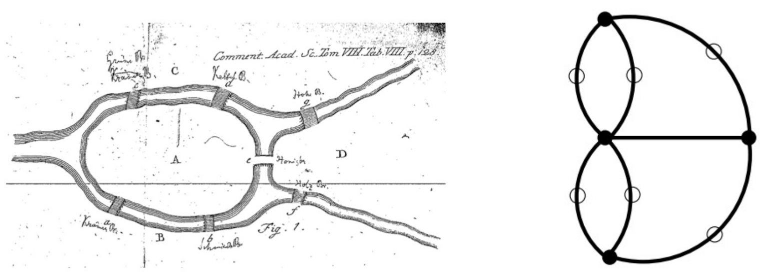
Final Project: Circuits in Graphs
Date: May 10, 2019
Author: Iris Liu, Taryn Waite, and Curtis Zhuang

This project explores two applications of paths and circuits in graphs, the Könisberg Bridge Problem and Hamiltonian paths in DNA reconstruction. All information was obtained from our course textbook [1].

1. Könisberg Bridge Problem and Eulerian circuits

An early application of what was to become graph theory was the **Könisberg Bridge Problem** in 1732, when a question about paths and bridges was posed in the town of Könisberg, Prussia. The town had seven bridges, and the question was whether it was possible to cross all of them without crossing any twice.



Map of the Könisberg bridges (left) and the corresponding graph [1]

As shown above, the situation can be modeled with a graph, where each land mass is represented by a vertex and each bridge between land masses is represented by an edge. Since there are some land masses adjacent to more than once bridge, we add a vertex in the center of each bridge.

Then, the problem is to determine whether there is an **Eulerian path** in the graph, which is a path that traverses each edge in the graph exactly once. Mathematician Leonhard Euler proved that, in the case of the Könisberg Bridge Problem, no such path existed.

The proof of the Königsberg Bridge Problem can done using contradiction. We suppose there exists a path that traverses each edge exactly once. In order for this to happen, for every vertex other than the starting and ending points, the path must enter and leave the vertex through two different edges, meaning that the vertices must be adjacent to an even number of edges. However, every vertex in the Königsberg Bridge graph is adjacent to an odd number of edges, so the path we are looking for must not exist.

From here, we can extend our question to the general conditions under which there exists an **Eulerian circuit**, which is an Eulerian path that is closed, in a graph. In order to answer this question, we will introduce the idea of **valence**, which is the number of edges that are not loops for which a vertex v is an endpoint plus twice the number of loops adjacent to v. It turns out that graphs in which each vertex has an even valence contain Eulerian circuits. This is theorem 9.7.2, which we will prove later.

2. Hamiltonian Path and DNA Reconstruction

The idea of paths in graphs is also applied to current biologists' efforts to reconstruct DNA sequences. Long DNA strands are cut into fragments in order to be readable by technology; however, the cutting process does not preserve the order of the sequence, so in order to reconstruct the initial sequence, we need to cut multiple copies of it and compare the resulting fragments. For example, consider the DNA sequence

$$TAATGCCA.$$

Possible fragments that multiple copies of this sequence could be cut into include:

$$TAA\ \ AAT\ \ ATG\ \ TGC\ \ GCC\ \ CCA.$$

Looking at the fragments, TAA has two AA bases, which only appear again in the AAT fragment, we therefore know that

$$TAA\ \ AAT\ \Rightarrow TAAT.$$

Following the exact same strategy, we are able to use all of the fragments to reconstruct the order of the original sequence:

$$\begin{aligned} TAAT\ \ ATG\ &\Rightarrow TAATG \\ TAATG\ \ TGC\ &\Rightarrow TAATGC \\ TAATGC\ \ GCC\ &\Rightarrow TAATGCC \\ TAATGCC\ \ CCA\ &\Rightarrow TAATGCCA \end{aligned}$$

Notice that we are able to reconstruct the sequence nicely because each time the overlapping bases only appear in one unvisited fragments; whereas, the presence of repeats could hinder the reconstruction process and lead to a wrong order. That is, at some point, we are not able to add unvisited fragments into the sequence and therefore to use every fragment, some kind of repetitions of fragments must be involved.

Observe that each fragment could be viewed as one vertex, and the overlapping bases, indicating possible adjacency of two fragments, could be viewed as directed edges, pointing from one fragment, or vertex, to another. The correct reconstruction should be a **Hamiltonian path**, which is a path where each vertex is visited only once, meaning that each fragment is only used once.

Notice that an Eulerian path crosses every edge exactly once, while a Hamiltonian path passes through each vertex exactly once. For a Hamiltonian path, there could be edges that are not visited, while an Eulerian path crosses every edge once but could pass through some vertices more than once.

Theorem 9.7.2. *Suppose that G is a nonempty finite connected graph such that every vertex has even valence. Then G contains an Eulerian circuit.*

This theorem allows us to generalize the idea of the Könisberg Bridge Problem to all graphs. We will first prove the theorem for graphs whose vertices all have valence 2. Note that this section of the proof is taken verbatim from [1]. Next, we will use this result to prove the theorem for graphs whose vertices all have even valence by using a proof by complete induction, where we induct on the number of vertices that have valence of at least 4.

Proof of Theorem 9.7.2. We begin by proving the theorem for graphs with the property that every vertex has valence 2. We prove this by minimal counter-example.

Assume that G is a nonempty finite connected graph such that every vertex has valence 2. Assume, for a contradiction, that G does not contain an Eulerian circuit. Out of all such counter-examples, we may assume that we selected G to minimize the number of vertices. Since G is non-empty, it has a vertex v. By hypothesis, v is the endpoint of two (not necessarily distinct edges) e_- and e_+. If $e_- = e_+$ then, since G is connected, it is a loop. The path v, v is then an Eulerian circuit. If $e_- \neq e_+$, let G' be the graph obtained from G by merging e_- and e_+ into a single edge e and removing v as a vertex. Notice that since e_- and e_+ had other endpoints, distinct from v as v had valence 2. Thus, G' is still connected (do you see why?) and nonempty. It also has one less vertex than G. Thus, G' cannot be a counterexample and so has an Eulerian circuit

$$\alpha = v_0, v_1, \ldots, v_n.$$

Since α traverses every edge of G' exactly once, there exists an $i \in \{0, \ldots, n-1\}$ such that the ends of α are the ends of e and these ends are $\{v_i, v_{i+1}\}$. Splitting e back into e_- and e_+, the path

$$v_0, v_1, \ldots, v_i, v, v_{i+1}, \ldots, v_n.$$

is then an Eulerian circuit in G. But this contradicts the fact that G is a counter-example to the theorem. Thus, the theorem holds for all nonempty finite connected graphs such that every vertex has valence 2.

Now, we will prove that the theorem holds for nonempty finite connected graphs G with all vertices having even valence. We will use complete induction on the number of vertices with valence ≥ 4.

<u>Base Case:</u> G has 0 vertices with valence ≥ 4. In this case, since all vertices in G must have even valence and G must be connected, either G consists of one vertex with valence 0 or all of the vertices in G have valence 2. In the first case, there is an Eulerian path consisting of the vertex. In the second case, as proved above, G has an Eulerian circuit.

Inductive Step: Assume that for some $k \in \mathbb{N}^*$, there exists an Eulerian circuit in all finite connected graphs that have $\leq k$ vertices with valence ≥ 4. We will show that if G has $k+1$ vertices with valence ≥ 4, then there exists an Eulerian circuit in G.

Let v be an arbitrary vertex in G with valence ≥ 4. "Let E be the set of non-loop edges having at least one endpoint at v. Since the valence of v is even, we may partition the endpoints of the edges in E into pairs. Form a new graph G' by splitting the vertex v into vertices $w_1, ..., w_n$, where $n = valence(v)/2$, such that each new vertex w_j has valence equal to 2. If e is a loop with its endpoints at v, in the new graph it shows up as a loop at a vertex not incident to any other edges"[1]. Note that if there is no pair of vertices v_1, v_2 in G such that the only paths between v_1 and v_2 include v, then the new graph G' will still be connected. If not, G will be split into two or more connected components. Thus, G' has $p \in \mathbb{N}$ connected components, $G_1, G_2, ..., G_p$ (see Figure 1); p will be 1 in the first case and greater than 1 in the second, and the maximum possible value of p is n.

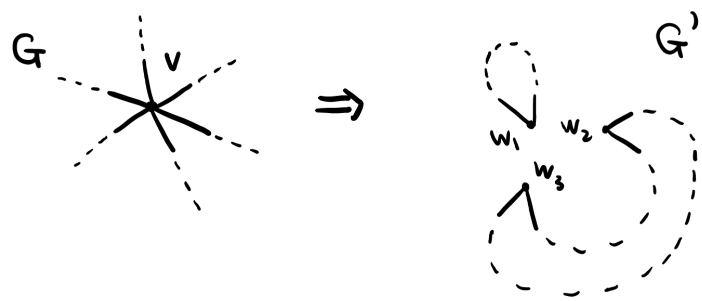


FIGURE 1. Splitting v into into vertices $w_1, ..., w_n$, resulting in connected component(s).

Each component G_i of G' is nonempty because it contains at least one vertex, w_j, and the two edges adjacent to w_j. Each G_i is also finite because each of its vertices is a vertex in G, and G is finite. Note that we obtained G' from G by replacing v, which had valence ≥ 4, with vertices w_j that each have valence 2, so the resulting G' has one fewer vertex with valence ≥ 4 than G, so it has $k+1-1 = k$ such vertices. Since each G_i is a component of G', it can have at most the same number of such vertices as G'. Thus, each G_i has $\leq k$ vertices with valence ≥ 4.

By the inductive hypothesis, each connected component G_i of G' has an Eulerian circuit with m_i vertices

$$\alpha_i = v_{i_0}, v_{i_1}, ..., v_{i_{m_i}}.$$

By a cyclic shift, let $v_{i_0} = v_{i_{m_i}} = w_j$. We can thus rewrite the Eulerian circuits as

$$\alpha_{ij} = w_j, v_{i_1}, ..., v_{i_{m_i - 1}}, w_j.$$

Now, we can merge the connected components of G' back into the original graph G by setting each vertex w_i equal to the original vertex v. We can thus rewrite the Eulerian circuits above as

$$\alpha'_{ij} = v, v_{i_1}, ..., v_{i_{m_i-1}}, v,$$

Each of which is a path in G that starts and ends at v. Now we will construct a new circuit α through G by combining each circuit α'_i:

$$\alpha = v, v_{1_1}, ..., v_{1_{m_1-1}}, v, v_{2_1}, ..., v_{2_{m_2-1}}, v, ..., v, v_{n_1}, ..., v_{n_{m_i-1}}, v.$$

Note that the only modification of G to form G' was splitting the vertex v into multiple vertices. Thus, G' has the same edges as G, and every edge in G is in one of the connected components of G'. Since each circuit α_i in G_i was Eulerian, each of them traversed each edge in the connected component G_i exactly once. Thus, α traverses each edge in G exactly once. Therefore, α is an Eulerian circuit in G. □

References

[1] Taylor, Scott A. *The Structures of Mathematics*, available online at `http://web.colby.edu/sataylor`