1.1 History of Graph Theory

In recent years one of the most interesting field of research is Graph Theory for its applications in diverse fields including Pure Mathematics, Chemistry, Physics, Biochemistry, Electrical Engineering, Operation research, Sociology, Computer Science and many more. The importance of Graph Theory is growing exponentially due to more and more applications of graph theoretical concepts and algorithms in many fields and solving several real life problems. The main objectives of using Graph Theory is representing a model for a certain problem for easy visualization and have a clear understanding of that problem. Graph Theory is the easiest way to represent any type of problems. By representing every problem using graph had made itself a pioneer in finding solutions for many unsolved problems of different areas such as Computer Science, Engineering, Microbiology, Physics, Chemistry, Mathematics, Business and Medical etc. A solution to a single graph theory problem can have many real life applications. For example, travelling sales man problem, has lots of applications in our real life such as vehicle routing, scheduling, finding shortest path, drilling of printed circuit boards and many more.
The era of Graph Theory started in 1736, when a famous mathematician Leonhard Euler had a problem of crossing the seven bridge of Konigsberg. This was called the first problem of Graph Theory. The objective of the Konigsberg bridge problem was to find whether a walk through the city could cross each bridge exactly once. Konigsberg Bridge problem was studied by Euler and constructed a graph structure called Eulerian graph to solve the problem [5] [6]. Euler formulated the problem as a multi-graph by considering the land areas and the bridges as vertices and edges respectively. Euler showed that the problem could be solved by tabulating the degree of the graph as if a graph has not more than two odd vertices, and then there exists some path traversing each edge once. All four vertices in the Konigsberg bridge problem (graph) are odd [1]. The Euler graph had been used in certain application problems starting from the problem of Konigsberg seven bridges to the current problem of DNA fragment assembly [100].

1.2 Some Applications of graph Theory

It has been found that the problems in Physics can be modeled with the help of graphs. One of the important problems in Physics is computing the current flow through the wires of an electrical circuit, which is a very critical for designing complex electronic devices. This problem was greatly simplified by the German Physicist and Mathematician Kirchhoff by representing the circuit by the graph in terms of trees and cycles. In 1845, Kirchhoff introduced the concept of trees and used the graph theoretical ideas to calculate the currents in electrical networks or circuits.
Graph theory has many applications in the field of Chemistry [73][76]. The chemical phenomena in Chemistry can be modeled by using the theoretical concepts of Graph Theory. The structural formulas of covalently bonded compounds which can be represented by some graphs called molecular graphs used to study the nature of molecules in chemistry. The graph theory has aided chemistry to represent many compounds in symbolic formulae. Graph theory has been used to explore the complete set of all possible monocyclic aromatic and heteroaromatic compounds [73]. Some special kind of graphs has been used in chemistry. Some of them are-Dualist Graph for benzenoids and diamond hydrocarbons to enumerate isomers [76], Reaction graphs and Synthon Graph for chemical synthesis [73][76]. The graph enumeration methods are used to enumerate some specific classes of chemical compounds used for classification of organic compounds.

In Operation Research, a major field of statistical mathematics is widely using the concepts of Graph Theory. The transportation problem, which is one of the most important problems of Operation Research, can be easily solved by using the graph theoretical concepts where a graph is used to model the transportation of commodity from one place to another. Here the main objective is to minimize the cost or time within the transport network. The planning and scheduling of large and complex project is the most successful and accepted application of graph in Operation Research.

Linguistics is another field where graph theoretical concepts are heavily used to represent natural language parsing diagrams. The vertices are used to represent words and word strings, and the edges are used to represent the syntactical relationships between
them. Graph based approaches are used in monolingual and cross-lingual word sense disambiguation [105], natural language processing [115] and linguistic annotation [114].

It has been found that the graph colouring concepts can be used to solve various real life applications like job scheduling, aircraft scheduling, computer network security, automatic channel allocation for wireless local area network, resource allocation by the operating system and many more. Many of the computer science domain like database design, data mining, neural network, clustering, image segmentation, artificial network etc are tasted enormous development using the graph theoretical concepts.

The concepts of path, walk and circuit are used in several real time applications such as shortest path problem, database design etc. The concepts of maximum flow, minimum clique minimum spanning tree etc have been used for image processing. The concepts of graph partitioning have been used in image segmentation, which is again one type of image processing, has an application in medical imaging.

Graphs and Hyper graphs are considered as two most important tools in the theory of databases. There are many applications of graph theory in database technology, beginning from data modeling to deadlock control, transaction processing, query processing, optimization and data distribution [101]. Functional dependency graphs have been used to represent the dependencies among the entities of a system and to find out all the candidate keys of a relational scheme [32]. Hyper graphs have been used to represent the Database schemes [36][60]. Wait for Graphs and Resource Allocation Graphs are used
in deadlock detection in distributed databases. Lots of graph based approaches have been applied in database recovery and concurrency control [95].

In recent years, one of the most interesting field of research is Graph databases which is used as a storage system and provide index-free adjacency, where the concepts of Graph theory has been used. A good number of graph databases have been developed and they have their own merits and demerits depending on the graph models used. Graph databases are capable of handling dynamic and highly heterogeneous data [98]. In [42], a comparative study of different graph data bases has been presented and it is proposed that Neo4j graph database is the most popular graph database. Shalini Batra et.al., has been provided a comparative study between the relational databases and the graph databases and also concluded that Neo4j graph database is the most popular graph database [44]. Again, a comparison between relational database systems and graph databases has been discussed based on the data structures, data model features and query facilities [119]. Yuanyuan Tian has been proposed different graph matching algorithms like SAGA, TALE and two aggregation operations – SNAP and k-SNAP [27]. Abdurashid Mamadolimov has been developed a few search algorithms for conceptual graph databases [107].

Data mining is another area where graph and graph theoretic approaches have been extensively used to solve data mining problems. Some of the specific fields of data mining are - semi-supervised learning [103], clustering [102][113], frequent pattern mining [57 ] etc.
It has been found that lots of graph based algorithms are applied in the field of software engineering [108][109], database management [11][17][24][30], data mining [102][103], distributed computing [40], knowledge discovery [104] including mobile computing and pervasive computing. There are many graph based algorithms in computer networks like routing management, network security [64]. In distributed computing lots of graph based algorithms has been used in distributed deadlock handling, commit protocols and load distribution.

These resulted in the development of new algorithms and theorems of Graph theory, which can be used in many more applications in different areas.

In this thesis, three problems from three different fields such as Pure Mathematics, VLSI design and Bioinformatics have been discussed and found some interesting results. Some theoretical results of a particular class of Euler graph, constructed from Euler diagram have also been discussed.

1.3 Graph Theory in Pure Mathematics

1.3.1 Introduction to Pure Mathematics
Pure Mathematics is a branch of mathematics that studies entirely abstract concepts. In this field the principle of mathematics has been developed for their own interest rather than for their immediate usefulness. The principles of pure mathematics have been used in different fields like astronomy, physics, economics, navigation, engineering and many more.

1.3.2 Related works on Pure Mathematics

The many unsolved problems in Graph Theory and the wide range of applications of it make a rich field of current mathematical research. The powerful combinatorial methods found in Graph Theory have been used to prove significant and well known results in a variety of areas in Mathematics [39]. There are lots of unsolved problems in Mathematics such as the Goldbach conjecture, the Riemann hypothesis, the twin prime conjecture, Determination of whether NP problems are actually P problems, determining if any odd perfect numbers exist, finding the formula for the probability that two elements chosen at random, proving which numbers can be represented as a sum of three or four cubic numbers and many more. It has been found that the application of matching, which is a part of Graph theory played an important role in Dharwadker’s proof of the four colour theorem [118] [80]. The graph theoretical concepts have been used to prove the mathematical theorem of Cantur- Schroder- Bernslein and Fermat’s little theorem [39]. In [55], has been discussed the existence and non existence of Fibonacci and super Fibonacci graceful labeling for certain type of graph. The theoretical explanations relating to pentagonal numbers, partition of numbers, graphical partition of perfect number have
been discussed from the two set of even numbers \{4n + 10 / n \geq 4\} and \{4n + 12 / n \geq 4\}[61]. Kalita et al have been discussed some properties of even numbers with various properties of graph and an algorithm has also been forwarded to know the various structure of graph relating to even numbers [77]. A direction of proof of Goldbach conjecture has been forwarded by Kalita [37] and the theoretical proof of it has also been forwarded in opposite direction [49]. A new direction of proof of Goldbach conjecture has also been discussed by Kalita using the Graph theoretical concepts [56].

**1.4 Graph Theory in Bioinformatics**

**1.4.1 Introduction to Deoxyribonucleic acid**

Deoxyribonucleic acid (DNA) is a chemical in the cells of every living organism, plays an important role as a storage medium for genetic information. DNA molecules are polymers that are built from simple monomer, called nucleotides consisting of three components such as sugar, phosphate and a base. Nucleotides may differ only by their bases, which are Adenine (C₅H₅N₃) (A), Guanine (C₅H₅N₅O) (G), Thymine (C₅H₆N₂O₂) (T) and Cytosine (C₄H₅N₂O) (C). The important feature of these bases is that its elements have pair wise affinity, i.e. Adenine (A) pair with Thymine (T) and Cytosine (C) pair with Guanine (G). This affinity is referred as the Waston- Crick complementarity and it is the central to the formation of double stranded DNA molecule. DNA sequencing, DNA assembling and DNA mapping are three main areas of interest in the field of DNA.
1.4.2 Introduction to DNA Sequencing

DNA sequencing is an approach to find the precise order of nucleotides within a DNA molecule, which is considered as an important problem of reconstructing the full strands of DNA. The process of rebuilding the original sequence of nucleotides from the set $\Sigma = \{A, C, G, T\}$ is called DNA sequencing. Sequencing by Hybridization (SBH) and the Sanger method are two popular methods used in DNA sequencing.

In DNA sequencing the main problem is to determine a sequence of nucleotides from an unknown DNA fragment. Generally the input data comes from a biological
hybridization experiment which is a set (called spectrum) of words (called fragments/oligonucleotides) over the alphabet \{A, G, C, T\}. These fragments may have fixed or varying length and usually have overlaps. The main aim is to reconstruct the original DNA sequence of a known length “n” on the basis of the overlapping fragments/words.

The concepts of Graph theory has been used in recently proposed approaches to the problem of DNA fragment assembly and DNA sequencing. This brings out the application of Graph theory in the field of Bioinformatics.

1.4.3 Uses of DNA sequencing

Every organism on the earth possesses a complete genetic material, which contains the full genetic information about the organism. DNA sequencing allows the scientists to use this genetic information in different research purposes.

**Medicine:** DNA sequencing can be used to detect the genes which are associated with some heredity or acquired diseases. Different techniques can be used to identify the defected genes and replace them with the healthy ones.

**Forensics:** DNA sequencing has been applied in Forensics science to identify a particular individual as every individual has unique sequence of DNA. DNA sequencing is also used to determine the paternity of a child.
**Agriculture**: The DNA sequencing and mapping of the whole genome of microorganisms has allowed the scientists to make them useful for the crops and food plants. For example, specific genes of bacteria have been used in some food plants to increase their resistance against insects and pesticide and as a result increase the productivity and nutritional value of the plants.

### 1.4.4 Related works on DNA and DNA sequencing

In bioinformatics, graphs are used for modeling gene-gene relationships [59]. Recently many theoretical research works has been done to realize general computation using DNA molecule. Adleman’s and Liptons have been demonstrated the use of DNA for representing information with the help of graph. A method has been discussed for solving a directed Hamiltonian path problem with seven cities using DNA molecule [31].

Graphs and graph theoretical concepts have been used to simulate and study biological systems [87], biological networks [82], human brains [79], genomics [33], protein analysis [63], DNA sequencing [91][117], DNA fragment assembly [89], graph data management for molecular biology [99].

Since 1970’s DNA sequencing technologies have been in existence and are still in the evolving stage. Sanger and Gilbert received Nobel Prize for DNA sequencing methods in 1980.
Hamori and Ruskin proposed a graphical representation to describe DNA sequences in 1983 [92] and later Nandy introduced some other graphical representations of DNA sequences [120]. A novel 2D graphical representation of DNA sequences of low degeneracy have been presents by Guo et al [93]. N. Jafarzadeh and A. Iranmanesh presented a new graphical representation of DNA sequence based on codons using the ALE- index for presenting numerical representation which is an invariant for DNA sequences [75].

It has been found that graph theoretical concepts are used to sequence, assemble and analyze the function and structure of genomes. Bains and Smith methods [66], Drmanac et al [67], Lysov et al methods [68] are some graph theory based methods for DNA sequencing using a directed graph where each vertex in the graph corresponds to each element of the spectrum which is a set of oligonucleotides or fragments.

In [66], the first algorithm has been proposed for reconstructing the original sequence on the base of a spectrum, which is a set of oligonucleotides or fragments. In 1999, Ludry and Waterman has been forwarded an algorithm for DNA sequencing by hybridization by using the concepts of Graph theory [96]. The graph theoretical approaches to DNA sequencing and fragment assembly have been discussed by Pevzner [84][85]. Another method has been forwarded to DNA sequencing with nanophore using the concepts of DNA graph [81]. In [89], has been discussed the use of De Bruijn graphs and Eulerian circuits for solving some problems in DNA sequencing and fragment assembly. Marta Kasprzak has been discussed various methods to solve the combinatorial part of DNA sequencing by hybridization, based on the graph theoretical concepts [91]. A method
has been forwarded to compare DNA sequences by constructing the weighted directed graph and the representative vector for a given set of DNA sequence [86].

1.5 Euler diagram and Euler graph

Euler diagram has a wide variety of uses from information visualization to logical reasoning. The ability to automatically layout of Euler diagrams brings considerable benefits to the application areas of Euler diagrams. There have been several recent developments towards the automatic generation and layout of Euler diagrams, which is the key of usefulness of Euler diagram for information visualization and visual languages. One significant feature of Euler Diagram is the capacity to visualize complex hierarchies. Euler diagrams are associated with various graphs, which play an important role in their automated layout. For example, Euler graph which can be constructed from the Euler diagram, where a vertex is each point where two or more curves meet and the edges are the curves segments that connect the vertices, plays an important instrumental role of automatic layout and modification of the Euler diagram. The recent application areas of Euler diagram includes software modeling, constraint diagrams, visualization of networks, database visualization, file system organization and many more.
1.5.1 Definition of Euler Diagram

An Euler diagram is a pair, \( d = (\text{Curve}, \text{l}) \), where

a) Curve is a finite collection of closed curves each with co domain \( \mathbb{R}^2 \), and

b) \( l: \text{Curve} \rightarrow \text{L} \) is an injective function that returns the label of each curve.

![Euler diagram showing the relationship between territories in the British Isles](http://www.wikimedia.org/wiki/File:British_Isles_Euler_Diagram_10.svg)

**Figure 1.2** Euler diagram showing the relationship between territories in the British Isles

(Source: http://www.wikimedia.org/wiki/File:British_Isles_Euler_Diagram_10.svg)

1.5.2 Tools for generating Euler Diagram

It is very difficult task to draw Euler Diagram automatically. Therefore, some software tools are used to draw Euler Diagram. Some of them are –

**Venn Master:** this tool is used for drawing area proportional Euler Diagram.
**Draw Euler**: this tool is used for drawing exact area proportional Euler Diagrams using polygons.

**Smart Draw**: this tool is used for drawing Venn and Euler Diagrams with the help of circles.

### 1.5.3 Related works on Euler Diagram and Euler Graph

A general definition of Euler diagram and its properties have been discussed by Stapleton et al [45]. It has been found that there are various methods for generating the Euler diagram and each method generates a particular class of Euler diagram having some particular set of properties, called well formed conditions [25]. Euler diagram and their extensions have been used in the area of information visualization [22] [23] [34] [47]. There are lots of application areas of Euler diagrams such as constraints diagram [88], software modeling [121], database visualization [51], visualization of networks [90], file system organization [83] etc. There are various methods for automatically generating Euler diagrams each of which concentrating on a particular class of Euler Diagrams [38] [46] [69]. Gem Stapleton et al have been forwarded several methods for generating Euler diagram by modifying the existing layouts, useful in domains where require updating or modifying diagrams such as in the logical reasoning context [26]. It has been found that by making some changes in the Euler graph, which is directly derivable from the Euler diagram, can generate a number of Euler diagram having same set or different set of properties. From these diagrams one can choose a diagram having a particular set of
properties according to their needs. It has been found that transformation of Euler diagram could be done by adding curves to an Euler diagram using a dual graph of the Euler graph [54]. In [52], has been discussed the modification of Euler diagrams by altering the corresponding Euler graph of the diagram and also discussed the measures that count the number of violations of properties of Euler diagrams. Euler graph has been involved for solving different problems, starting from the famous problem of Konigsberg’s seven bridges to the recent problem of DNA sequencing and fragment assembly to reconstruct full strands of DNA based on the data recorded. In [58], a novel algorithm has been proposed for segmenting an image into different regions using Euler graph.

1.6 Graph theory in VLSI design

1.6.1 Introduction to VLSI

VLSI is the process of creating an integrated circuit by combining thousands of transistors into a single chip. The era of VLSI began in 1970’s when communication technologies and complex semiconductors were being developed.

1.6.2 Typical steps of VLSI chip design

In a VLSI system, a designer will typically go through six major steps-

**Specification** - In this step a functional specification of the system has been produced.

**Logic design** - In this step the functional specification is translated into a logical representation typically via Boolean expression.
**Circuit design** – In this step the logic representation is converted into a circuit by using various components from a variable library modules such as AND, NOT or OR gates, standard cells or building block macros.

**Physical design** – In this step the circuit design has been translated into a physical package representation, specified through a set of mask descriptions, which define how the individual layers of the integrated circuit are to be produced. At this step the circuit representations are converted into geometric representation of shapes, which is called integrated circuit layout.

**Fabrication** – In this step the physical package representation has been used to fabricate an actual integrated circuit.

**Testing** - In this step determine whether there are any manufacturing errors that prevent the integrated circuit from implementing the functional specification.

**Floorplanning** is the first major step in physical design. The major objective of floorplanning is to allocate modules of a circuit into a chip to optimize some of the design metrics such as wire length, area and timing.

**1.6.3 Related works on VLSI design and Floorplanning**

In the field of VLSI design, it has been found that there are many unsolved problems related to graph theoretic concepts. Graphs are important tools to analyze
electronic circuit design, short circuit testing for printed circuit board [122][10][28][29], and floor planning [10][14]. Few works have been studied in the field of VLSI design technology by Y. Lai and S. M. Leinwand [3], W. R. Heller et al [8], Kalita B.[2] and many others.

In VLSI design, floorplanning is one of the most important phase, where the main concern is placement of rectangular modules of arbitrary size and shape in such a way that the total area covered by the modules and interconnection should be minimum. It has been found that a floorplan can be classified into two categories, Slicing structure and Non-slicing structure. The Slicing structure was first proposed by Otten [13]. Many authors W. R. Heller et al [8], Maling et al [9], Sadiq M Sait and Habib Youssef [15], R. H. J. M. Otten [13], B. Preas and C. S. Chow [12] have been discussed various results related to floorplan design. It has been found that the area minimization problem for floorplan of VLSI design with rectangular module is an NP complete problem [16]. The graph theoretical rectangular dualization method has been used in floorplan design. It has been found that every dual graph of a rectangular floorplan is a planner triangulated graph [112]. B. G. et al have been proposed that a floorplan of a circuit dual graph can be designed in such a way that the floorplan consists of only I-module and L-module [14]. E. F. Y. Young et al have been presented a method to handle different kinds of placement constraints in floorplan by augmenting the constraint graphs with edges of positive, negative or zero weights [53]. In [50], an analytical method has been discussed for general floorplan and optimization based on a mixed integer programming model and application of standard mathematical software.
There are two popular approaches to floorplanning, Simulated annealing and Analytical formulation, typically used for solving the floorplanning problem. Simulated annealing based floorplanning relies on the representation of the geometric relationship among the modules and analytical approaches usually capture the absolute relationship among the modules directly. In simulated annealing approach, first has to encode a floorplan as a solution, called a floorplan representation, which model the geometric relation of the modules in a floorplan.

It has been found that Floorplan representation becomes an important issue in floorplanning as it has a great impact on the flexibility and complexity of the floorplan design. Wong et al has been proposed a mechanism that transforms a binary tree in post order, called Polish expression to represent a Slicing floorplan [111]. It has been found that there is various floorplan representation of Non-slicing structure such as Normalized Polish expression [111], Sequencing Pair [41], O- tree [71], Corner Block list [18], B\(^+\) - tree [21], transitive closure graph [19]. Chang Tzu Lin et al has been proposed a new representation of VLSI floorplan, which can efficiently reuse some area that cannot be utilized by Normalized Polish expression and also able to represent Non slicing structure of floorplan [20].

1.7 Thesis outline

In this thesis, the following topics are discussed.
1.7.1 Consecutive Even Number Finding Graph (CENFG) Related To Gold Bach Conjecture

In this chapter, the proof of Gold Bach conjecture, which proposes that every even number greater than 2 (two) can be expressed as sum of two primes, has been forwarded with the help of consecutive even number finding graph (CENFG). Two new definition and two theorems have been established. The consecutive even number finding graph has been discussed in this chapter. An algorithm, Construction of adjacency matrix for consecutive even number finding graph (CENFG), has been proposed which gives the adjacency matrix with self loop for consecutive even number finding graph.

1.7.2 Shortest Superstring in DNA Sequencing

In this chapter, a new algorithm has been developed to find the shortest superstring of a given DNA spectrum having variable or fixed length of fragments. In this algorithm the graph theoretical approach has been used which gives all the possible shortest superstrings that may be present in a given DNA spectrum of fixed or varying length of fragments. The main objective of the algorithm is finding a maximum weighted path from a directed weighted graph which has been drawn from a given DNA spectrum having fixed length or varying length.

1.7.3 Properties of Some Euler Graphs constructed from Euler Diagram
In this chapter, a particular class of Euler graph has been constructed from the Euler Diagram having some certain set of properties. Here, the various theoretical properties of the Euler graph, dual graph and the intersection graph have been discussed.

1.7.4 NGPE: A New Generalized Polish Expression for VLSI Floorplan Problem

In this chapter, a new approach for representing the Slicing and Non Slicing floorplan having rectangular as well as L shaped modules have been proposed. In this approach the concepts of Polish expression and four new operators have been used to represent a floorplan. An algorithm has been developed to determine the shape and the layout of the modules in the floorplan. Two operations for any modification of a given floorplan, which can be performed on the NGPE- tree of the corresponding floorplan have also been discussed. From this new approach any Slicing and Non slicing floor plan containing I- shape and L- shape module can be represented by a NGPE or NGPE – tree. The NGPE – tree can be constructed by scanning the NGPE from left to right and each module and the operator corresponding to leaf and the internal node respectively. The NGPE can also be obtained from the NGPE – tree by traversing the NGPE – tree in postorder.

1.7.5 Conclusion and Future Scope

In this chapter, the conclusion and future scope of our research finding have been discussed.