorithm – Virtual network mapping algorithm which is based on detection of subgraph isomorphism. The goal of visualization is to execute numerous distinct applications on the equivalent shared physical resources. Virtual network is constructed by virtual network mapping, to make valuable use of underlying resources.

In this paper, authors have also proposed an additional backtracking algorithm. This algorithm is designed on the basis of the subgraph isomorphism search method which maps, links and nodes at the time of the same stage.

Experimental results shown in the paper are: subgraph isomorphism based outcomes were improved and were superior to the previous two stage approach. Authors have discussed about the following terms:

1. Network
2. Virtual Network Mapping
3. Virtual Network Mapping Costs
4. Virtual Network Request
5. Residual Graph

**Jinsoo Lee, et. al(2013)**

In [55], authors have advised a query language for the graph database that supports any type of attribute on node edge and graph. A graph is a primary element of information in the query language. Each query is capable of managing more than one graph. The notion of the formal language of graph domain is carried out.

They have started with extending graph algebra from relational algebra. In the selection, operator is generalized to graph arrangement pairing and composed operators for rewriting the matched graphs.

The graph query approach is used in various domains listed below:

- Finding a complete heterocyclic chemical compound that contains aromatic rings and side chains. Both the rings and chains are indicated as a graph with the atom as nodes and bonds as edges
- Finding complete protein structure

The concept of GraphQL is proposed in this paper. GraphQL is an abbreviation of Graph Query Language and graph QL is used for “graph pattern for the basic operation unit.”

Authors have proposed their work as a formal language for graph this language is useful for managing a graph. This language supports graph as the basic unit of information.

**John W. Raymond, et. al(2002)**

This paper [56] comes under the category of maximum common graph algorithm for matching chemical structures 2D and 3D graphs. It provides a review and classification of maximum common graph algorithm.

**Algorithm1: Exact Algorithm** - This algorithm is designed for calculating an exact solution of the maximum common graph.

**Algorithm2: Maximum Clique-Based Algorithm** - Maximum number of clique are reduced that depend on graph capability. In image processing, graph compatibility is also known as graph association, in the maximum clique algorithm, maximum
common graphs are disconnected.

**Algorithm 3: Backtracking Algorithm** - This algorithm makes use of first two algorithms for iteration of backtracking procedures, the two important algorithms used here are: McGregor and Wong, these both algorithms reduce the number of the backtracking instance by getting the possible solutions in depth of search, this algorithm is different from other maximum common edge algorithm which makes use of dynamic programming for maximum common edges subgraph problem, in mathematical programming, methods are called as dynamic programming

**Algorithm 4: Approximate algorithm** - This algorithm is responsible for finding the computational difficulties related to maximum common subgraph within compatible time complexity, it is not suitable for better performance.

**Algorithm 5: Genetic Algorithm** - Genetic algorithm is one of the best algorithms used for maximizing the objective function, chemical graphs are used in Genetic Algorithms, Parallel Genetic Algorithm used for calculating the maximum common edge sub-graph in FPGA graph, some techniques are used for saving time, they are - Combinatorial optimization, fragment storage and Adhoc procedures.

**Algorithm 6: 3D-Specific Algorithms** - This algorithm has used some techniques which are problem prediction and screening procedure.

**Jose, e. al(2011)**

**Basic operations on queue are [57]**

Enqueue: Adding on an element in the queue is known as enqueue
Dequeue: Deleting elements from the queue is known as dequeue,
isFull(): Checks is queue is full or not,
isEmpty(): Checks if the queue is empty.
Whenever, we want to perform add i.e. enqueue operation. First, we check whether the queue is full if is full then the operation cannot be performed. It is not full then the rear will be incremented by 1 new element will be inserted in the queue. Whenever, we want to perform dequeue i.e. delete operation, we check for empty condition. If the queue is empty then the element cannot be deleted and if it is not empty, then front will be incremented by 1 to the element i.e. to be deleted.

**Types of Queue are:** Linear queue, Circular queue, Dequeue (double ended queue),
Priority queue.

**Applications of Queue:**

- All types of customer service like railway ticket reservation, airline ticket reservation etc. designed using a queue to store customer information
- Job Scheduling is done using a queue
  
  a) First Come First Serve
  
  b) Shortest Job First
  
  c) Round Robin

**J.R. Ullman (1976)**

The fastest general algorithm to find a subgraph isomorphism is a matching algorithm developed by Ullman [58]. Ullman algorithm found an improvement and made some method which reduces backtracking procedure. Some difficulties of the algorithm are described. The fact that all matrices generated at run time, which are represented a subgraph isomorphism. The algorithm of backtracking is necessary to run in the dead end.

Decision tree based approach to overcome a backtracking problem at run time is also discussed here. Some methods for detection of the graph and subgraph isomorphism are discussed. The theoretical result of the trimming technique is studied in experiments with unspecified graph generation. Ullman algorithm returns a match-set which is the set of vertex pairs in the subgraph isomorphism. The algorithm finds the match set by translating the problem into a search.

Potential vertex pairs incrementally added to a match set until either a subgraph isomorphism is discovered or no more potential pairs exist. Because of the potential for failing paths in the search tree, this search requires the use of backtracking. Because Ullman algorithm uses backtracking search hence partial matches can cause searching to be computationally expensive. This also leaves many opportunities for parallelism, as any branch in the search could be handled as a new thread of computation. The subgraph isomorphism problem has a number of applications. In chemical engineering, it is used to find particular chemical structures in large molecules.
Kuramochi, et. al(2001)

The term discovery is used for describing their process, since semantically same nodes are obtained by data mining techniques [59]. Authors have designed and implemented a search framework for filtering and verification for achieving higher efficiency and adaptability. Analysis of heterogeneous information networks and mining plays a key role in various large scale systems.

The fast evolution of extracting social and information network are analyses through several on-line databases. A deep study has been done on various problems that includes clustering, ranking, search of entity similarity, ranking and prediction of relationships in an information network. In this paper, a newly composed discovery system is proposed: Subgraph query is given as an input and a list of sub-graphs are extracted from information networks which satisfies query criterion.

Two query types are described for searching and processing purpose:

- Xquery
- AQUA Query

A Tree Grep is searching and filtering algorithm is implemented in XML search engine and is called as XML query by example. XMLs are detected in an XML database that approximately match or contains the query tree. The example XML query and matching XML that contains query are displayed using the internet explorer browser. The query tree and XML matching tree are displayed with the help of Java applet and the matching portion of the XML tree are marked with bullet.

Nasrabadi, et. al(2007)

Pattern recognition is used for recognition of patterns and identifying the hidden information in the data [60]. A person can differentiate between the sounds, numbers, and characters, but in the case of Computer, it is difficult for a computer programme to differentiate the different items.

In any case, if there should be an occurrence of a computer programme, it is extremely hard to take care of this kind of perceptual issue. Application potential is an important aspect of pattern recognition:

- Familiar patterns are easily recognised by the pattern recognition system
• Unfamiliar objects are characterized by the system
• Shapes and objects are distinguished by the system from the various angles
• Hidden parts are recognised by the system

Nowadays applications like pattern recognition make use of the concepts of graphs and sub-graphs isomorphism. Likewise, these concepts are used in many applications.

Some applications are listed below:
• Graph Grammars
• Semantic Network
• Chemical Structures

**Navarro, et. al(2002)**

Pattern matching is used for checking and situating the particular arrangements of information of some pattern among raw data or a sequence of tokens [61]. Pattern matching is more accurate matching system than a pattern recognition system. Pattern matching is an essential paradigm of different programming languages. Examples of patterns are tree structures and sequences, etc.

Facebook, twitter and linked-In are using pattern recognition or pattern matching for social communication and also find social position during run time. Pattern matching has some problems in the social networking because real time of graphs is too large.

In Facebook, millions of users (called as a node) and billions of links (called as edges) creates very large graphs. To perform the pattern matching algorithms on large graph, specialized algorithms are created.

**Nian Zhang, et. al(2003)**

In this paper [62], Authors have used fuzzy attributed graph (FAG) to solve the problem of extraction of sub-circuit. This FAG approach is used widely in various fields such as image understanding, recognition of pattern and fuzzy graph match problems.

Extraction of a sub-circuit is one of the most crucial applications in several areas of CAD, specially for the advance design of Very Large Scale Integration
Every year the design of VLSI draws larger and larger hence semiconductor industries need to ensure that the manufacturing chips should be defect free.

For solving this extraction question, various methods have been already designed. Some of them are given below:

- Logical tree representation of circuit was used by Yokomizo
- Partitioning and relabeling algorithm for recognition of sub-circuits are used by SubGemini
- For identification of sub-circuits, partitioning and genetic algorithms are employed by SUBGEN
- Decomposition of circuit and partitioning is used by Ling
- Recursive scheme is adopted by DECIDE algorithm for achieving identification operation

Tsai and Fu introduced this fuzzy attributed approach of the graph. It represents structural patterns forthrightly. Terminology such as fuzzy attributed graph and core nodes, goal nodes, free nodes are also defined in this paper.

- FAG matching algorithm for extraction of sub-circuit
- Similarity of FAG pair

As per the experimental results, FAG algorithm is very productive techniques for implementation of sub-circuit extraction issue

It is able to detect all the occurrences of pattern circuits. NAND gate is used as their pattern circuits

**Niusvel Acosta-Mendoza, et. al(2011)**

In [63], Authors have invented two methods for subgraph mining: Exact matching and approximate matching. Exact matching is used for identifying two graphs are exact same or not. But in case of other problems, this approach is not good enough for better result, hence it becomes a tedious job so alternate method was used to find the graph similarity on the basis of approximate approach.

A new method is designed by Authors and used quad-tree for representation of the image in the tree form. The quad-tree involves division of an image in 4 same-size quadrants. Every quadrant further can be divided into four sub-quadrants and so on till the limit for splitting is met.
Then, after gathering image quad-tree, a graph is generated for representation of image with the objective of controlling the constructural information. In this paper, Authors have shown the effectiveness between exact algorithm vs. approximate algorithm and approximate graph mining algorithm on the classification of an image. Pattern identified through this algorithm allows the linguistic variations that are not treated by exact approaches. It allows the variations in position of objects that cannot be considered by APGM, because it considers linguistic variations on vertex labels.

Another proposed solution was also invented for a collection of images, therefore classification results were obtained by APGM. Exact algorithms are lower than APGM via VEAM by all angles.


Databases are structured to facilitate the storage, retrieval, modification, and deletion of data in conjunction with various data-processing operations [64]. Generally the data are arranged in such a way that it supports processing of the required information. It is stored as a file or a set of files on magnetic disk or tape, optical disk, or some other secondary storage device. The information in these files may be broken down into records, each of which consists of one or more fields. Fields are the basic units of data storage, and each field typically contains information pertaining to one aspect or attribute of the entity described by the database.

Records are also organized into tables that include information about relationships between its various fields. Although the database is applied loosely to any collection of information in computer files, it can provide with referencing capabilities. Using keywords and various sorting commands, users can rapidly search, rearrange, group, and select the fields in many records to retrieve or create reports on particular aggregates of data. The end user cannot directly get access to the database instead it requires some system that can act as an interface between the end user and the database. This system is called as the Database Management system.

**Peixiang Zhao, et al(2010)**

SPath algorithm is proposed in the paper [65] to solve the graph query in large
networking and makes use of graph indexing method to graph query problem in large network. In this paper, a new method is proposed to efficiently find graph structures in large computer network, graph indexing mechanism for high performance. Graph query is one of the hard problems to identify if both the graphs are same or not. Reconstruction of shortest path and orientation in big network is the key role of SPath.

Optimization of graph in large network involves three frameworks which are given below:

- Pattern based graph indexing
- SPath – new graph based indexing technique
- Query graph optimization

Pattern based indexing performs their operation in three different steps:

- Base Line Algorithm
- Graph Indexing
- Pattern Evaluation Model

SPath is path base indexing techniques used in a big network. It performs two steps:

- Define neighborhood structure of the graph
- Implementation of SPath

The advantages of graph query processing are:

- Processes the query
- Optimize graph in path-at-a-time

Query processing works in three different steps

- Composition of query
- Path selection and joining
- Path installation


Importance and applications of graph theory is discussed in paper [66] explains the. In 1735, concept of graph theory was applied for creating the path of bridge in Koinsber city.

Graph theory has various applications in different type of areas such as
chemical, networking, electronics, medical, computer science etc.

In the chemical field, graphs are created to represent the properties of chemical compounds and also used to compare the properties of two chemical compounds.

In computer science field, complex problems can be solved by applying algorithm on different types of graph. This paper has proposed the process of solving natural problems by using graph theory.

Robison, et.al(2013)
Cilk++ language could handle the process of snapshot and logging [67]. Contribution of the project as follows:

- This project is compared the performance of several data structures for sub-graph isomorphism
- It implemented a faster match than current state-of-the-art sub-graph isomorphism match algorithm.
- It also gave a detailed a general approach to dealing with large data structure copying for spawns in cilk platform
- It speculated on useful language features to enable conditional copying in cilk.

Rusu, et. al(2010)
Organization of structural description into network is a vital step for search optimization [68]. Distinguishing recursive index of description in networks for the usage of interpretation tree is mandatory, since the recursive indexing approach sustained via network possess contrast performance that aids much more competency. An essential approach for recognition of 3D images of 2D images is a view description.

Recursive indexing is a useful approach for recognition of objects on the basis of description network. In a recognition system, initially line segments are detected in the image and further developed into encouraging matches to the simpler 2 dimensional model nodes available in the network.

The following steps are to be executed:

1. Verify sorted 2D matches on the basis of match type
2. Evaluate and incorporation an outcome of 2D or 3D matches into present state
3. Select the finest 2D match for extending and verifying in the upcoming cycle
4. Select individual matches for extension or verification
5. Select match combination pairs.

Sanders, et. al (2010)

CUDA programming is particularly appropriate to deal with the issues that can be represented as data-parallel estimations [69]. Data-parallel model can be used for processing huge datasets applications to increase the performance of the calculations. Hence data-parallel processing distributes data items to each parallel thread.

The initial phase of constructing a data parallel programme is to divide and allocate the data to multiple threads where each thread executes each piece of data. CUDA architecture is designed with two inbuilt CPU and GPU architectures where CPU and GPU guarantee consistent multithreading.

In this case CPU is known as Host machine and GPU is known as Device machine. Memory and cache latency is reduced by CPU architecture for each thread. The CUDA GPU model conceals latency with calculation from other thread wrap.

Samsi, et. al (2017)

In [70], authors have discussed static challenges of graph. The rise of graph analytic system has created an essential way for measurement and comparison of the capabilities of the system. The proposed subgraph isomorphism graph challenge drawn upon prior challenges from machine learning, high performance computing, and visual analytics for creating a graph challenge which is reflexive to various real world graph analytics processing system. Subgraph isomorphism graph challenge is an aggregated specification with several combined kernels that can execute either synchronously or individually.

The computations are too simple where prediction of the performance can be done on the basis of simple computing hardware model. The encompassing kernel provides context for the entire kernel that allows precise determination of input as
well as the output for each kernel. Moreover, it can be used for measurement and quantitative comparison over a huge range of current day and later system. The focus on graph analytics allows the subgraph isomorphism graph challenge to draw upon significant work from the graph benchmarking community. The Graph500 benchmark provides as a scalable power-law graph generator which is used to build the world’s largest graphs with the goal of optimizing the rate of building a tree of the graph. The Firehouse benchmark simulates computer network traffic for performing real-time analytics on network traffic.

The Page rank pipeline benchmark uses a Graph500 generator and provides many multi-lingual reference implementations to allow users to optimize the rate of computing Page rank on a graph. Finally, mini Tri takes an arbitrary graph as input and optimizes the time to count triangles.

Seong, et. al(1993)

In [71], Authors have proposed an incremental clustering based system and incremental clustering algorithm. Criterion function for minimizing attributed graph of incremental clustering is proposed in this paper. And attributed graph is used for representation of structural and semantic information.

Entropy minimization is the one of the best solution technique for the clustering problem in optimization of energy. This criterion function is used for evaluation of the different clustering in the incremental clustering algorithm.

**Brief points of incremental clustering algorithm are:**

Step 1: Input given is an attributed graph.

Step 2: Five operators are used by the proposed algorithm:

- Creation
- Merging
- Splitting
- Termination

The creation operation involves usage of two attributed random graphs in the operation of the clustering algorithm.

The four tests are introduced in the second operator of the algorithm. In the second operator, searching of possible hierarchies is performed, that includes
searching, creating, merging, and splitting test.

The third operator has introduced five operators which are used to build a hierarchy incrementally. The clustering algorithm performs their operation in a top-down manner. The proposed algorithm can apply best operator for achieving results from the test performed on entropy minimization.

**Shijie Zhang, et. al(2010)**

In this paper [72], Authors have introduced an index based algorithm known as SUMMA. SUMMA is a subgraph matching in a massive graph which occupies global as well as local information about database graph. SUMMA selects an uncommon combination of labels as the local index and plan to make usage of capturing relations among vertex that needs space for storing without expansion.

One more algorithm invented by Authors for matching purposes for query processing and they used the global index approach for estimation of distance within vertex.

When the working memory is not capable of storing database graph, at such time, SUMMA is considered as one of the leading algorithm that can implement subgraph matching problems.

SUMMA consists of:

1. To indicate uncommon combinations of labels as local index
2. To indicate shortest path tree as global index
3. Offers an inventive algorithm for matching graphs with the help of its local indexes and global as well

SUMMA algorithm has shown improved performance over the existing technique. It is capable to attain processing of query effortlessly.

Some basic definitions are also discussed in this paper are: Labeled Graph and Shortest Distance. To stimulate the processing of query, two indexes is constructed as global and local.

Local Index -Set of vertex label that is arranged in lexicographical order, Local index is used to determine vertex to initiate.

Global Index-Global index is used to find bounded graph that can be stored in working memory. It has used the shortest path tree as global index due to following
reasons:

1. This index provides information of global distance for eventual match.
2. Easier to manage since index size is linear to size of vertex.

Steps of an algorithm:

1. Matching of vertex by local indexing
2. Find bounded graph by global indexing
3. It includes estimation of distance & search algorithm
4. The Verification algorithm to detect exact match.

Results show that SUMMA is an efficient and accurate method.

Shirui Pan, et. al(2013)

In paper [73], Authors have discussed about the framework for the classification of labeled and unlabeled graph using the graph streaming technique. A framework is developed by using multiple techniques and algorithms. This framework used the concept of partitioning the graph into a number of chunk graphs. Graph classification creates a classification model which predicts the matching accuracy with input graph.

The graph classification process consists of two different methods as given below.

- Feature based method.
- Distance based method.

Classification of the graph is done on the basis of features of the subgraph. To create effective and efficient framework, algorithm have used different factors:

- Identifying informative sub-graphs with minimum redundancy
- Capturing concept drifting in streams

This framework follows processes as Instance Weighting, Selection of subgraph features and updating ensemble. It uses three algorithms for the classification of labeled and unlabeled graphs.

- Algorithm1: to find the minimum redundancy of sub-graph
- Algorithm2: to check minimum redundancy of sub-graph
- Algorithm3: combines the instance weighting and minimum redundancy for classification of graph streaming
Stephen C, et. al(2013)

Memory management is handled by Cgraph solely, once it is called in a programme, Cgraph data structures are allocated and deallocated automatically [74]. A parent graph declares numerous vertices and edges, subgraph and also defines other related data of graphs. The graph can be directed or undirected and called by a name. Hence, all this information is stored in one root graph and cannot be used by other root graph or independently. For declaring a parent directed graph Agraph_t object will be used with agopen() function.

The first parameter of this function is any string. The second parameter of this function is a graph type. If it is a directed type of a graph, one of the arguments will be used as Agdirected, Agstrictdirected, Agundirected, or Agstrictundirected. The third argument of this function is NULL or optional. agnameof(g) this function is used to recall the name of the graph. The Subgraph is an essential part of Cgraph. They are expected for sorting out subsets of graph objects and can be utilized reciprocally with best level graphs in all Cgraph functions.

A subgraph is a combination of vertices or edges of its root graph. The hierarchy of a graph is made by subgraph which is also called as tree, therefore with the help of Cgraph functions we can search, create or iterate on the subgraphs. Graph or subgraph arguments are returned by Agparent() function. Cgraph gives some polymorphic capacities and macros that apply to all Cgraph objects. Some of the error functions are handled by Cgraph, hampered by the absence of special cases in C.


In [75], Authors have introduced DNA-based graph conceals algorithm. It is a technique used to fix few intractable graph difficulties like a subgraph isomorphism problem which is known as NP-Complete problems. This paper is categorized into various sections:

- Introduction of subgraph isomorphism problem.
- Provided detailed study about Adleman Lipton model
- Introduction of DNA-based algorithms

Several algorithms discussed in the paper are as follows:

- The Adleman Lipton model
• Sticker based solution space
• Sequence generation
• Permutation generation

These algorithms provide the solution for graphs problem. The DNA based algorithm used to solve the subgraph isomorphism problems and more than one common subgraph problem depends on bio-logical operation. The author developed a web server that generates DNA sequences that produces the solution space of the graph problem. Algorithms are developed to generate graphs from DNA samples.

**Van Harmelen, et. al(2008)**

In Computer Science, graph theory assumes the critical job since it gives a simple and deliberate approach to demonstrate numerous issues [76]. Various graph algorithms are useful for solving the graph problems, some of them can be used for serial and parallel problems.

The graph is a data structure which is represented by two components:

The first component of a graph is vertex also called a node and second component of a graph is an edge also called link. The day-to-day life applications are represented by graphs, which may include network and social networks. The network graphs are used for finding a path in a city or in a circuit network or in a telephone network. In the social networking like Facebook, LinkedIn, etc. a person’s structure is represented by a node. The node structure consists of, name of a person, ID, address, gender, etc.

The basic need of graph representation is to find out how much memory space is required for storage of information. Graphs are stored in the form of 0’s and 1’s.

**Vincenzo Bonnici, et. al(2012)**

This paper [77] presented algorithms and their application relevant to subgraph isomorphism. With the help of graph theory technique, medical science makes use of graph for representation of biological networks of molecular, protein etc. In this paper they described the existing algorithm. The aim of this paper is to eliminate unsuccessful mappings. Authors proposed subgraph isomorphism algorithm and also implemented the search technique. As per author’s approach the algorithm is
used to reduce search space without complicated rules and domain reduction procedure.

Comparison of existing and developed efficient sub-graph isomorphism algorithm is also discussed here. The applications in biological network system such as molecular level, protein to protein interaction, and metabolic interaction presented here. Networks are represented by graphs and molecular components, edges represent relationship between them.

Algorithms that finds network motif using following points:

1. All possible sub-graphs enumerate in network
2. Classify non-isomorphic sub-graphs
3. Generate random graphs and enumerate sub-graph

Authors have mentioned other applications that represents graph which shows small and large proteins molecular components. In this paper they have represented aspects of algorithms of sub-graph isomorphism methods - Search strategy and reduce the search space. Simple enumeration algorithm is also discussed by authors to find all occurrences of subgraph isomorphism in a main graph and implementation part is used for comparison of synthetic and bio-chemical data using subgraph isomorphism.


In this paper [78], authors have proposed various algorithms with their corresponding applications using pattern matching. It focuses on graph pattern matching use for social networking analysis. Different algorithms of this paper are discussed below:

**Algorithm 1: Graph pattern matching used in traditional way.**
This algorithm has used the old traditional technique to match pattern of graph and data graphs are used in pattern matching algorithm whose expected outcome was not perfect.

**Algorithm 2: Graph Pattern Matching Revised.**
This algorithm makes use of Bounded Simulation for matching graph. This algorithm finds the edge relationship with data graph, pattern graph and matching graph. Authors have considered the graph topology and query techniques for big social graph. Pattern matching of big graph pattern matching is too expensive.
Following methods are used in the query technique:

- Incremental method
- Graph Compression Query Preserving methods
- Distributed Graph Pattern Matching

West, et. al(2001)

There are two ways to traverse a graph: Breadth First Search (BFS) and Depth First Search (DFS)[79]. Tree is a non-linear data structure. A tree is a collection of elements called as “nodes” such that there is one special node known as root node.

Root: It is special node in a tree and the entire tree is referred through it. It doesn’t have parent.

Parent: It is any node expects the root node that has one edge upward to a node called as parent.

Path: Path refers to the sequence of nodes along the edges of tree.

Child: All immediate successors of node are known as child node.

Leaf: The node which doesn’t have any child is known as leaf.

Siblings: Nodes with same parent node is known as siblings.

A tree is binary if each node of the tree can have maximum of two children. One child is called as “left” child and the other is called as “right” child.

The representation of tree is done through array. Binary Tree can be represented using 1D array, the nodes are numbered sequentially level by level from left to right. Even, the empty nodes are numbered and the root of tree is stored at first location i.e. at tree[a] is the index position.

Whittaker, et. al(2009)

Following are the characteristics of a graph [80].

- The number of triangles
- The clustering coefficient
- The number of isomorphism
- Graph diameter

Several algorithms are available which performs some operation on a graph like finding degree, finding connectivity between its two clustering coefficient
(neighbors), path finding of nodes, shortest path finding between two nodes, etc.

Graphs are of following types:-

- Directed graphs
- Undirected graphs
- Random graphs
- Bipartite graphs.

Some graphs which are directed graphs, graphs with edge weights and non-directed graphs are considered in this research. The undirected graph has no information about the direction between nodes or flow, whereas on the other hand directed graph include direction information and the weighted graph add some additional feature of the edge.

Isomorphic property is the main property of graphs, which mentions appropriate match between the two given graphs, dimensionality, structure, corresponding nodes mapping and connectivity, etc.

Woodward, et. al(1979)

Complexity theory is a part of the hypothesis of calculation dealing with the assets required during calculation to tackle a given issue. The execution of a programme is the measure of computer memory and the time expected to run a programme [81]. There are two ways to deal with decide the execution of a programme. One is explanatory, and the other is trial. In execution examination logical strategies are utilised, while in execution estimation tests are leaded.

If a problem is given, try to find the solution. For finding the solution, algorithm are identified carefully and for this it is necessary to confirm that whether an algorithm is appropriate algorithm for the given problem or not. In such a case, the most common resources for identification of an algorithm are time and space complexity. With the help of complexity computation, one can conclude whether given algorithm is an efficient or not.

The complexity of an algorithm C is the g(n) function which gives the running time and additionally storage room necessity of the algorithm as far as the size 'm' of the input. Generally, the storage space needed by an algorithm is essentially different from the information measure 'm'.

35
The g(n) function, represents the execution time of an algorithm which is based on both size of input and a specific data.

**Xiao Yu, et al.(2012)**

In this paper [82], authors have given demonstration on query driven discovery of linguistically same sub-structures in heterogeneous network. Architecture of S⁴ framework is divided into 2 parts: Off-line analysis and indexing, on-line query answering. Various modules of the system are described below:

**Entity Similarity Analysis** – Several techniques are used to study one of the ultimate principal in heterogeneous information network, which calculates the entity similarity. This module is responsible for calculation of entity similarity using various measurements that are based on meta-paths. Results of prominent entities are stored as similarity indices, Structure Index Selectivity Analysis: Similarity index is responsible for retrieval of similar entity list with distinct semantic meaning and conveniently finds the entity combinations of query in the network. Structure index requires time as well as space to develop, and it is not possible to off-line the index subgraph structure, candidate Validation: This module is the final step of on-line method. Finally retrieved results are calculated after verification of partial results which are obtained from the structure index and similarity index. There are two principles that have pursued in the system for making the system to response fast, Shorten the candidate validation number, Diminish the sub-graph structure size, Implementation Detail: The system is able to work well with minimum resources like memory and hard drive. The system will generate a small index when there is limited hard drive. Aim of limited hard drive is to make most frequent cases fast and dynamically searching un-indexed entity.

As per the conclusion, S⁴ system prototype is applicable to various outlines; some of them are listed below:

- New endorsement system for B2C
- Searching of multidimensional candidates while company recruitment
- Exploring professional community on precise fields for scientific researcher
- Search engines of multi-dimensional similarity for journalist and lawyers