

# Euler paths and circuits on digraphs and genome sequencing

Nilay Bhatt

Central Connecticut State University

*nilaybhatt@my.ccsu.edu*

June 1, 2017

# Overview

- 1 Definitions
- 2 Eulerian Graphs
- 3 Applications
- 4 Genome Sequencing

# Definitions

## Definition (Graph)

A **graph**  $G$  consists of a non-empty finite set  $V(G)$  of elements called **vertices**, and a finite 'family'  $E(G)$  of unordered pairs of (not necessarily distinct) elements of  $V(G)$  called **edges**.

# Definitions

## Definition (Graph)

A **graph**  $G$  consists of a non-empty finite set  $V(G)$  of elements called **vertices**, and a finite 'family'  $E(G)$  of unordered pairs of (not necessarily distinct) elements of  $V(G)$  called **edges**.

- Too formal? Think of a graph as a bunch of tennis balls connected with a thread.

# Definitions

## Definition (Graph)

A **graph**  $G$  consists of a non-empty finite set  $V(G)$  of elements called **vertices**, and a finite 'family'  $E(G)$  of unordered pairs of (not necessarily distinct) elements of  $V(G)$  called **edges**.

- Too formal? Think of a graph as a bunch of tennis balls connected with a thread.

## Definition (Digraph)

A **digraph**  $G$  consists of a non-empty finite set  $V(G)$  of elements called **vertices**, and a finite 'family'  $E(G)$  of ordered pairs of distinct elements of  $V(G)$  called **directed edges**.

# Definitions

## Definition (Graph)

A **graph**  $G$  consists of a non-empty finite set  $V(G)$  of elements called **vertices**, and a finite 'family'  $E(G)$  of unordered pairs of (not necessarily distinct) elements of  $V(G)$  called **edges**.

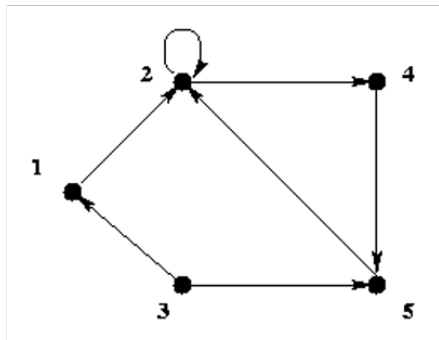
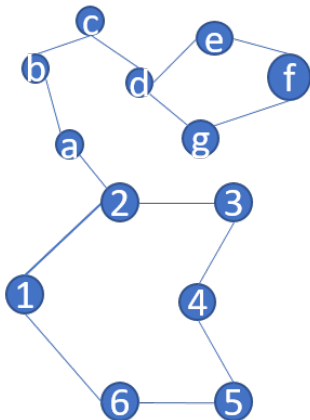
- Too formal? Think of a graph as a bunch of tennis balls connected with a thread.

## Definition (Digraph)

A **digraph**  $G$  consists of a non-empty finite set  $V(G)$  of elements called **vertices**, and a finite 'family'  $E(G)$  of ordered pairs of distinct elements of  $V(G)$  called **directed edges**.

- NYC grid with all the streets being one-way, which then restricts our movement on the graph.

# Graphs and Digraphs



# Euler Paths vs. Circuits

## Definition (Euler Path)

An **Euler path** on a graph  $G$  is a special walk that uses each edge exactly once, and it starts and ends at **different** vertices.

## Definition (Euler Circuit)

An **Euler circuit** on a graph  $G$  is a walk that uses each edge exactly once, and it starts and ends at the **same** vertex.



# Criterion for an Euler path or circuit on a graph

## Euler Path

A given graph  $G$  has an Euler path if and only if the graph is connected and **all but 2 vertices in the graph are of odd degree.**

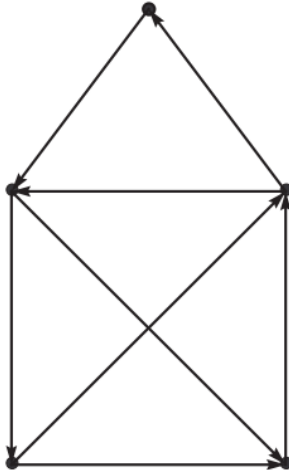
# Criterion for an Euler path or circuit on a graph

## Euler Path

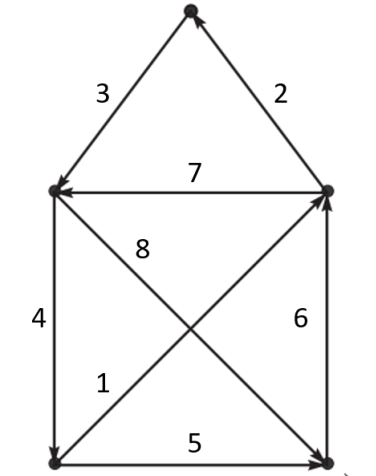
A given graph  $G$  has an Euler path if and only if the graph is connected and **all but 2 vertices in the graph are of odd degree.**

- What extra condition do you think we need to add for a digraph to have an Euler path?

# Can we find an Euler Path?



# YES WE CAN!



# Criterion for an Euler path on a digraph

## Euler Path

A given graph  $G$  has an Euler path if and only if the graph is connected and **all but 2 vertices in the graph are of odd degree.**

## Euler Path on a digraph

A digraph  $D_g$  has an Euler path iff the graph is connected and **all but 2 vertices in the graph are of odd degree.** plus "the  $|indegree - outdegree| = 1$  for given two vertices in  $D_g$ .

# Criterion for an Euler circuit on a digraph

## Euler Circuit

A graph  $G$  has an Euler circuit if and only if all the vertices are of **even degree** and is connected.

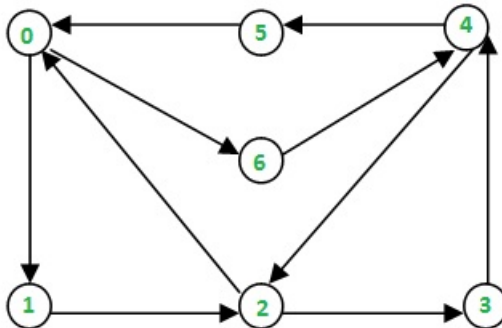
# Criterion for an Euler circuit on a digraph

## Euler Circuit

A graph  $G$  has an Euler circuit if and only if all the vertices are of **even degree** and is connected.

- What extra condition do you think will be needed on a digraph for it to have an Euler circuit?

# Euler Circuit on digraph



**Euler Circuit : 0 -> 6 -> 4 -> 5 -> 0 -> 1 -> 2 -> 3 -> 4 -> 2 -> 0**



## Criterion for an Euler circuit on a digraph

### Euler Circuit

A graph  $G$  has an Euler circuit if and only if all the vertices are of **even degree** and is connected.

### Euler circuit on a digraph

A graph  $D_g$  has an Euler circuit if and only if all the vertices are of even degree and is connected and the **indegree = outdegree** for all vertices.

- A formal proof done as part of the final project in Senior Seminar Spring 2017.

# How to find an Euler path and/or circuit on a graph?

Recall: For an Euler Path all but 2 vertices must be odd and an euler circuit all vertices must be of even degree.

# How to find an Euler path and/or circuit on a graph?

Recall: For an Euler Path all but 2 vertices must be odd and an euler circuit all vertices must be of even degree.

While traversing (removing edges once traversed) if removal of a single edge from the remaining connected graph makes it disconnected then such an edge is called a **bridge**.

- To traverse an Euler path, start at either of the odd vertex.

# How to find an Euler path and/or circuit on a graph?

Recall: For an Euler Path all but 2 vertices must be odd and an euler circuit all vertices must be of even degree.

While traversing (removing edges once traversed) if removal of a single edge from the remaining connected graph makes it disconnected then such an edge is called a **bridge**.

- To traverse an Euler path, start at either of the odd vertex.
- Follow edges one at a time and if you come across a bridge and a non-bridge:
  - Bridge-Non-Bridge: Choose Non-bridge

# How to find an Euler path and/or circuit on a graph?

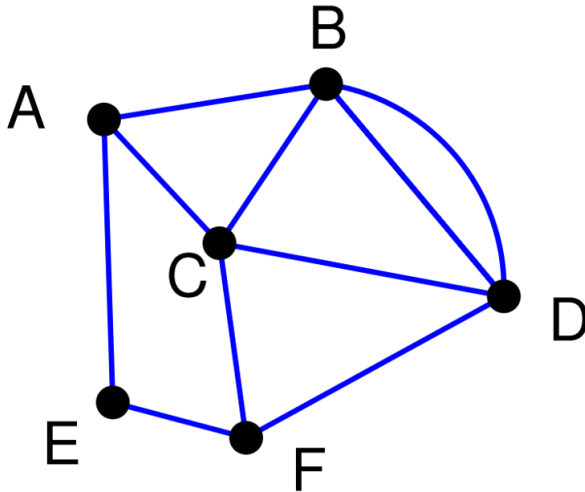
Recall: For an Euler Path all but 2 vertices must be odd and an euler circuit all vertices must be of even degree.

While traversing (removing edges once traversed) if removal of a single edge from the remaining connected graph makes it disconnected then such an edge is called a **bridge**.

- To traverse an Euler path, start at either of the odd vertex.
- Follow edges one at a time and if you come across a bridge and a non-bridge:
  - Bridge-Non-Bridge: Choose Non-bridge

This is called **Fleury's Algorithm**

## Find an Euler Path?



# What's next?

This is cool, but for all the applied mathematicians in the room and computer engineers like me the question becomes:

**How does that help me?**

# Applications

## ① Mathematical modeling



# Applications

- 1 Mathematical modeling
- 2 DNA fragment reconstruction

# Applications

- 1 Mathematical modeling
- 2 DNA fragment reconstruction
- 3 Postman problem

# Applications

- 1 Mathematical modeling
- 2 DNA fragment reconstruction
- 3 Postman problem
- 4 Traveling salesman problem
- 5 many more ....

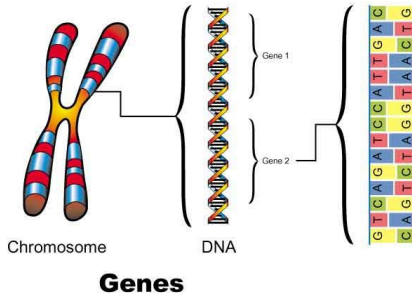
# Applications

- ① Mathematical modeling
- ② **DNA fragment reconstruction**
- ③ Postman problem
- ④ Traveling salesman problem
- ⑤ many more ....

# Genome Sequencing

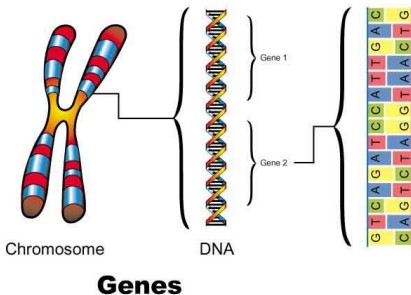
## DNA

- Protein that stores the genetic information of a living organism



# Genome Sequencing

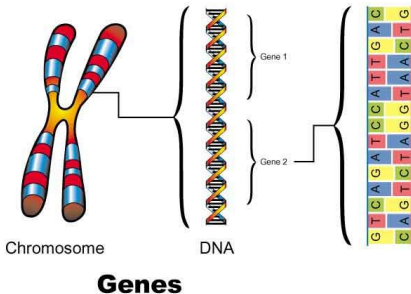
## DNA



- Protein that stores the genetic information of a living organism
- Four components to it: A,T,G,C. A connects to T and G connects to C

# Genome Sequencing

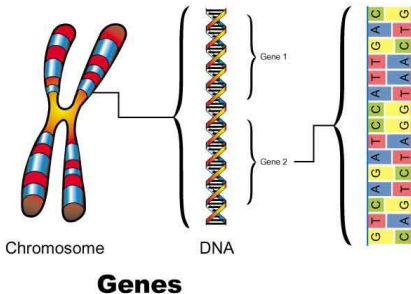
## DNA



- Protein that stores the genetic information of a living organism
- Four components to it: A, T, G, C. A connects to T and G connects to C
- Fragments of DNA is used in genetic research and discovery
- Strings of such **ATGC** pairs have this information stored in them

# Genome Sequencing

## DNA



- Protein that stores the genetic information of a living organism
- Four components to it: A, T, G, C. A connects to T and G connects to C
- Fragments of DNA is used in genetic research and discovery
- Strings of such **ATGC** pairs have this information stored in them



# Genome Sequencing

- Reads are taken of a known DNA strand (Sanger, 1977)
- They vary in size from 30-800 nucleotides
- Reads are taken in an overlapping form
- Reconstructing the specific genome with the overlapping to create a superstring is NP-complete

# Shortest Super String Problem

Problem: **Given a set of strings, find a shortest string that contains all of them** - NP Complete

## The Shortest Superstring problem

Set of strings: {000, 001, 010, 011, 100, 101, 110, 111}

Concatenation  
Superstring 000 001 010 011 100 101 110 111

Shortest  
superstring

000  
0 0 0 1 1 0 1 0 0  
001  
111  
101  
100

# De Bruijn Sequence Approach

- Given a string, split it into  $k$ -mers

# De Bruijn Sequence Approach

- Given a string, split it into  $k$ -mers
- Then split each  $k$ -mer into  $k - 1$ -mer

# De Bruijn Sequence Approach

- Given a string, split it into  $k$ -mers
- Then split each  $k$ -mer into  $k - 1$ -mer
- Now assign the two  $k - 1$ -mer as vertices and the  $k$ -mer as the name of the directed edge

# De Bruijn Sequence Approach

- Given a string, split it into  $k$ -mers
- Then split each  $k$ -mer into  $k - 1$ -mer
- Now assign the two  $k - 1$ -mer as vertices and the  $k$ -mer as the name of the directed edge
- Consolidate the duplicate edges (collapse them, while conserving the directed edges)

# De Bruijn Sequence Approach

- Given a string, split it into  $k$ -mers
- Then split each  $k$ -mer into  $k - 1$ -mer
- Now assign the two  $k - 1$ -mer as vertices and the  $k$ -mer as the name of the directed edge
- Consolidate the duplicate edges (collapse them, while conserving the directed edges)
- The final directed Euler Path is the super string

# What a $k$ -mer composition of a given genome string

Given a string : ***TAATGCCATGGGATGTT***, what is a 3-mer composition?

$= \{TAA, AAT, ATG, TGC, GCC, CCA, CAT, ATG, TGG, GGG, GGA, GAT, ATG, TGT, GTT\}$

- We use these 3-mers as labels of directed edges.



# What a $k$ -mer composition of a given genome string

Given a string : **TAATGCCATGGGATGTT**, what is a 3-mer composition?

$= \{TAA, AAT, ATG, TGC, GCC, CCA, CAT, ATG, TGG, GGG, GGA, GAT, ATG, TGT, GTT\}$

- We use these 3-mers as labels of directed edges.

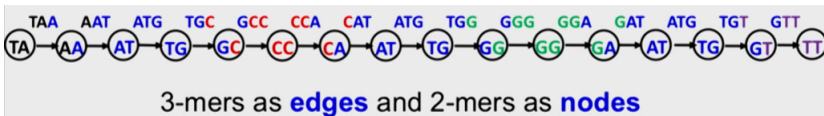


# What a $k$ -mer composition of a given genome string

Given a string : **TAATGCCATGGGATGTT**, what is a 3-mer composition?

$= \{TAA, AAT, ATG, TGC, GCC, CCA, CAT, ATG, TGG, GGG, GGA, GAT, ATG, TGT, GTT\}$

- We use these 3-mers as labels of directed edges.



# Table

Treatments	Response 1	Response 2
Treatment 1	0.0003262	0.562
Treatment 2	0.0015681	0.910
Treatment 3	0.0009271	0.296

Table: Table caption

# Theorem

Theorem (Mass–energy equivalence)

$$E = mc^2$$

# Verbatim

## Example (Theorem Slide Code)

```
\begin{frame}
\frametitle{Theorem}
\begin{theorem}[Mass--energy equivalence]
 $E = mc^2$ 
\end{theorem}
\end{frame}
```

# Figure

Uncomment the code on this slide to include your own image from the same directory as the template .TeX file.

# Citation

An example of the `\cite` command to cite within the presentation:

This statement requires citation [Smith, 2012].

# References



John Smith (2012)

Title of the publication

*Journal Name* 12(3), 45 – 678.



# The End