**Graph Theory and Analysis of Biological Data in Computational Biology**

**WEEK 1: Assignment**

**Introduction:**

Graph theory is widely used as effective modeling, analysis and computational tool, in biological mathematics to deal with various biology problems. The focus of this assignment is to clear the basic concepts about Graph Theory and Analysis of Biological Data in Computational Biology. This is applied in various fields as :

* Microbiology: It can be used in such a way that the molecular structure, where cell, gene or protein can be denoted as a vertex, and the connect element can be regarded as an edge.
* Mathematical biology: It is used to deal with various modeling and calculation problems.
* Microscopic biology: DNA and other protein molecular structure can be represented as a graph. Given below are problems classified of several different domains which use graph theory for knowledge extraction from data:
* Modeling of Bio molecular networks: This is basically modelling methods of metabolic networks, protein interactions, transcription regulatory networks.
* Metabolic networks:The main role within a metabolic network is played by the enzymes, since they are the main determinants in catalyzing biochemical reactions. The collection of pathways, holding information about a series of biochemical event s and the way they are correlated, is called a metabolic network.
* Protein-protein interaction (PPI) networks**:** Hold information of how different proteins operate in coordination with others to enable the biological processes a within the cell.

**Measurement of centrality and importance in bio-molecular networks**

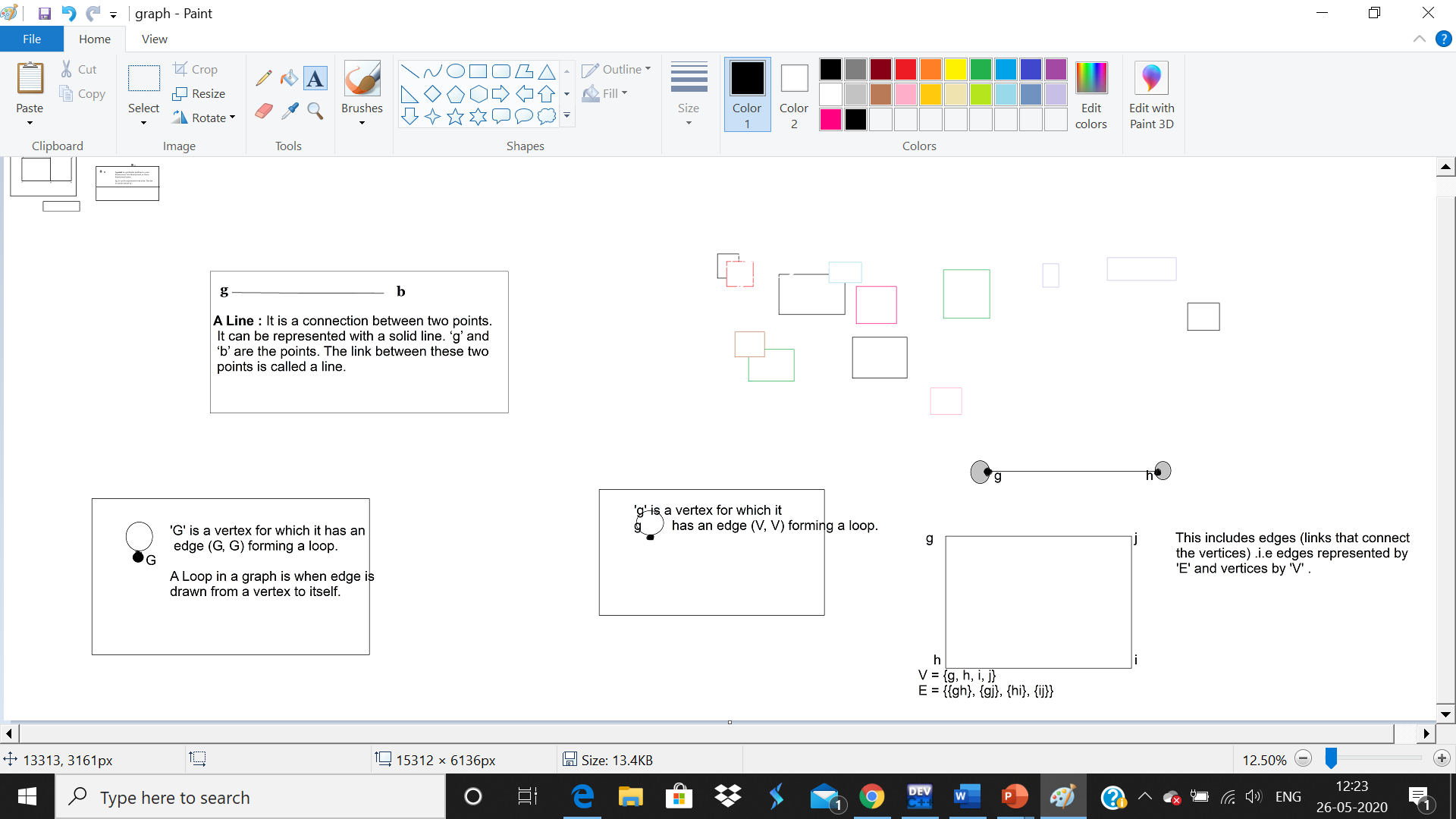
A large number of different molecules and extremely complicated consequence of the action that interact in many different ways .i.e. biological function. A functional reflection between Genomic associations genes and their products is seen. The strength of the functional associations correlates with the strength of the genomic associations. The genes which frequently co-occur in same operon in a diverse set of species are more likely to physically interact than genes that occur together in an operon in only two species and proteins linked by gene fusion or conservation of gene order are more likely to be subunits of a complex than are proteins that are merely encoded in the same genome.

Path, path length and connected graph:

Assume j, k be two vertices in a graph G. Then a sequence of vertices j = k1 , k2 ,..., kn = j, such that for i = 1,..., n-1, is said to be a path of length n-1 from j to k. The geodesic distance, or simply distance, d(j, k), from j to k is the length of the shortest path from j to k in G. If no such path exists, then we set d(j, k) = 1. If for every pair of vertices, (j, k), in graph G, there is some path from j to k, then we say that G is connected.

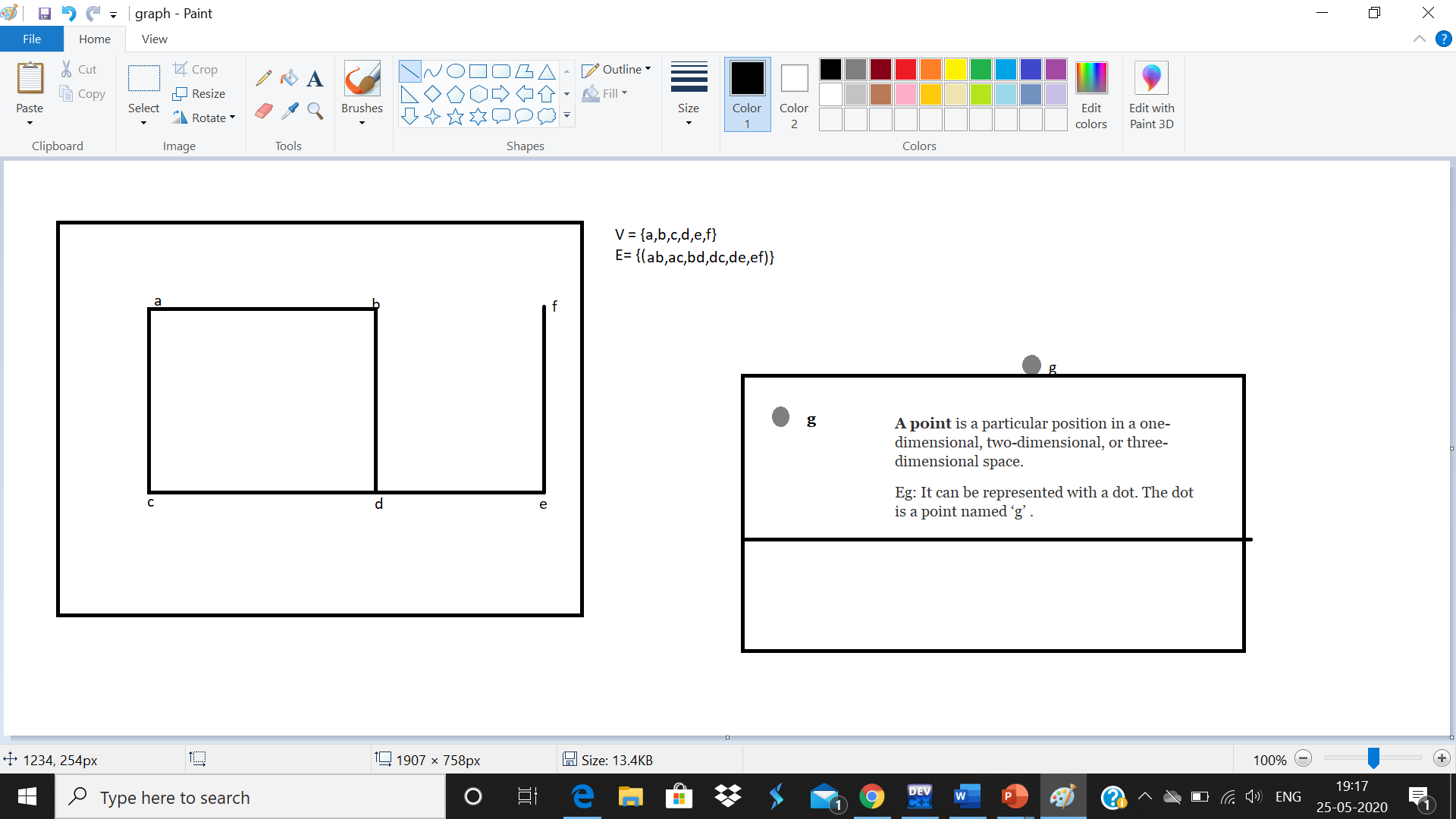
**Definitions and Mathematical Preliminaries:**

* Graph Concept: A picture designed to express words, particularly the connection between two or more quantities and relationship between two numbers or measurements in the form of a grid is a **graph.** The study of graphs that points out the relationship among edges and vertices is **graph theory.**

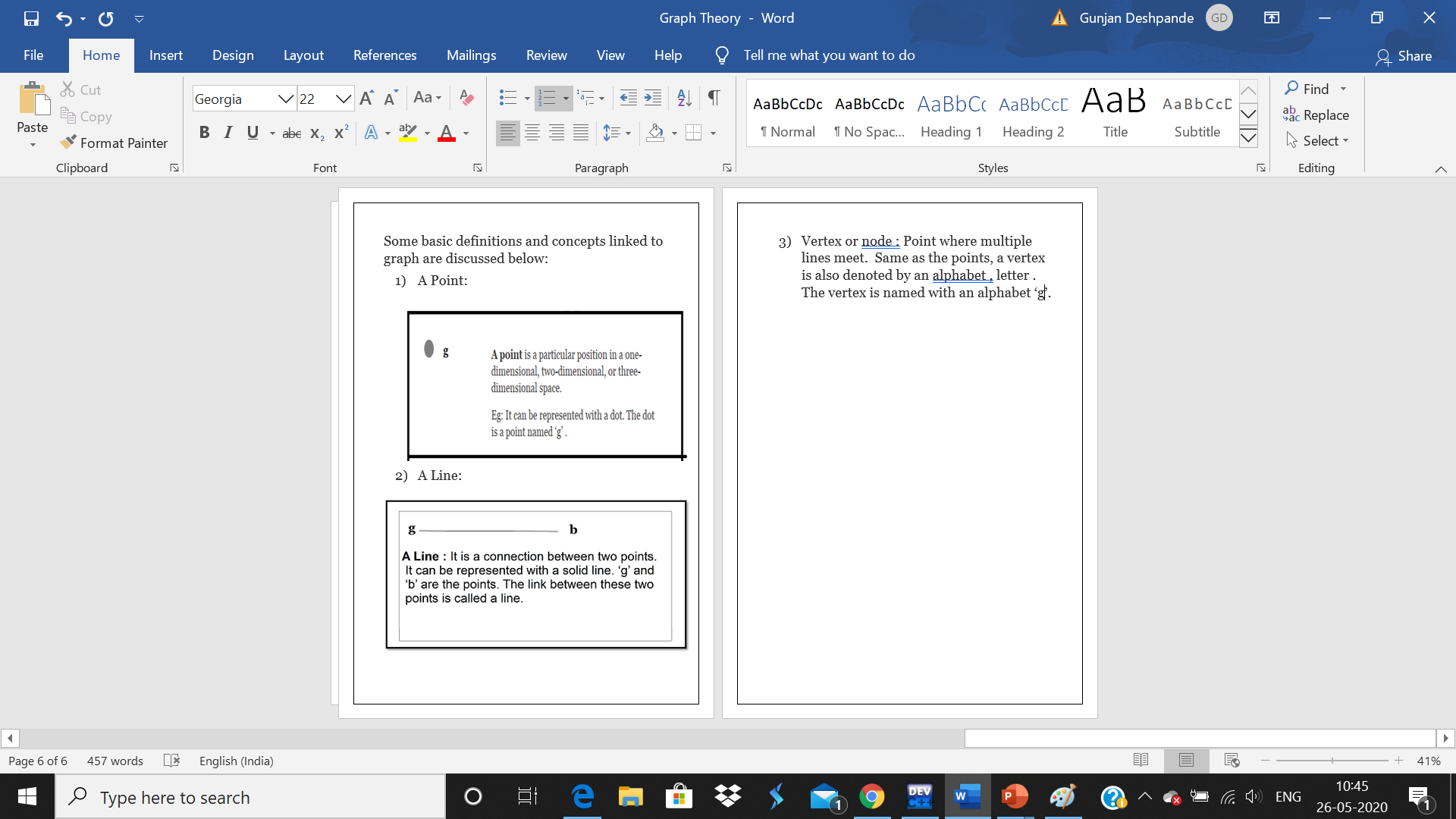


* Some basic definitions and concepts linked to graph are discussed below:

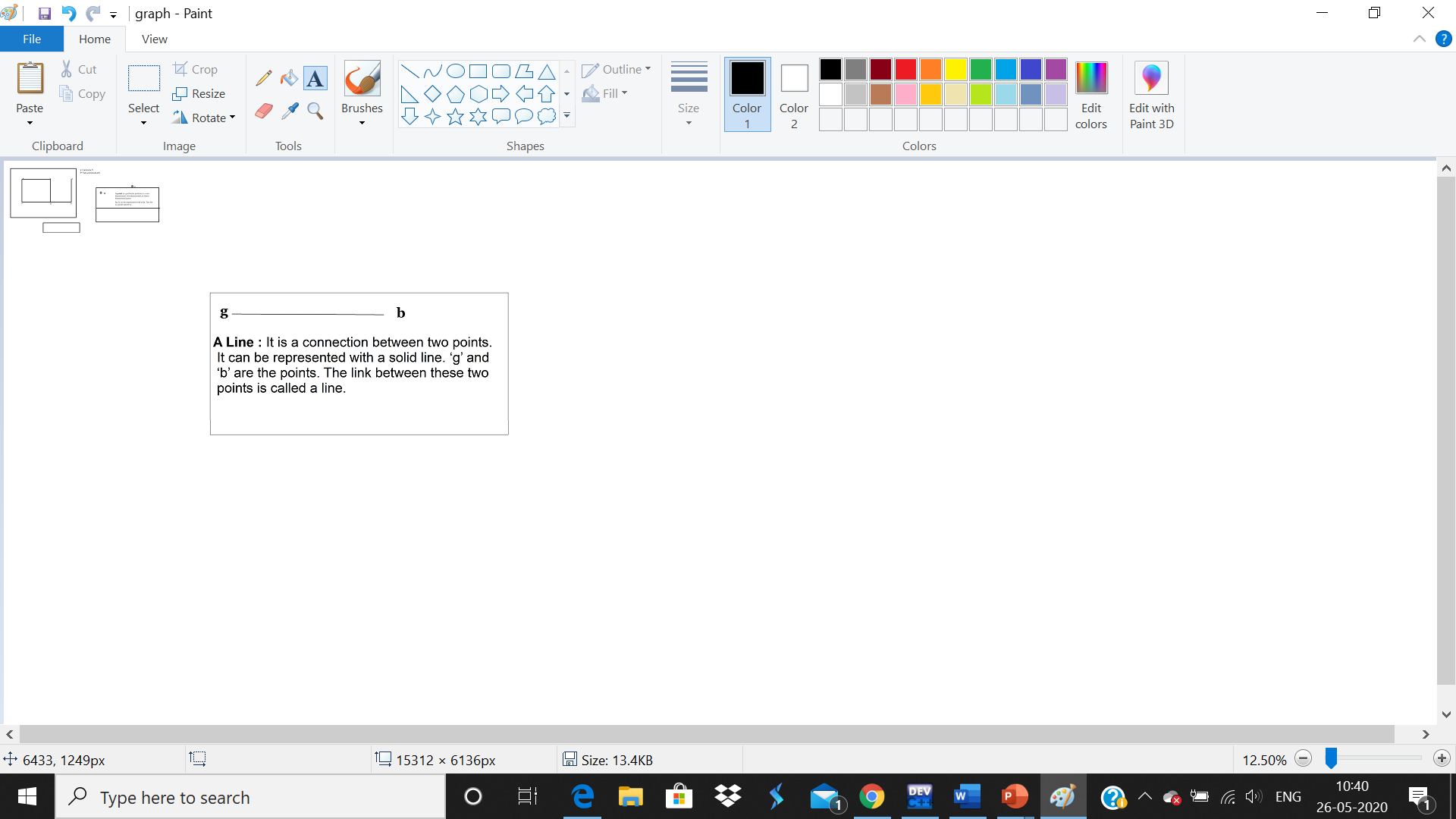
1. A Point: Basically known as a dot.

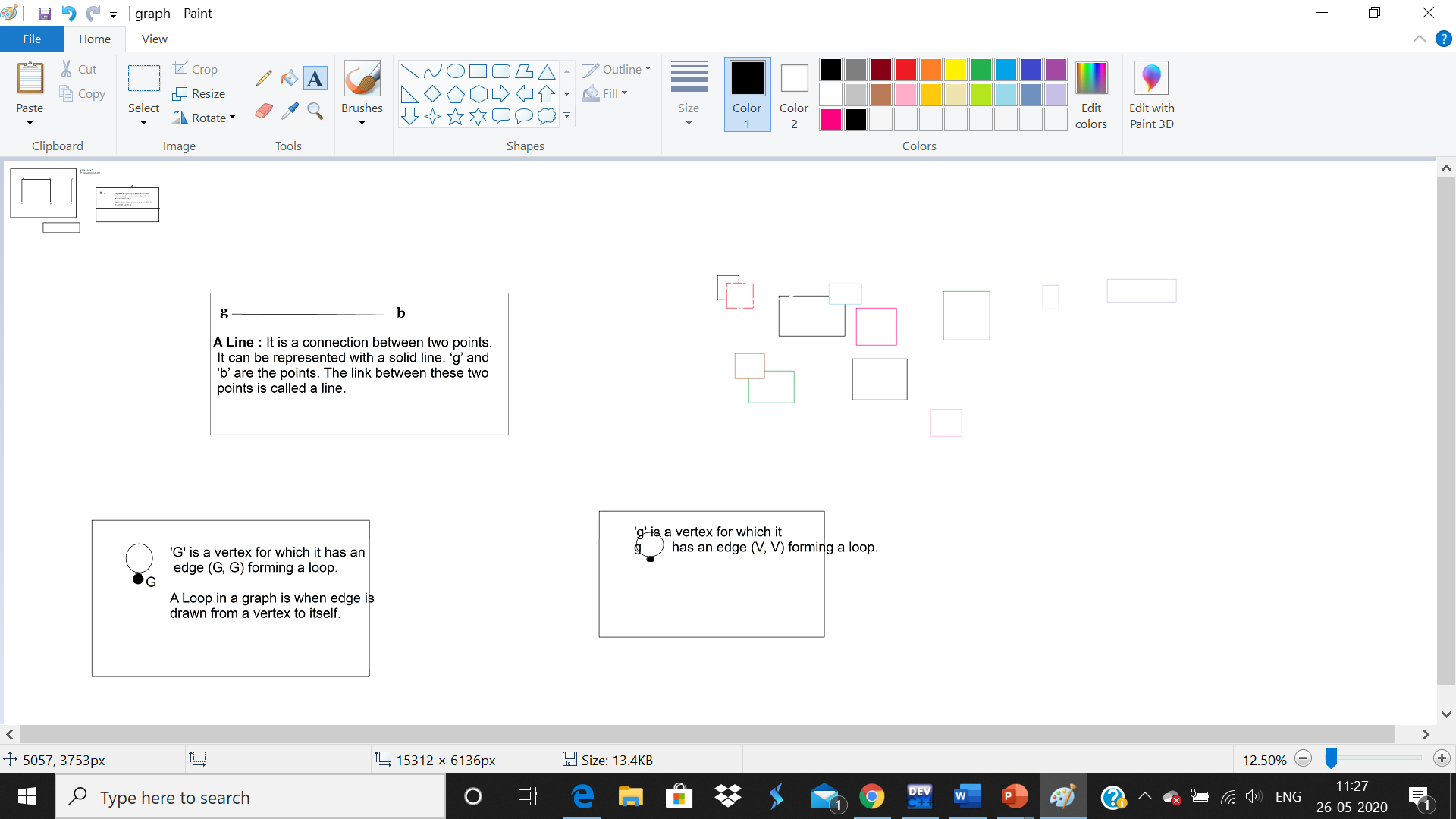


1. Vertex or node : Point where multiple lines meet. Same as the points, a vertex is also denoted by an alphabet , letter . The vertex is named with an alphabet ‘g’.

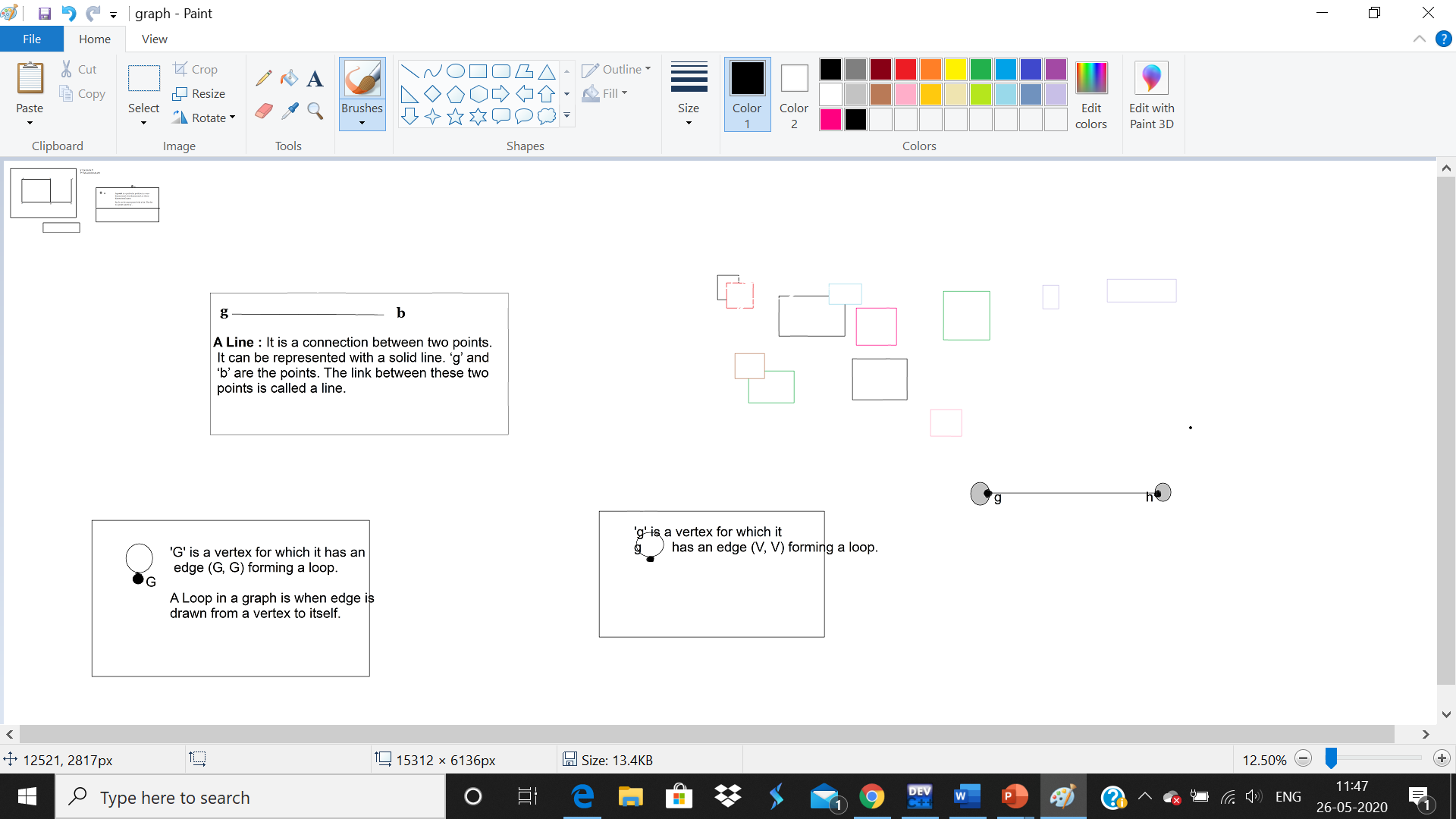


1. A Line:



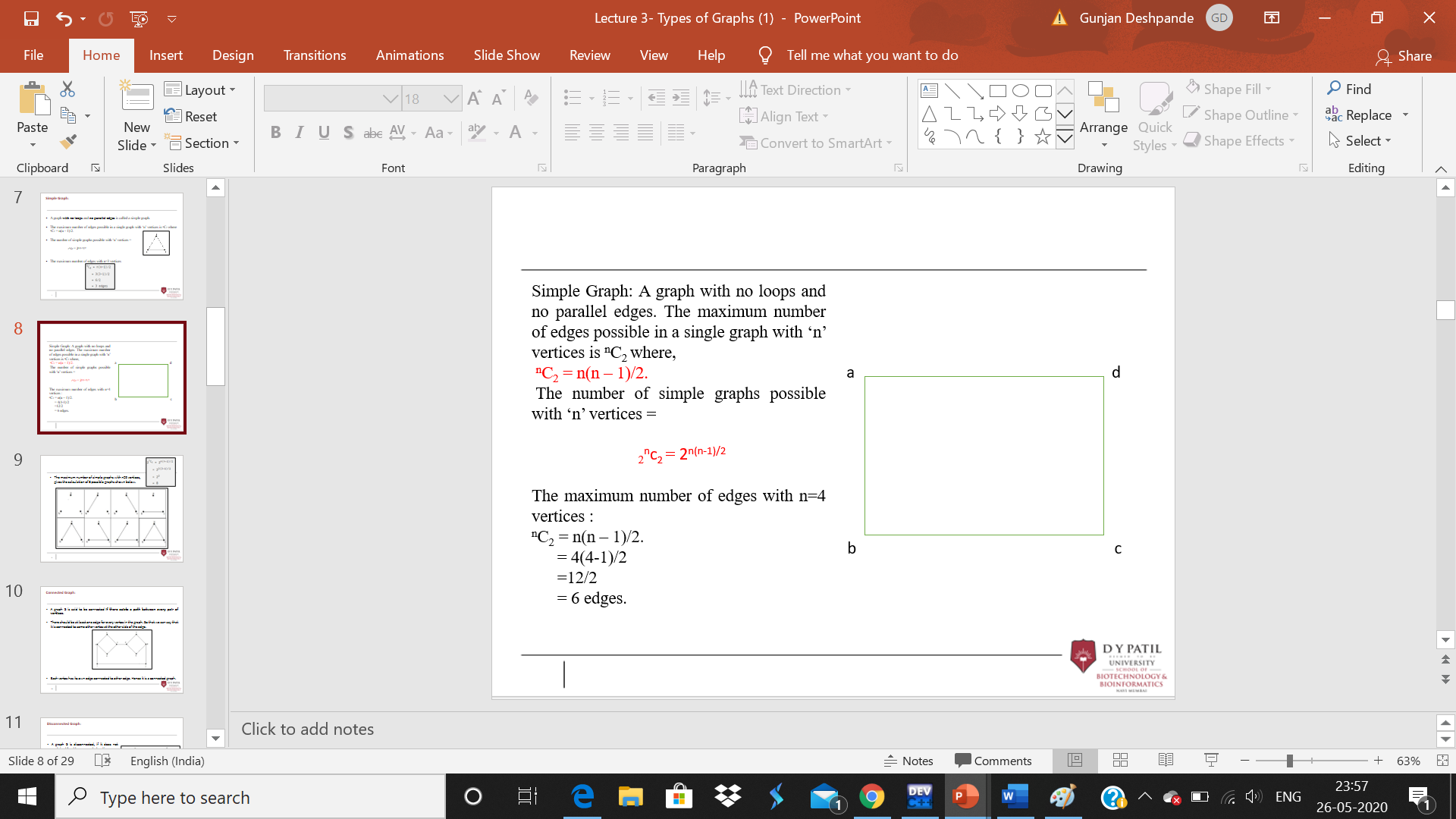
1. Loop:

Example : Where two loops are formed at the vertex ‘g’ and vertex ‘h’ is represented as :

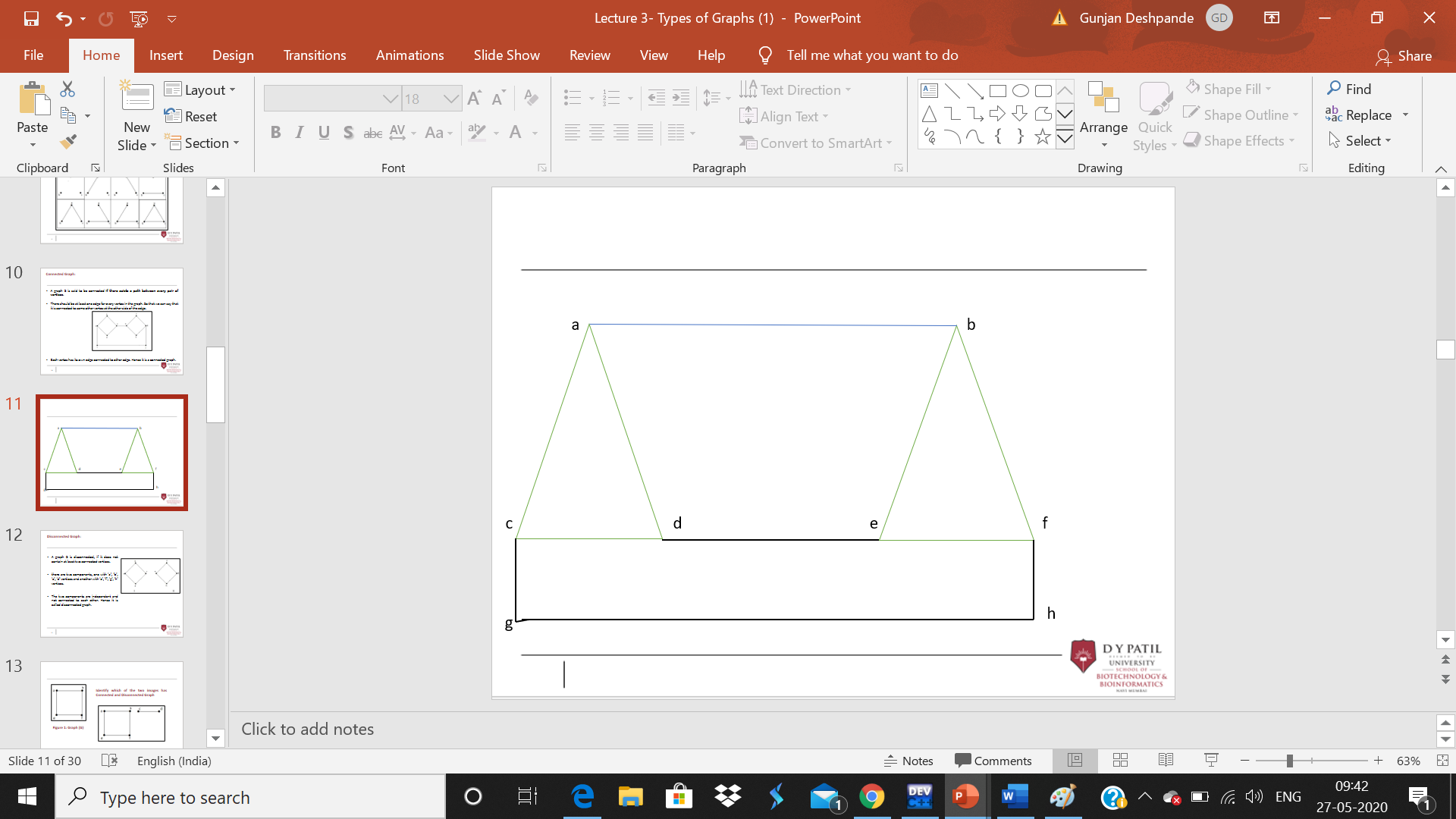


**There are many types of graphs :**

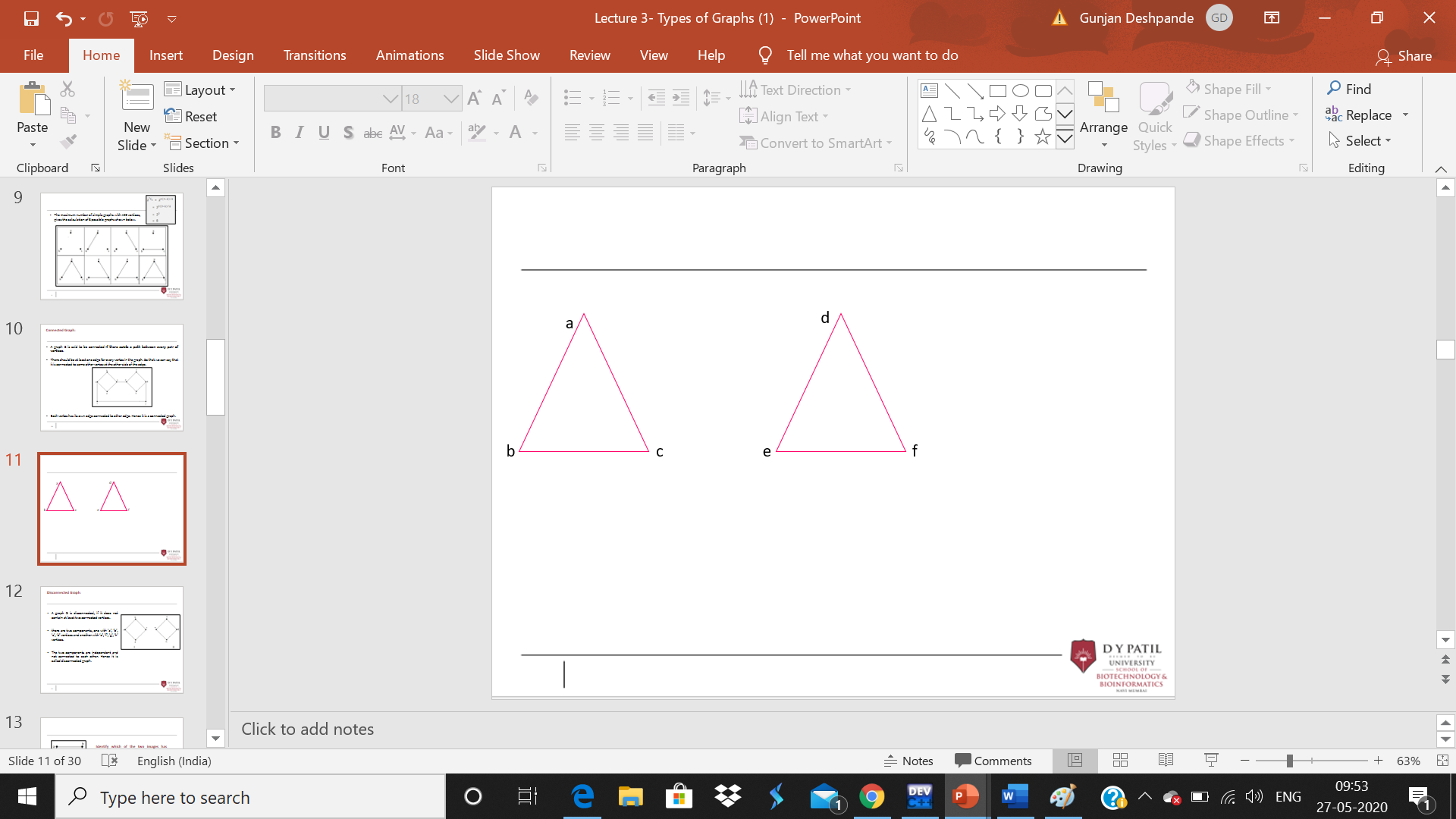
1. Simple Graph:



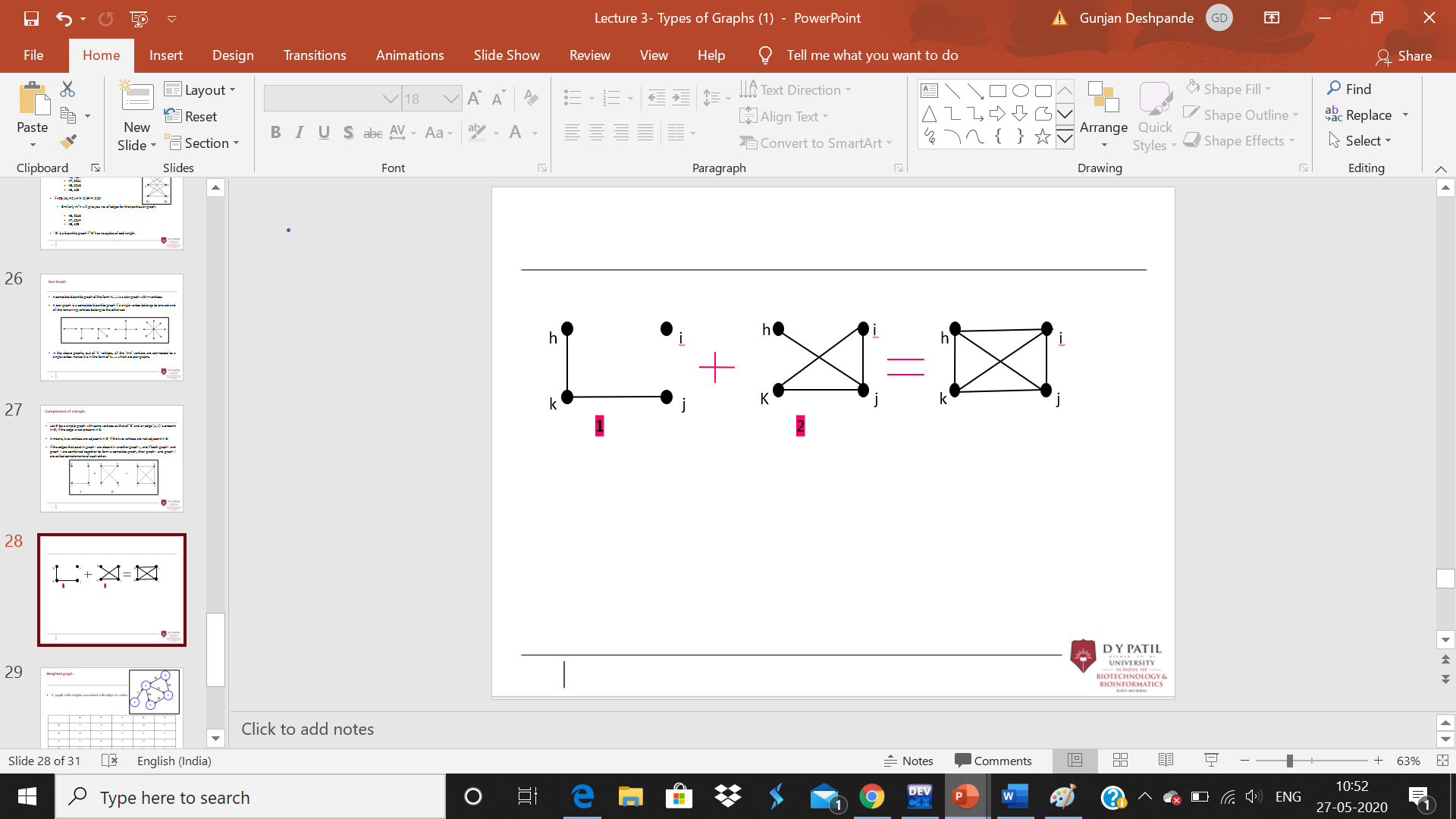
2.Connected Graph**:** Any graph will be called connected only if there exists a path between every pair of vertices. At least one edge for every vertex in the graph should be present. It is a connected graph because each vertex has its own edge connected to other edge.



3.Disconnected Graph: Any graph can be said to be disconnected only if it does not contain at least two connected vertices. There are two triangles shown above one with ‘a’, ‘b’, ‘c’ vertices and another with ‘d’, ’e’, ‘f’ vertices. The two components are independent and not connected to each other. Hence it is called disconnected graph.



4.Complement of a Graph**:** Complements of each other as said to be when :if the edges that exist in graph 1 are absent in another graph 2, and if both graph 1and graph 2 are combined together to form a complete graph, so here graph 1 and graph 2 are complement of graph.



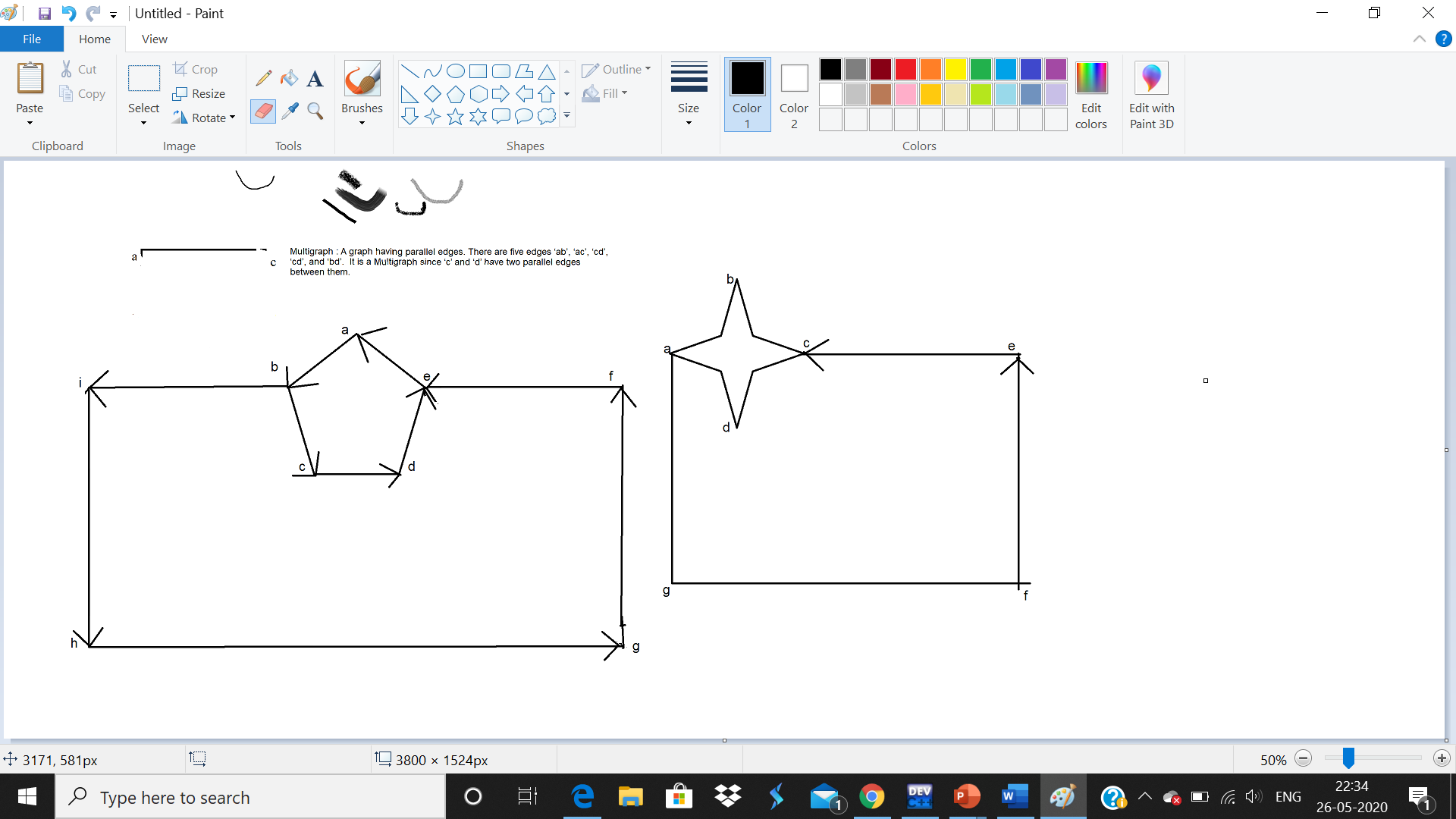
5.Non-Directed Graph/ Undirected Graph:

The edges in a non-directed graph are not directed ones.



6.Directed Graph:

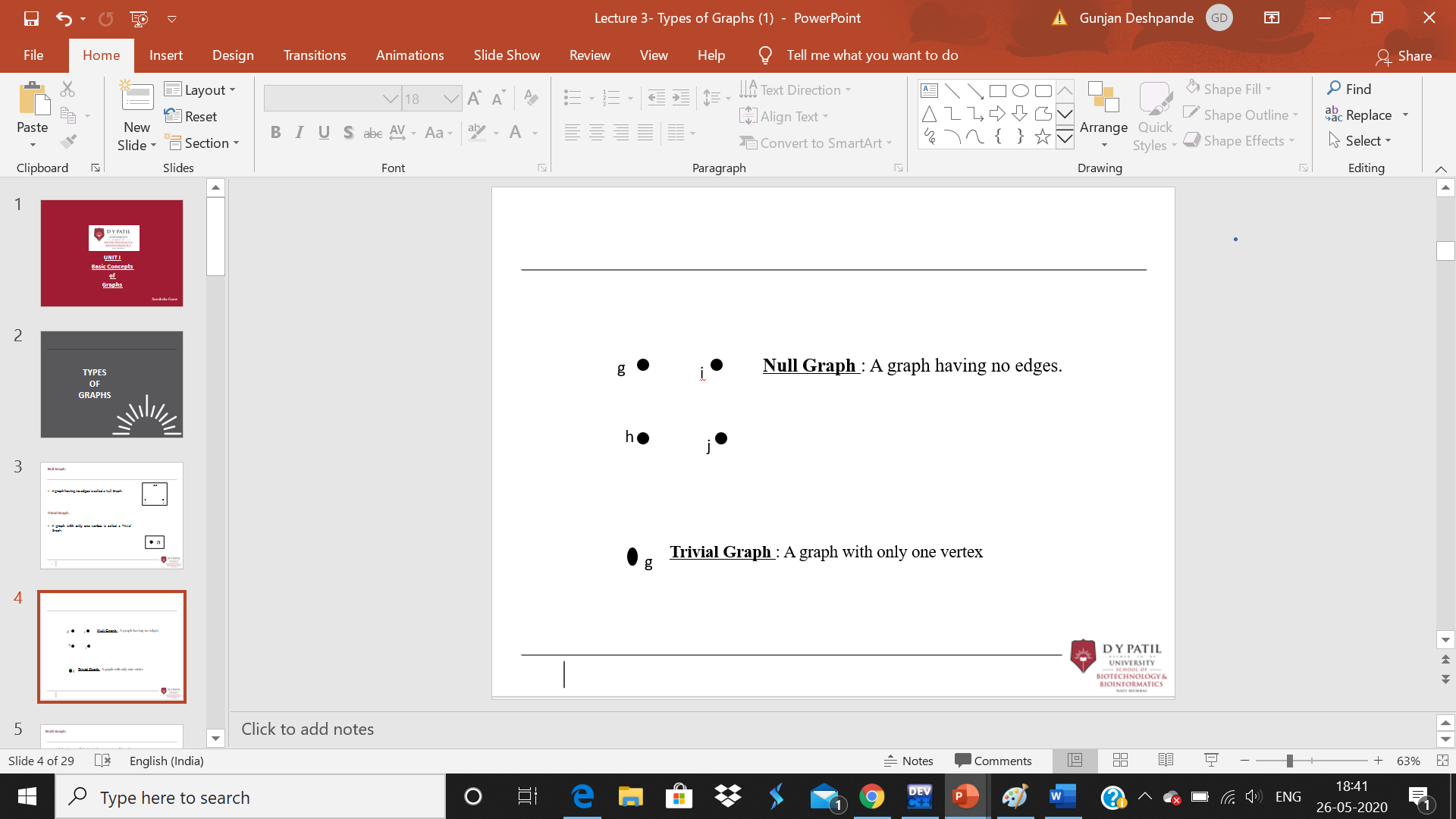
Unlike the Non-Directed graph / Undirected graph here each edge has a direction.

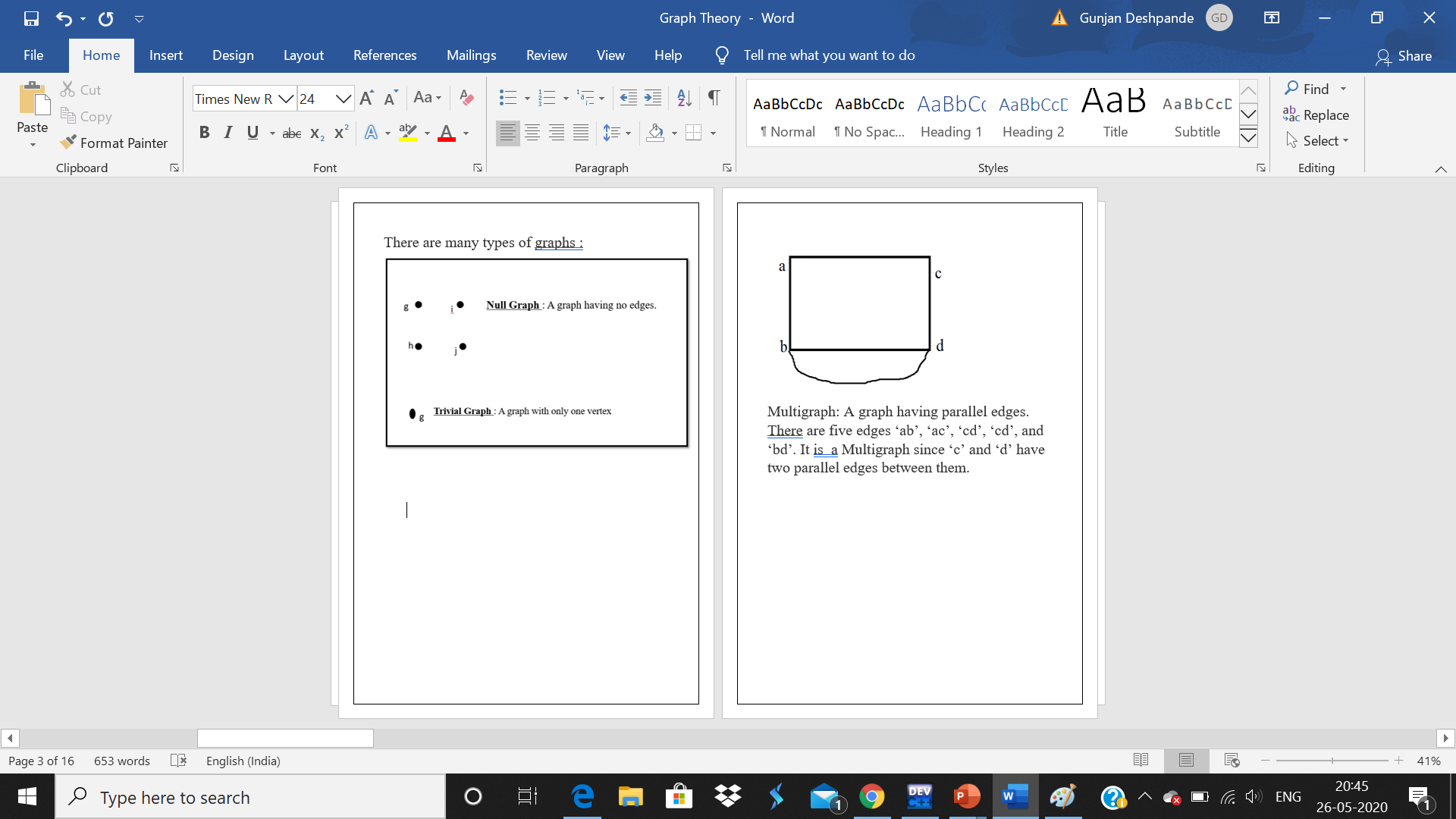


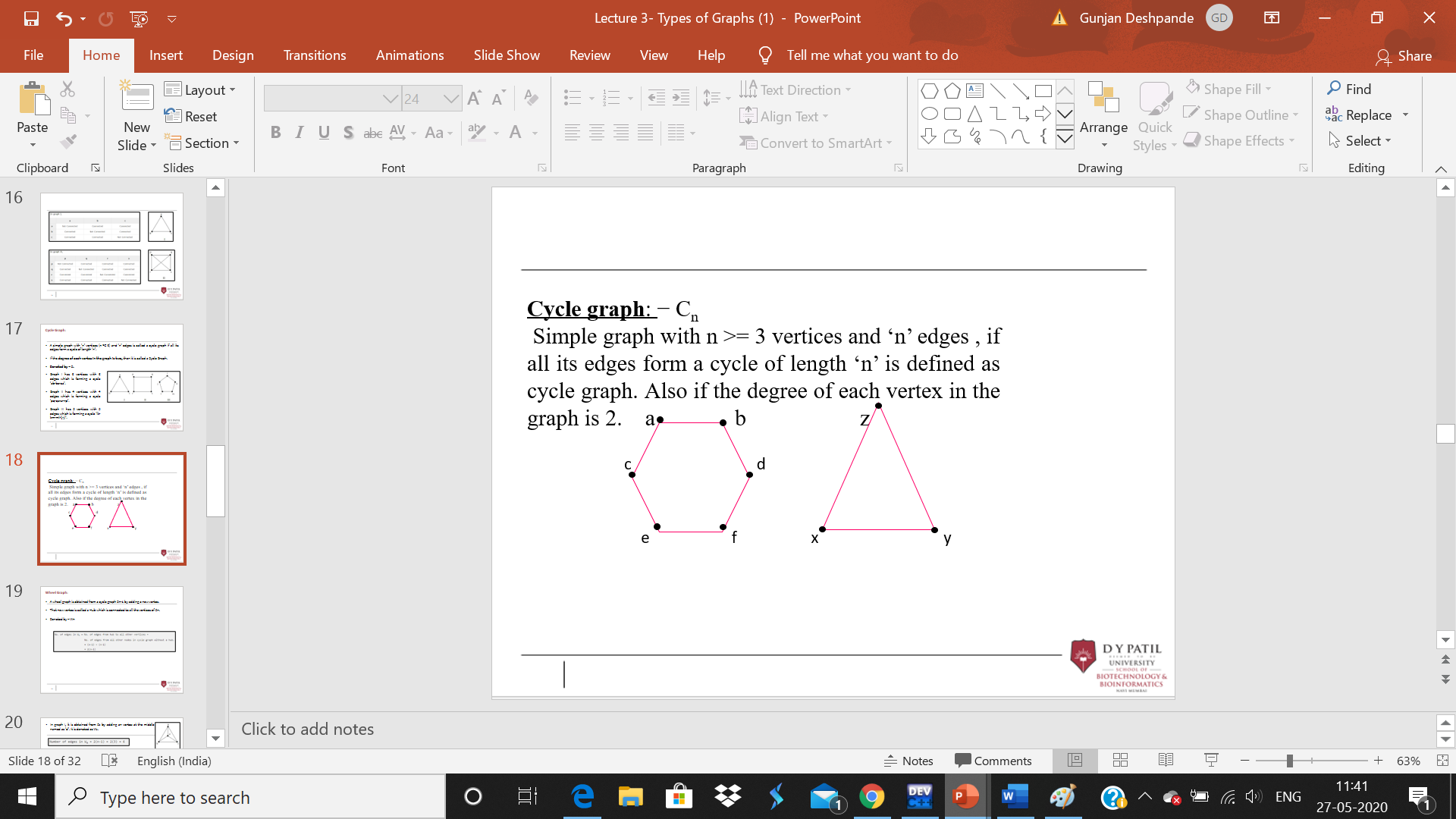
Given below are some basic types of graph for remembering :

1. Null graph
2. Trivial graph
3. Multi graph
4. Cycle graph
5. Acyclic graph and Eccentricity of a Vertex

**Some basic types of graphs:-**







**Cyclic Graph**: A graph with atleast one cycle given example:

a b

e

c d

Here we have two cycles in the above triangle:

1. a-c-e-a
2. b-e-d-b

**Acyclic Graph:** Any graph with no cycles the below example graph we do not have any cycles so it is a non-cyclic graph or acyclic graph.

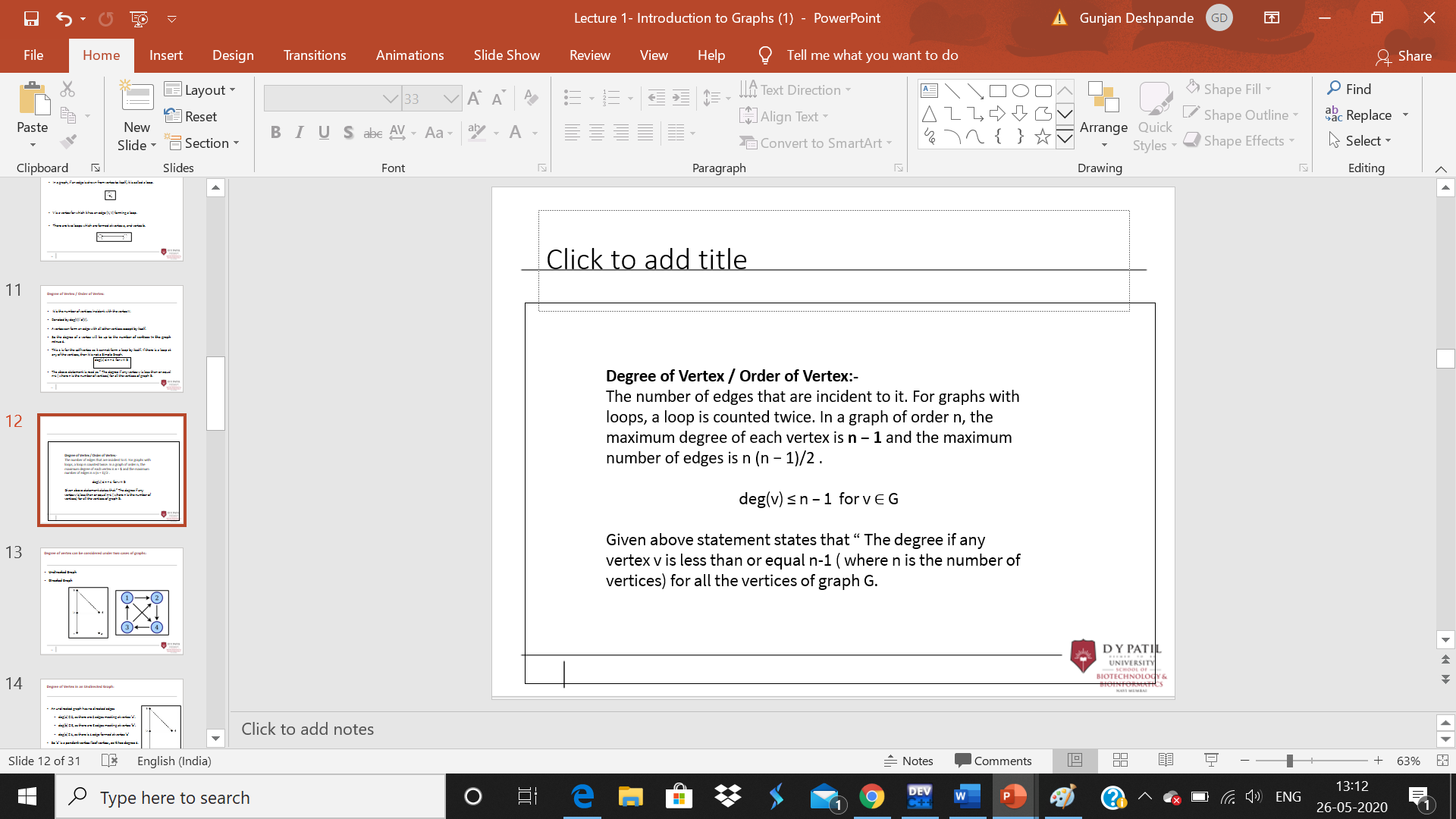
a b

c d e



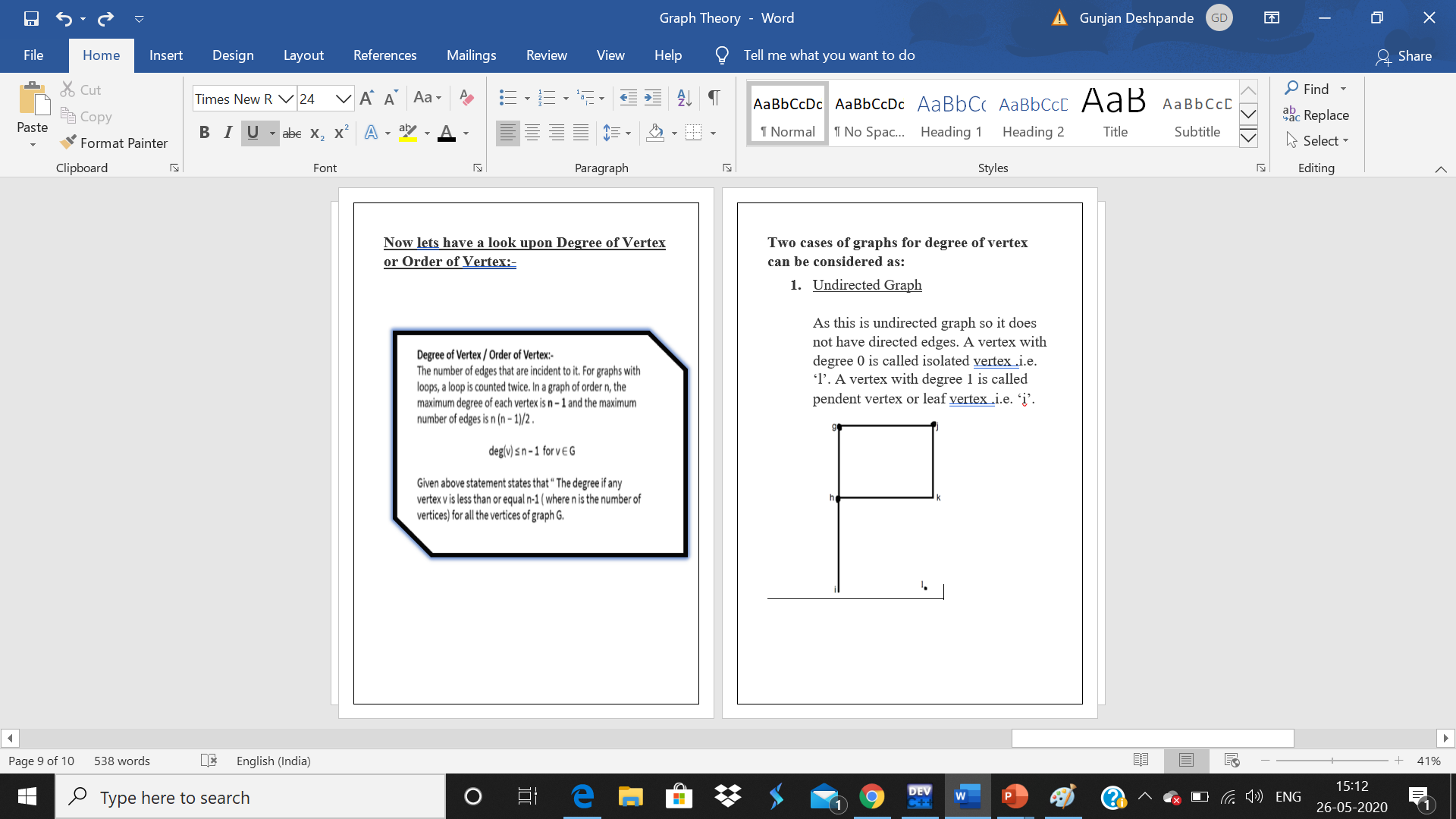
* + from ‘g’ to ‘l’ is 1 (‘gl’),
  + from ‘g’ to ‘k’ is 1 (‘gk’),
  + from ‘g’ to ‘h’ is 1 (‘gh’),
  + from ‘g’ to ‘j’ is 2 (‘gh’-‘hj’) or (‘gk’-‘kj’),

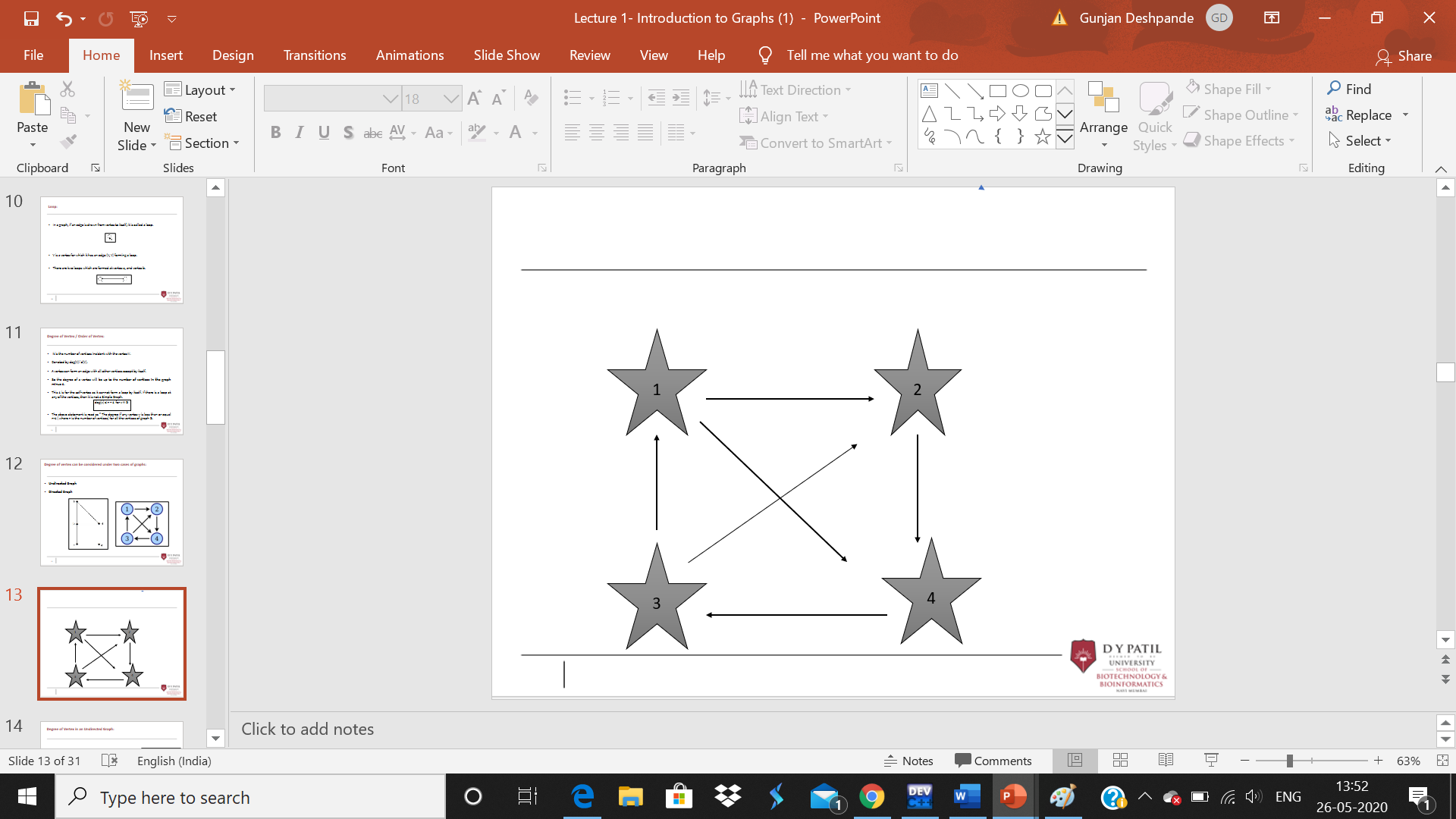
**Now lets have a look upon Degree of Vertex or Order of Vertex:-**



**Two cases of graphs for degree of vertex can be considered as:**

1. Undirected Graph

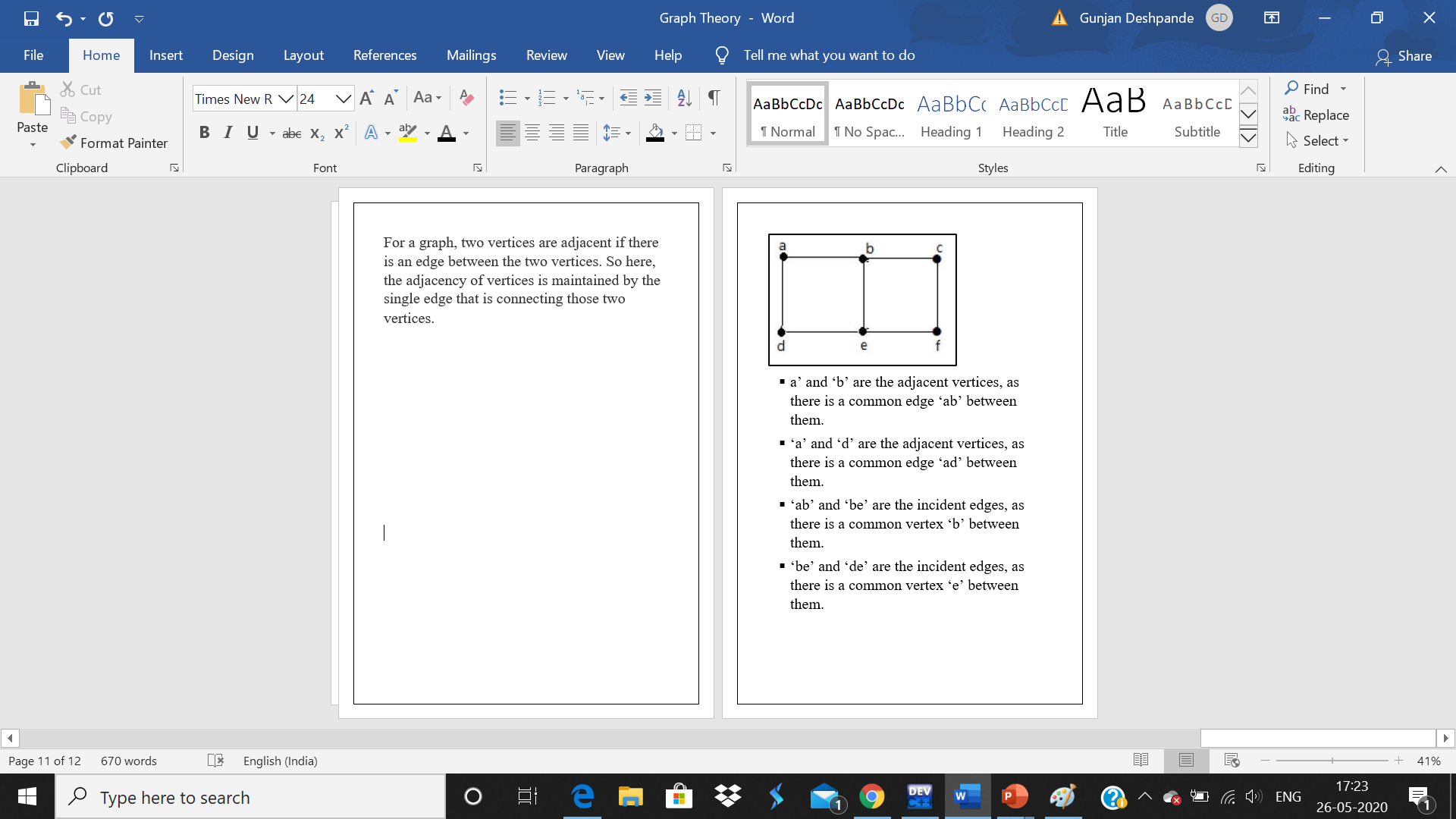


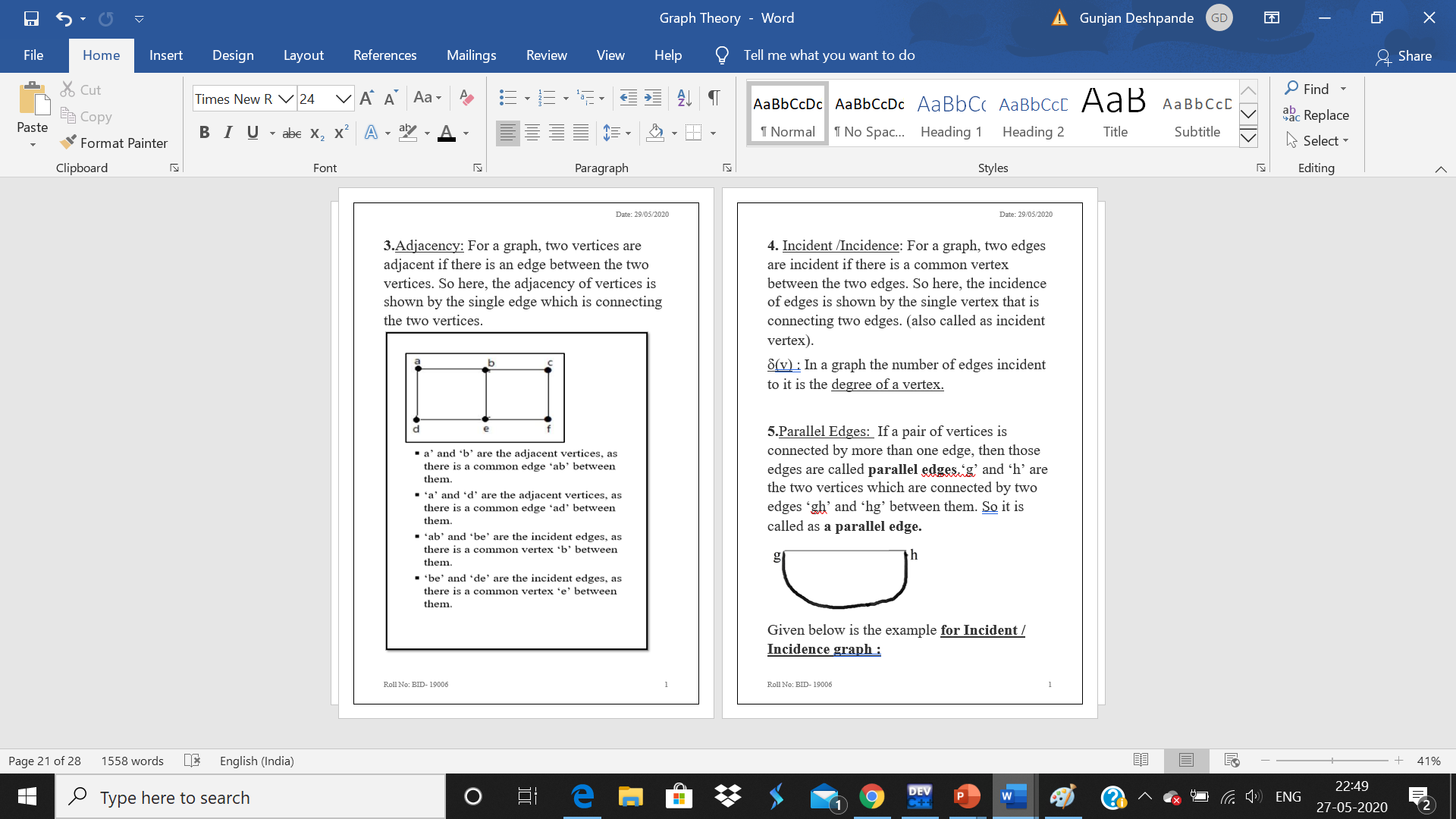
1. Directed Graph

Degree of Vertex in a Directed Graph:

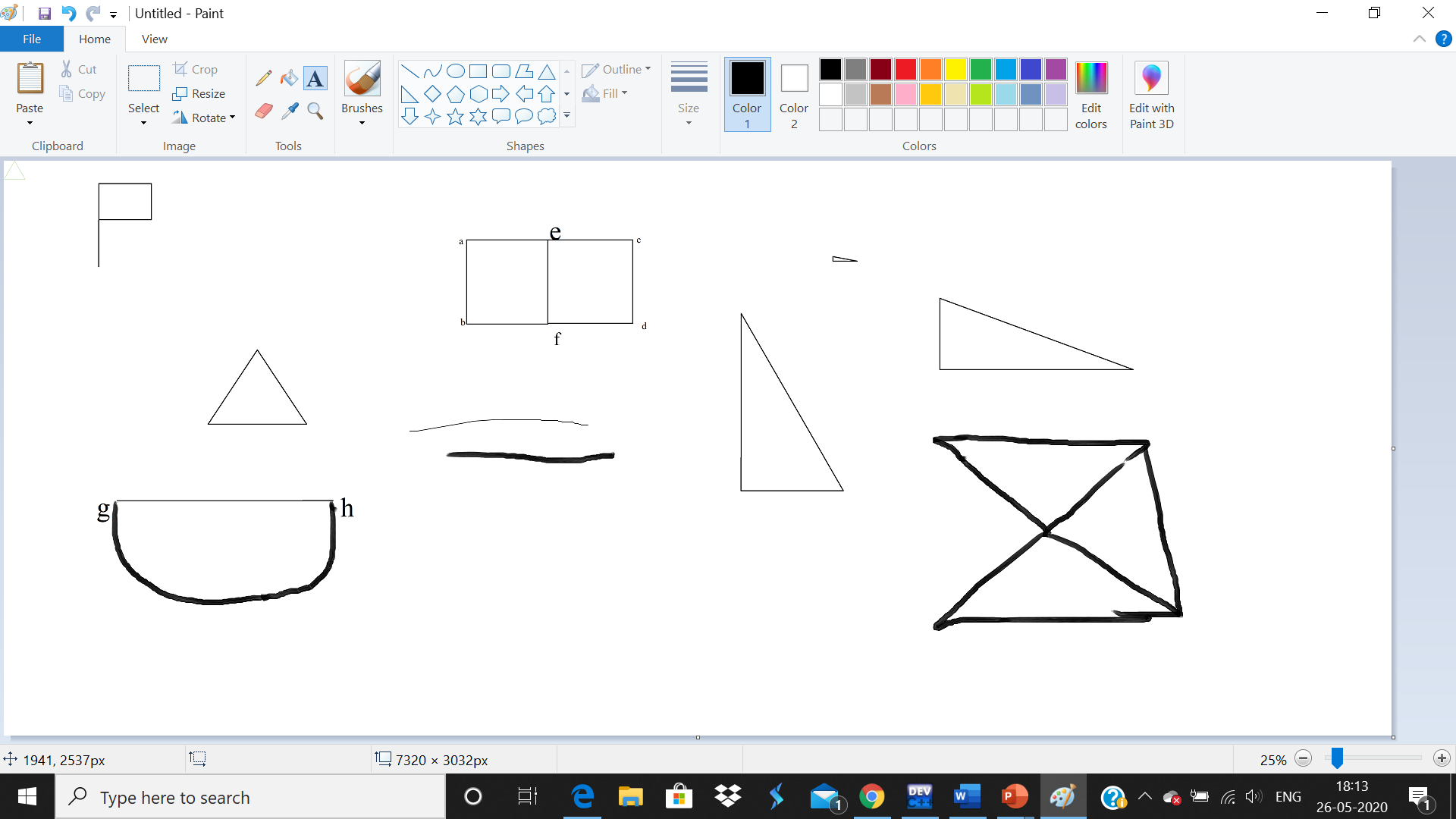
So for a directed graph, each of the vertex has indegree and outdegree.

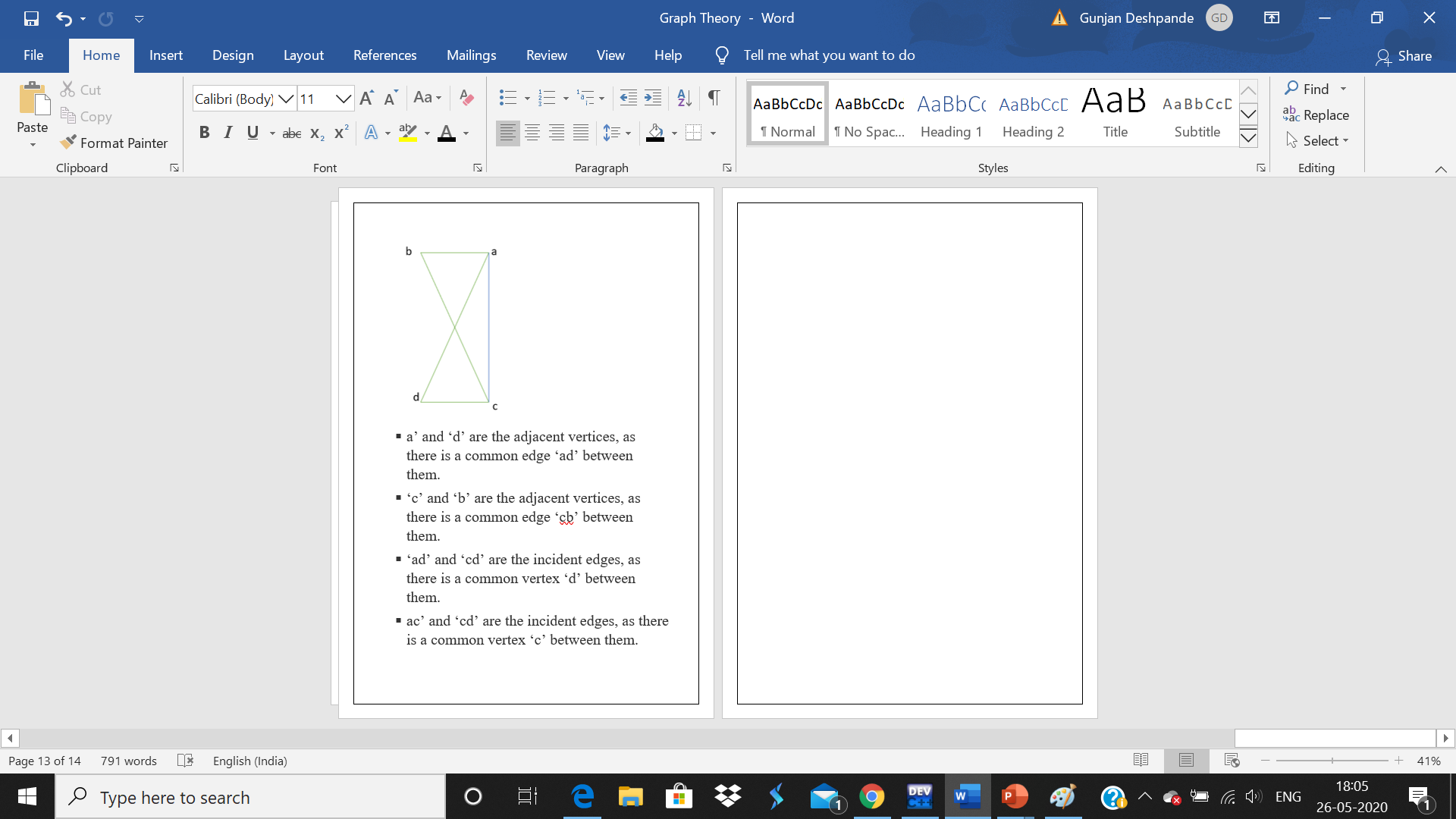
* Indegree of a Graph: The number of edges which are coming into the vertex V is indegree of vertex V.Denoted as deg+(V).
* Outdegree of a Graph: Number of edges which are going out from the vertex V is the outdegree of vertex V.Denoted as deg-(V).

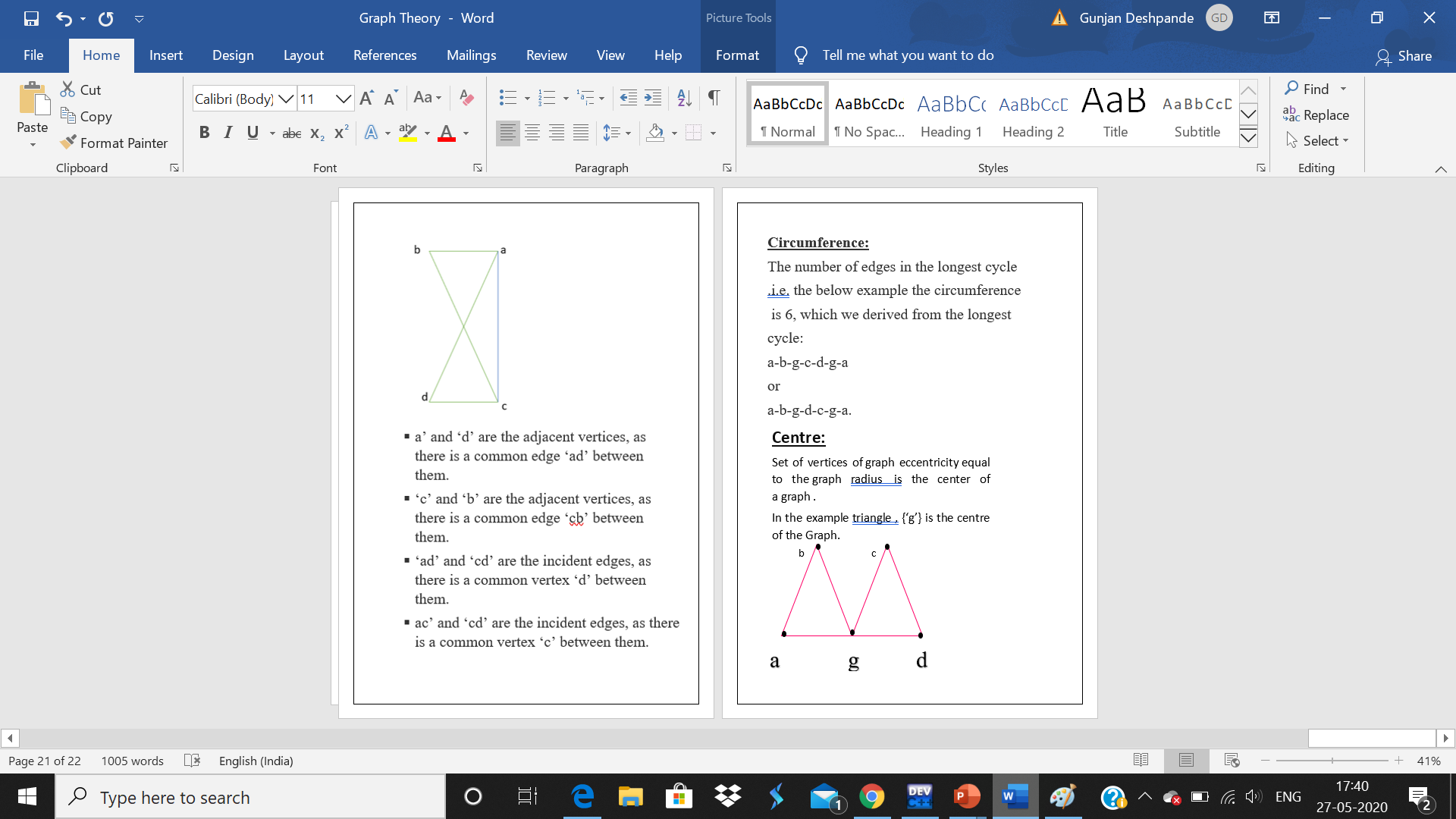
**3.**Adjacency: For a graph, two vertices are adjacent if there is an edge between the two vertices. So here, the adjacency of vertices is shown by the single edge which is connecting the two vertices. 



**5.**Parallel Edges: If a pair of vertices is connected by more than one edge, then those edges are called **parallel edges**.‘g’ and ‘h’ are the two vertices which are connected by two edges ‘gh’ and ‘hg’ between them. So it is called as **a parallel edge.**



Given below is the example **for Incident / Incidence graph** 



**Girth:** g(G)

The number of edges in the shortest cycle of ‘G’ is called its Girth. In the below example graph, the Girth of the graph is 4, which we derived from the shortest cycle:

a b c

d e

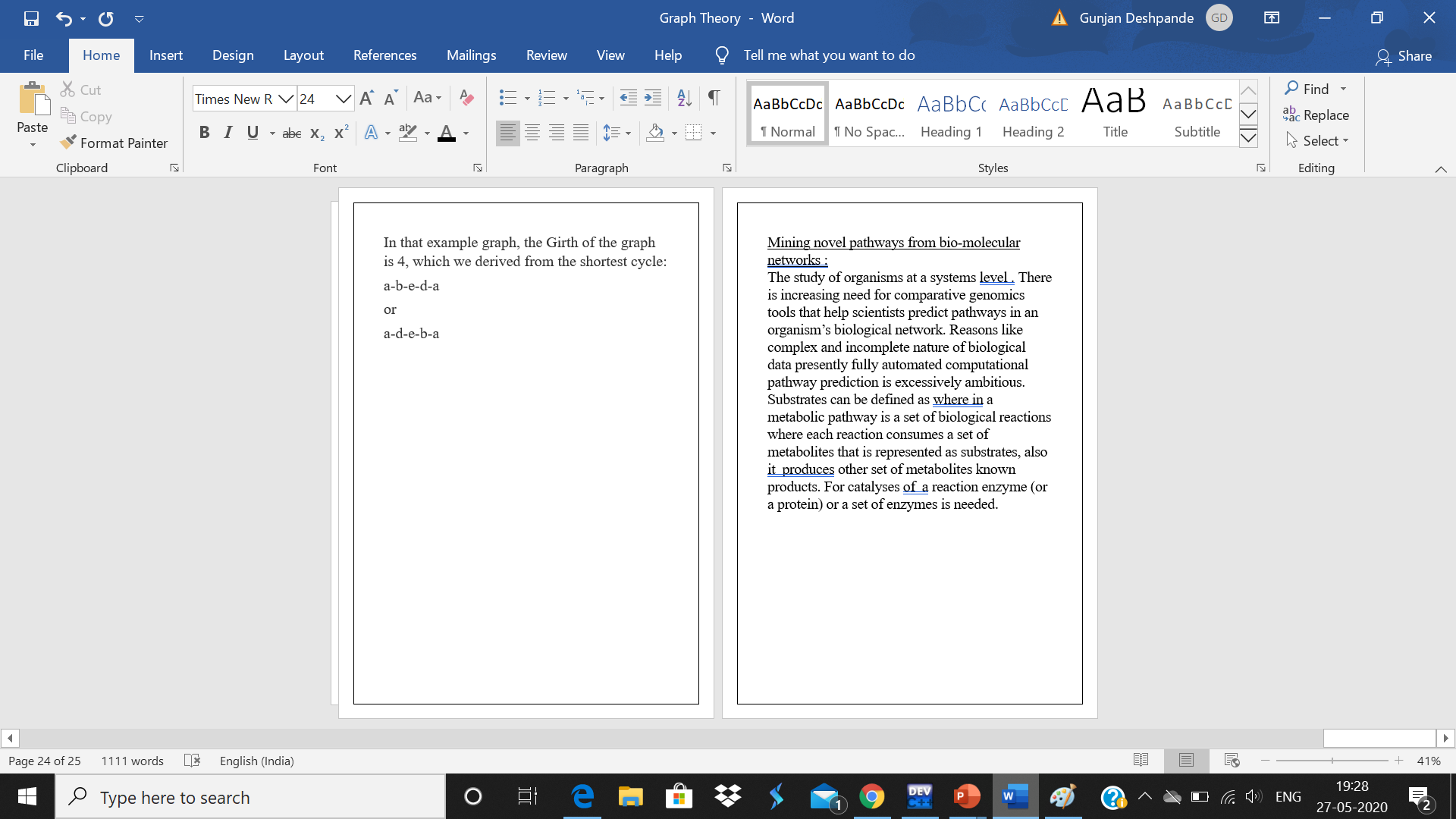
f g

In that example graph, the Girth of the graph is 4, which we derived from the shortest cycle:

a-b-e-d-a

or

a-d-e-b-a



Identifying motifs or functional modules in biological networks :

The engineering systems, which are traditionally described by networks such as flow charts can readily be compared biological systems. Then what will happen is biological networks and engineered networks are seen to share structural principles such as modularity and recurrence of circuit elements. Modularity both the biological systems function and engineering are organized. At different levels subroutines in software and replaceable parts in machines can be a decomposed into a functional modules by Engineering systems. There is no consensus on the precise groups of genes and interactions that form modules in biological network, is clear that they possess a modular structure .

patterns again and again throughout a network.

Conclusion: An increasing speed is adapted by bio-molecular interactions that is becoming available at huge rate that enables a glimpse into complex rganisms also the availability of unprecedented amounts of data, has led to considerable gcellular networks. A straightforward way to represent this information is by Mathematical graph theory also many graph-based models can exploit global and local characteristics of these networks relevant to cell biology. A great way for a more systematic approach to the analysis of living orowth of activity in the theory and analysis of complex biological networks in recent years . Advantages by network graphs are that they are very simple to reason about, and correspond by and large to the information that is globally available today on the network level. In Biology networks are ubiquitous occurring at all levels from biochemical reactions within the cell up to the complex webs of social and sexual interactions that govern the dynamics of disease spread through human populations. There are many biological processes appear to require more detailed model although the binary relation information does represent a critical aspect of interaction networks. For diseases such as Cancer a detailed comprehensive understanding of these networks is needed to develop more sophisticated and effective treatment strategies.