Gromov-Hausdorff Distance Between Metric Graphs

Jiwon Choi
St. Mark's School

January 2, 2019

Abstract

In this paper we study the Gromov-Hausdorff distance between two metric graphs. We compute the precise value of the Gromov-Hausdorff distance between two path graphs. Moreover, we compute the precise value of the Gromov-Hausdorff distance between a cycle graph and a tree. Given a graph X, we consider a graph Y that results from adding an edge to X without changing the number of vertices. We compute the precise value of the Gromov-Hausdorff distance between X and Y.

1 Introduction

Gromov-Hausdorff distance, a concept that was first introduced by Gromov in his study of convergence of manifolds [2], is a very important concept in modern differential geometry because this notion allows us to evaluate the distance between two distinct metric spaces.

Suppose we have a metric space. The standard notion of distance allows us to determine the distance between two subspaces of a metric space. However, when we are considering the distance between two subspaces that intersect, this definition of distance does not give us enough information to distinguish between the two subspaces. This definition of distance was later improved by Hausdorff. Hausdorff proposed a way of finding the distance between two subspaces of the metric space. Out of all the Hausdorff distances that result from different isometric embeddings and different metric spaces, the infimum of all of them is the Gromov-Hausdorff distance. This notion allows us find the distance between two metric spaces, thus allowing us to compare the two metric spaces.

It is often challenging to compute the exact value of Gromov-Hausdorff distances since the process involves finding the infimum of all possible Hausdorff distances. It is relatively simple to find the upper bound for the Gromov-Hausdorff distance. Because the Gromov-Hausdorff distance is the infimum of all the Hausdorff distances resulting from different methods of isometric embeddings, we only need to exhibit one specific isometric embedding such that the condition is satisfied in order to find the upper bound. Nevertheless, it is more challenging to get the exact lower bound for Gromov-Hausdorff distance because the condition has to told true for all possible isometric embeddings.

In this paper, we compute the exact value of the Gromov-Hausdorff distance between specific graphs. The main difficulty in this paper is finding the lower bound of the Gromov-Hausdorff distance. We start out with Lemma 1, which gives the lower bound for the Hausdorff distance between two metric graphs. From Lemma 1, we find out that the Hausdorff distance can be controlled by the distances between points within the two metric graphs that are being considered. While finding the lower bound for the Gromov-Hausdorff distance between specific graphs, we can use these controls to obtain the optimal lower bound. We compute the exact value of the Gromov-Hausdorff distance between path graphs, which are graphs that can be drawn so that all of their vertices and edges lie on a single straight line. Path graphs are subgraphs of many more complicated graphs. Therefore, the result we get for path graphs prove to be significant when we later compute the Gromov-Hausdorff distance between more complicated graphs. We consider other graphs such as cycle graphs and trees.

Finding the Gromov-Hausdorff distance between two graphs has various applications. In the biostatistics field, these results may be helpful when comparing large data sets in high dimensions. H Lee et al in their paper on brain networks [3] discuss possible applications in the biostatistics field by introducing the Gromov-Hausdorff distance as an already well-established method in shape analysis for comparing shapes of networks. Moreover, they [3] argue that the concept can also be used for comparing the shapes of neural networks in brains. The Gromov-Hausdorff distance between graphs has further applications in computer science. In particular, it can be used as a promising method for shape matching and comparison. Such applications may help improve algorithms for face recognition, pattern recognition, or matching of articulated objects [4].

In Section 2, basic definitions are discussed. More specifically, the definitions of the Hausdorff distance and the Gromov-Hausdorff distance are discussed. In Section 3, the proof of the fundamental lemma is given. Section 4 and Section 5 are devoted to computing the exact value of the Gromov-Hausdorff distance between graphs using the lemma. In Section 6, future directions are discussed.

2 Basic definitions

Definition 1. Let X be an arbitrary set. A function $d : X \times X \to \mathbb{R} \cup \{\infty\}$ is a *metric* on X if the following conditions are satisfied for all $x, y, z \in X$:

(1) Identity: $d(x, y) = 0 \iff x = y$,
(2) Commutativity: $d(x, y) = d(y, x)$,
(3) Triangle Inequality: $d(x, z) \leq d(x, y) + d(y, z)$.

Elements of X are called *points* of the metric space; $d(x, y)$ refers to the *distance* between points x and y. The Euclidean distance in $\mathbb{R}^2$ is an example of a metric.

Definition 2. Given a metric space X, a map of $f : X \to Z$ is called an *isometric embedding* if for all $x_1, x_2 \in X$: $d_X(x_1, x_2) = d_Z(f(x_1), f(x_2))$. The metric on Z is denoted as d_Z.

An isometric embedding of X into metric space Z is distance preserving. Using isometric embedding, we can compute the distance between two different spaces by embedding them into a metric space Z.

Definition 3. A *graph* G is an ordered pair $G = (V, E)$ consisting of a set V of vertices, nodes or points together with a set E of edges, which are 2-element subsets of V.

Remark 1. All graphs mentioned in this paper are finite graphs.

Definition 4. Given a point x and a graph Y embedded into a metric space Z, the *distance* between point x and graph Y is defined as

$$d_Z(x, Y) = \inf_{y \in Y} d_Z(x, y).$$

Definition 5. Given two graphs X, Y isometrically embedded into a metric space Z, for all $x \in X, y \in Y$, the *Hausdorff distance* between X and Y is

$$d_H(X, Y) = \max(\sup_{x \in X} d_Z(x, Y), \sup_{y \in Y} d_Z(X, y)).$$

In a metric space Z, there exists a shortest distance from point $x \in X$ to Y. Similarly, there exists a shortest distance from point $y \in Y$ to X. Out of all these distances, the Hausdorff distance between X and Y is the longest distance.

Remark 2. We will say that vertices x and y *correspond* to each other if $d(x, y) = d_H X, Y$.

Definition 6. The *Gromov-Hausdorff distance* between two graphs X and Y is

$$d_{GH}(X, Y) = \inf_{i: X \hookrightarrow Z, j: Y \hookrightarrow Z} d_H(i(X), j(Y)).$$

The Gromov-Hausdorff distance between X and Y is defined to be the infimum of all numbers $d_H(i(X), j(Y))$ for all metric spaces Z and all isometric embeddings $i : X \to Z$ and $j : Y \to Z$. There are infinitely many isometric embeddings of X and Y into any metric space Z and different Hausdorff distances resulting from them. Out of all possible Hausdorff distances, the Gromov-Hausdorff distance between X and Y is the infimum.

Remark 3. When proving the lower bound for the Gromov-Hausdorff distance, the lower bound must hold true for all possible Hausdorff distances. When proving the upper bound for the Gromov-Hausdorff distance, showing one specific isometric embedding that exhibits a Hausdorff distance bounded by the upper bound.

Definition 7. Take two graph X, Y with each n and m vertices Let us denote the vertices of X as x_i for all $1 \leq i \leq n$ and the vertices of Y as y_j for all $1 \leq j \leq m$. Then, isometrically embed graphs X, Y into a metric space Z. We say that x_i *corresponds* to y_j if

$$d_Z(x_i, y_j) = d_Z(x_i, Y).$$

In the rest of this section, we will define specific graphs that will play an important role in the paper later.

Definition 8. We say that graph P_m is a *path graph* of length $m-1$ if it consists of m vertices and $m-1$ edges. Denote the vertices as v_i for all $1 \leq i \leq m$. All vertices v_i other than v_1 and v_m are connected to v_{i-1} and v_{i+1}. We call v_1 and v_m *endpoints*, and they are only connected to v_2 and v_{m-1} respectively.

Figure 1 is a path graph with 6 vertices and 5 edges.

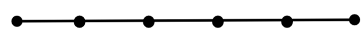

Figure 1: P_6.

Definition 9. A graph C_n is a *cycle graph* if it consists of $n \geq 3$ vertices $v_1, v_2, \ldots v_n$ and n edges, and there is an edge between v_m and v_{m+1} for all $1 \leq m \leq n-1$ and an edge between v_1 and v_n.

Figure 2 is a cycle graph with 5 vertices and 5 edges.

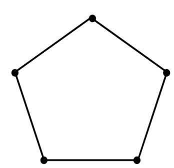

Figure 2: C_5.

Definition 10. A graph X is a *tree* if it is a set of connected edges containing no cycle graphs as a subgraph.

In other words, a tree graph is a connected acyclic graph. Figure 3 shows examples of a tree graph.

Figure 3: Tree graphs.

3 Lower Bound for the Hausdorff Distance

In this section, we will prove a fundamental lemma that will help prove the subsequent theorems in this paper. We use basic definitions to prove this lemma, which gives us the lower bound for the Hausdorff distance between two graphs X and Y. Notably, we use the definition of correspondence and the definition of isometric embedding.

Lemma 1. *Take two graphs X, Y and isometrically embed them into a metric space Z. If $x_i \in X$ corresponds to $y_i \in Y$ and $x_j \in X$ corresponds to $y_j \in Y$, then $d_H(X,Y) \geq \frac{d_X(x_i,x_j)-d_Y(y_i,y_j)}{2}$.*

Proof. Because X and Y are isometrically embedded into a metric space Z,

$$d_X(x_i, x_j) \leq d_Z(x_i, y_i) + d_Y(y_i, y_j) + d_Z(x_j, y_j).$$

Because x_i corresponds to y_i and x_j corresponds to y_j,

$$d_X(x_i, x_j) \leq d_Z(x_i, Y) + d_Y(y_i, y_j) + d_Z(x_j, Y).$$

Moreover,

$$d_X(x_i, x_j) \leq d_H(X, Y) + d_Y(y_i, y_j) + d_H(X, Y).$$

Therefore, we get

$$d_X(x_i, x_j) - d_Y(y_i, y_j) \leq 2d_H(X, Y),$$

and $\frac{d_X(x_i,x_j)-d_Y(y_i,y_j)}{2}$ is the lower bound for $d_H(X, Y)$.

□

Remark 4. In order to find the best possible lower bound for $d_H(X, Y)$, we want to maximize $d_X(x_i, x_j)$ and minimize $d_Y(y_i, y_j)$.

4 Gromov-Hausdorff Distance Between Two Path Graphs

We consider the Gromov-Hausdorff distance between two path graphs. To find the best possible lower bound, we know from Lemma 1 that we must maximize $d_X(x_i, x_j)$ and minimize $d_Y(y_i, y_j)$. In order to do so, we consider the endpoints of the two path graphs. Finding the Gromov-Hausdorff distance between path graphs is significant because path graphs are subgraphs of many more complicated graphs. We compute the precise value of the Gromov-Hausdorff distance between two path graphs.

We prove that $d_{GH}(P_m, P_n) \geq \frac{|m-n|}{2}$ by contradiction, and we prove that $d_{GH}(P_m, P_n) \leq \frac{|m-n|}{2}$ by exhibiting a special isometric embedding of X, Y into a metric space Z.

Theorem 2. *Given two graphs P_m and P_n, the Gromov-Hausdorff distance between P_m and P_n is exactly $\frac{|m-n|}{2}$.*

Proof. Assume $m \geq n$.

First, let us prove that $d_{GH}(P_m, P_n) \geq \frac{m-n}{2}$.

By the definition of the Gromov-Hausdorff distance between two spaces X, Y, in order to prove $d_{GH}(P_m, P_n) \geq \frac{m-n}{2}$, we need to prove that $d_H(P_m, P_n) \geq \frac{m-n}{2}$ holds for any isometric embedding of P_m, P_n into Z. By the definition of Hausdorff distance, we need to prove that either $d_Z(x, P_n) \geq \frac{m-n}{2}$ or $d_Z(P_m, y) \geq \frac{m-n}{2}$ for some point $x \in P_m$ and some point $y \in P_n$.

In order to prove this statement, we want to show that $\sup_{x \in P_m} d_Z(P_n, x) \geq \frac{m-n}{2}$. To maximize the lower bound, we consider the endpoints x_1 and x_m. In other words, we want to show that either $d_Z(P_n, x_1) \geq \frac{m-n}{2}$ or $d_Z(P_n, x_m) \geq \frac{m-n}{2}$. We can prove this by contradiction.

Assume that $d_Z(P_n, x_1) < \frac{m-n}{2}$ and $d_Z(P_n, x_m) < \frac{m-n}{2}$. Then, for some i, j such that $1 \leq i, j \leq n$, $d_Z(x_1, y_i) < \frac{m-n}{2}$ and $d_Z(x_m, y_j) < \frac{m-n}{2}$. Because P_m and P_n are path graphs, each edge is of length 1, thus $d_Y(y_i, y_j) = |i - j|$ and $d_X(x_1, x_m) = m - 1$. Since $d_Z(x_1, y_i) < \frac{m-n}{2}$ and $d_Z(x_m, y_j) < \frac{m-n}{2}$,

$$d_Z(x_1, y_i) + d_Y(y_i, y_j) + d_Z(x_m, y_j) < m - n + |j - i|.$$

Because isometric embedding is distance-preserving,

$$d_X(x_1, x_m) \leq d_Z(x_1, y_i) + d_Y(y_i, y_j) + d_Z(x_m, y_j),$$

and

$$m - 1 < m - n + |j - i|.$$

However, this is a contradiction. Thus, either $d_Z(P_n, x_1) \geq \frac{m-n}{2}$ or $d_Z(P_n, x_m) \geq \frac{m-n}{2}$, and we can conclude that $d_{GH}(P_m, P_n) \geq \frac{m-n}{2}$.

Second, let us prove that $d_{GH}(P_m, P_n) \leq \frac{m-n}{2}$.

In order to prove this, we want to show that there exists an isometric embedding of P_m, P_n into a metric space such that the $d_H(P_m, P_n) \leq \frac{m-n}{2}$.

Consider the isometric embedding shown in Figure 4. We first build a pyramid with edges and vertices as shown below. Construct the pyramid so that the horizontal edges are each of length 1, and the diagonal edges are each of length $\frac{1}{2}$.

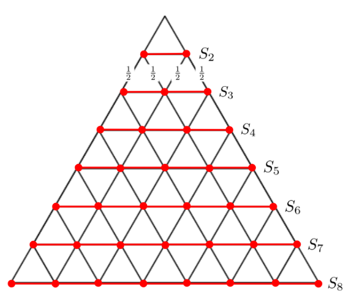

Figure 4: Isometric Embedding for $d_{GH}(P_m, P_n) \leq \frac{m-n}{2}$

Consider the bases of the triangles in each layer as path graphs. The first layer exhibits P_2, the second layer exhibits P_3, and so on. In this isometric embedding, $d_H(P_m, P_n) \leq \frac{m-n}{2}$.

Thus, since $d_{GH}(P_m, P_n) \geq \frac{m-n}{2}$ and $d_{GH}(P_m, P_n) \leq \frac{m-n}{2}$, $d_{GH}(P_m, P_n) = \frac{m-n}{2}$. □

5 Gromov-Hausdorff Distance Between Graphs that Contain Path Graphs as Subgraphs

We consider the Gromov-Hausdorff distance between a cycle graph and a tree. Burago [1] claims in his book that the Gromov-Hausdorff distance between two subspaces of two metric spaces is equal to the Gromov-Hausdorff distance between the two metric spaces. Since path graphs are subgraphs of a cycle graph, we divide up the cycle graph into multiple path graphs and use Lemma 1 to find the distance between those path graphs and the path graphs found in the tree.

We prove the lower bound for the Hausdorff distance by contradiction, and the upper bound is given by exhibiting a specific isometric embedding of the two graphs into a metric space. To find the best upper bound possible, we want to show an isometric embedding that minimizes the Hausdorff distance. Intuitively, we want to divide up the vertices of the cycle graph as evenly as possible to correspond to each vertex of the tree.

Theorem 3. *If C_m is a cycle graph with m vertices and X is a tree where n vertices with $m > n$, then $d_{GH}(C_m, X) = \frac{\lceil \frac{m}{n} \rceil - 1}{2}$.*

Proof. First, we will prove that $d_{GH}(C_m, X) \geq \frac{\lceil \frac{m}{n} \rceil - 1}{2}$. Given all isometric embeddings of C_m, X into a metric space Z, we can find an isometric embedding where there are at least $\lceil \frac{m}{n} \rceil$ vertices of C_m that correspond to one vertex of X, using the pigeon hole argument. One resulting Hausdorff distance is greater than or equal to the Hausdorff distance between a path graph of length $\lceil \frac{m}{n} \rceil$ and a path graph of length 1.

Since $\lceil \frac{m}{n} \rceil$ vertices correspond to a single vertex in X, the longest possible distance between two vertices that correspond to the same vertex is greater than $\lceil \frac{m}{n} \rceil$. This condition holds for any isometric embedding of C_m and X. Therefore, if x_i and x_j are the two vertices that are farthest apart that correspond to the same vertex y_i, then $d_X(x_i, x_j) > \lceil \frac{m}{n} \rceil$. Using Lemma 1,

$$2d_H(C_m, X) \geq \left\lceil \frac{m}{n} \right\rceil - 1,$$

and thus

$$d_H(C_m, X) \geq \frac{\lceil \frac{m}{n} \rceil - 1}{2},$$

for all possible isometric embeddings of C_m and X into a metric space. Therefore, $d_{GH}(C_m, X) \geq \frac{\lceil \frac{m}{n} \rceil - 1}{2}$.

Second, we will prove that $d_{GH}(C_m, X) \leq \frac{\lceil \frac{m}{n} \rceil - 1}{2}$. In order to prove this statement, we need to exhibit an isometric embedding of C_m and X into metric space Z that satisfies the condition.

Consider the following isometric embedding of C_m and X into a metric space Z. Let $p = \lceil \frac{m}{n} \rceil$. Call the vertices of C_m v_i where $1 \leq i \leq m$. Then, take the vertices $v_1, v_{1+p}, \ldots, v_{1+kp}$ where k is an integer such that $1 + kp \leq m$, and let them each correspond to a single vertex of X. Make the correspondence so that each vertex of C_m corresponds to a different vertex of X. Then, take the remaining vertices and construct the isometric embedding such that two adjacent vertices of C_m correspond to vertices that are no longer than p apart. For all vertices of X, at least one vertex of C_m should correspond to a single vertex of X. Moreover, no more than p vertices of C_m must correspond to the same vertex of X. This isometric embedding of C_m and X gives us a Hausdorff distance less than or equal to $\frac{\lceil \frac{m}{n} \rceil - 1}{2}$. Therefore, $d_{GH}(C_m, X) \geq \frac{\lceil \frac{m}{n} \rceil - 1}{2}$.

Figure 6 shows the steps of how to construct such an isometric embedding of C_8 and X. The dotted red lines represent correspondence.

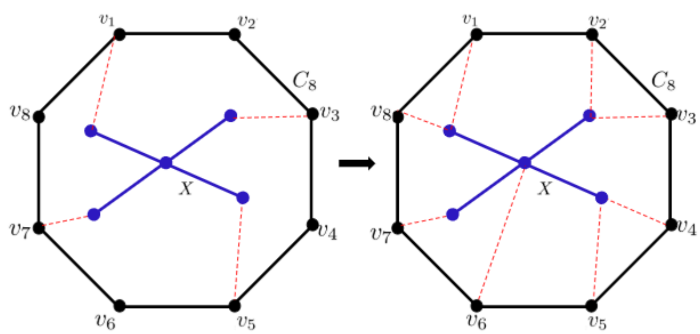

Figure 5: Isometric Embedding of C_8 and X

□

Previously, we had used Lemma 1 to prove Theorem 3 by finding the distance between path graphs that are subgraphs of the cycle graph and the tree. Now we consider the Gromov-Hausdorff distance between two graphs X and Y with the same number of vertices and different number of edges. Here, we use a different approach because the length of the longest path graphs for both X and Y are equal. We prove the lower bound by contradiction. To prove the upper bound, we exhibit a specific isometric embedding of X and Y into a metric space where the vertices of X and Y have a one-to-one correspondence.

Theorem 4. *Consider a graph X with m vertices $v_1, v_2, \ldots, v_m$. Add an edge between v_i, v_j where $1 \leq i, j \leq m$ and $d_X(v_i, v_j) = 2$ to create graph Y. Then, $d_{GH}(X, Y) = \frac{1}{2}$.*

To prove Theorem 4, we will prove the lower bound of d_{GH} using contradiction by Lemma 1. To prove the upper bound, we will show an isometric embedding that exhibits $d_H = \frac{d_X(v_i,v_j)-1}{2}$.

Proof. First, we will prove that $d_{GH}(X,Y) \geq \frac{1}{2}$. In order to prove this, we need to prove that $d_H(X,Y) \geq \frac{1}{2}$ for all possible isometric embeddings of X and Y.

Assume that $d_H(X,Y) < \frac{1}{2}$. Using Lemma 1, we get

$$d_H(X,Y) > \frac{d_X(v_i,v_j) - d_X(w_i,w_j)}{2}.$$

However, $d_X(v_i,v_j) = 2$ and $d_Y(w_i,w_j) = 1$. Then, we get $d_H(X,Y) > \frac{1}{2}$, and this is a contradiction. Therefore, the Hausdorff distance between X and Y is always greater than or equal to $\frac{1}{2}$ for all possible isometric embeddings of X and Y. Thus, $d_{GH}(X,Y) \geq \frac{1}{2}$.

Second, we will prove that $d_{GH}(X,Y) \leq \frac{1}{2}$. Consider an isometric embedding of X and Y where all vertices $v_i, v_2, \ldots, v_m$ of X each respectively correspond to $w_1, w_2, \ldots, w_m$ of Y, and vice versa. In this isometric embedding, there is a two-way one-to-one correspondence between all the vertices of X and Y. Figure 5 shows an example of such an isometric embedding. The red dotted lines represent the two-way correspondence between the vertices. In this isometric embedding, $d_H(X,Y) \leq \frac{1}{2}$.

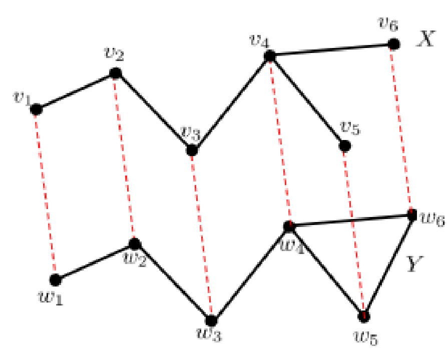

Figure 6: Isometric Embedding of X and Y where $m = 7$

Because $d_{GH}(X,Y) \geq \frac{1}{2}$ and $d_{GH}(X,Y) \leq \frac{1}{2}$, $d_{GH}(X,Y) = \frac{1}{2}$.

□

6 Conclusion and Future Directions

One possibility for future work is a deeper investigation into the Gromov-Hausdorff distance between two graphs that have the same number of vertices and a different number of edges. For example, let us call the original graph G and the new graph with additional edges but the same number of vertices G'. Unlike in Theorem 4, we can add more than one edge to create G'. For these graphs, can we also show that the one-to-one correspondence between the vertices of the two graphs is the only isometric embedding that works.

7 Acknowledgements

I am grateful for the MIT PRIMES Program for the opportunity to work on mathematical research with such an amazing mentor, Ao Sun. I want to thank my mentor, Ao Sun, from MIT for offering great guidance and support throughout the year. I also thank Tanya Khovanova for the guidance she offered for writing quality research papers.

References

[1] D. Burago, Y. Burago, S. Ivanov. A course in metric geometry. Graduate Studies in Mathematics, 33. American Mathematical Society, Providence, RI, 2001.

[2] M. Gromov. Metric structures for Riemannian and non-Riemannian spaces. Modern Birkhuser Classics. Birkhuser Boston, Inc., Boston, MA, 2007.

[3] H. Lee, M. Chung, H. Kang, B. Kim, D. Lee. Computing the Shape of Brain Networks using Graph Filtration and Gromov-Hausdorff Metric. University of Wisconsin, Madison, WI 53705, USA, 2011.

[4] F. Memoli. On the use of Gromov-Hausdorff Distances for Shape Comparison. Eurographics Symposium on Point-Based Graphics, Stanford, CA, USA, 2007.