University of Maribor
Faculty of Natural Sciences and Mathematics

Doctoral dissertation

Distance-based Invariants and Measures in Graphs

December, 2019 Aleksander Kelenc

UNIVERSITY OF MARIBOR

FACULTY OF NATURAL SCIENCES AND MATHEMATICS

DOCTORAL DISSERTATION

DISTANCE-BASED INVARIANTS AND MEASURES IN GRAPHS

DECEMBER, 2019 ALEKSANDER KELENC

MENTOR: ASST. PROF. DR. ANDREJ TARANENKO

CO-MENTOR: ASSOC. PROF. DR. ISMAEL GONZALEZ YERO

UDC: 519.173(043.3)

Univerza v Mariboru

Fakulteta za matematiko in naravoslovje

Doktorska disertacija

Na razdaljah osnovane invariante in mere v grafih

December, 2019

Aleksander Kelenc

Mentor: doc. dr. Andrej Taranenko

Somentor: izr. prof. dr. Ismael Gonzalez Yero

UDK: 519.173(043.3)

Zahvaljujem se mentorju dr. Andreju Taranenku in somentorju dr. Ismaelu Gonzalez Yeru za nasvete, pomoč in usmerjanje pri raziskovalnem delu in nastajanju doktorske disertacije. Posebej bi se rad zahvalil dr. Ismaelu Gonzalez Yeru za topel sprejem v času mojega raziskovalnega obiska v Španiji.

Hvala tudi družini in prijateljem, ki so me podpirali in spodbujali pri študiju. Še posebej hvala partnerki Tini za vso pozitivno energijo, pomoč in ljubezen.

Contents

ABSTRACT

This doctoral dissertation is concerned with aspects on distance related topics in graphs. We study three main topics, namely a recently introduced measure called the Hausdorff distance of graphs and two new graph invariants - the edge metric dimension and the mixed metric dimension of graphs. All three topics are part of the metric graph theory since they are tightly connected with the basic concept of distance between two vertices of a graph.

The Hausdorff distance is a relatively new measure of the similarity of graphs. The notion of the Hausdorff distance considers a special kind of common subgraph of the compared graphs and depends on the structural properties outside of the common subgraph. We study the Hausdorff distance between certain families of graphs that often appear in chemical graph theory. Next to a few results for general graphs, we determine formulae for the distance between paths and cycles. Previously, there was no known efficient algorithm for the problem of determining the Hausdorff distance between two trees, and in this dissertation we present a polynomial-time algorithm for it. The algorithm is recursive and it utilizes the divide and conquer technique. As a subtask it also uses a procedure that is based on the well-known graph algorithm for finding a maximum bipartite matching.

The edge metric dimension is a graph invariant that deals with distinguishing the edges of a graph. Let $G = (V(G), E(G))$ be a connected graph, let $w \in V(G)$ be a vertex, and let $e = uv \in E(G)$ be an edge. The distance between the vertex w and the edge e is given by $d_G(e, w) = \min\{d_G(u, w), d_G(v, w)\}$. A vertex $w \in V(G)$ distinguishes two edges $e_1, e_2 \in E(G)$ if $d_G(w, e_1) \neq d_G(w, e_2)$. A set S of vertices in a connected graph G is an edge metric generator of G if every two distinct edges of G are distinguished by some vertex of S. The smallest cardinality of an edge metric generator of G is called the edge metric dimension and is denoted by $\dim_e(G)$. The concept of the edge metric dimension is new. We study its mathematical properties. We make a comparison between the edge metric dimension and the standard metric dimension of graphs while presenting some realization results concerning the two. We prove that computing the edge metric dimension of connected graphs is NP-hard and give some approximation results. Moreover, we present bounds and closed formulae for the edge metric dimension of several classes of graphs.

The mixed metric dimension is a graph invariant similar to the edge metric dimension that deals with distinguishing the elements (vertices and edges) of a graph. A vertex $w \in V(G)$ distinguishes two elements of a graph $x, y \in E(G) \cup V(G)$ if $d_G(w, x) \neq d_G(w, y)$. A set S of vertices in a connected graph G is a mixed metric generator of G if every two elements $x, y \in E(G) \cup V(G)$ of G, where $x \neq y$, are distinguished by some vertex of S. The smallest cardinality of a mixed metric generator of G is called the mixed metric dimension and is denoted by $\dim_m(G)$. In this dissertation, we consider the structure of mixed metric generators and characterize graphs for which the mixed metric dimension equals the trivial lower and upper bounds. We also give results on the mixed metric dimension of certain families of graphs and present an upper bound with respect to the girth of a graph. Finally, we prove that the problem of determining the mixed metric dimension of a graph is NP-hard in the general case.

Math. Subj. Class. (2010): 05C12, 05C85, 05C90, 05C40, 05C05, 05C70, 05C76.

KEYWORDS: Hausdorff distance, distance between graphs, graph algorithms, trees, graph similarity, edge metric dimension, edge metric generator, mixed metric dimension, metric dimension.

UDC: 519.173(043.3)

Povzetek

V doktorski disertaciji se posvetimo nekaterim temam, ki so povezane z razdaljami v grafih. Osredotočimo se na tri glavne teme, in sicer na pred kratkim vpeljano Hausdorffovo razdaljo med grafi in na dve novi grafovski invarianti - povezavno metrično dimenzijo grafa in mešano metrično dimenzijo grafa. Vse tri obravnavane teme spadajo v metrično teorijo grafov, saj so tesno povezane s konceptom razdalje med dvema vozliščema grafa.

Hausdorffova razdalja med grafi je relativno nova mera podobnosti grafov. Določitev Hausdorffove razdalje med dvema grafoma temelji na posebnem skupnem podgrafu primerjanih grafov, ki ga določimo na podlagi strukturnih lastnosti zunaj samega skupnega podgrafa. V disertaciji obravnavamo Hausdorffovo razdaljo med nekaterimi družinami grafov, ki se pogosto pojavljajo v kemijski teoriji grafov. Poleg rezultatov za splošne grafe izračunamo formule za Hausdorffovo razdaljo med potmi in cikli. Do sedaj ni bil poznan noben učinkovit algoritem za reševanje problema določitve Hausdorffove razdalje med dvema drevesoma, v tej disertaciji pa predstavimo algoritem, ki reši omenjen problem v polinomskem času. Algoritem je rekurziven in uporablja strategijo reševanja problemov "deli in vladaj". Algoritem za reševanje enega od podproblemov uporablja tudi metodo, ki temelji na dobro poznanem algoritmu za iskanje največjega prirejanja v dvodelnem grafu.

Povezavna metrična dimenzija je grafovska invarianta, ki se nanaša na razlikovanje povezav grafa. Naj bo $G = (V(G), E(G))$ povezan graf, naj bo $w \in V(G)$ vozlišče grafa in naj bo $e = uv \in E(G)$ povezava grafa. Razdalja med vozliščem w in povezavo e je določena z $d_G(e, w) = \min\{d_G(u, w), d_G(v, w)\}$. Vozlišče $w \in V(G)$ razlikuje povezavi $e_1, e_2 \in E(G)$, če $d_G(w, e_1) \neq d_G(w, e_2)$. Množica vozlišč S v povezanem grafu G je povezavni metrični generator za G, če za vsaki dve različni

povezavi grafa G velja, da ju razlikuje neko vozlišče iz množice S. Moči najmanjšega povezavnega metričnega generatorja grafa G rečemo povezavna metrična dimenzija in jo označimo z $\dim_{\rm e}(G)$. Povezavna metrična dimenzija je nov koncept. V disertaciji proučujemo njene matematične lastnosti. Skozi predstavitev rezultatov o obstoju grafov z vnaprej določeno povezavno metrično dimenzijo in standardno metrično dimenzijo naredimo primerjavo med obema. Dokažemo, da je izračun povezavne metrične dimenzije povezanih grafov NP-težek problem in podamo nekaj rezultatov o približnih rešitvah. Poleg tega predstavimo še meje in natančne formule za povezavno metrično dimenzijo številnih družin grafov.

Mešana metrična dimenzija grafa je grafovska invarianta, ki je podobna povezavni metrični dimenziji. Nanaša se na razlikovanje elementov grafa (vozlišč in povezav). Vozlišče $w \in V(G)$ razlikuje dva elementa grafa $x, y \in E(G) \cup V(G)$, če $d_G(w, x) \neq d_G(w, y)$. Množica vozlišč S v povezanem grafu G je mešani metrični generator za G, če za vsaka dva elementa $x, y \in E(G) \cup V(G)$ grafa G, kjer $x \neq y$, velja, da ju razlikuje neko vozlišče iz množice S. Moči najmanjšega mešanega metričnega generatorja grafa G rečemo mešana metrična dimenzija in jo označimo z $\dim_{\rm m}(G)$. V disertaciji obravnavamo strukturo mešanih metričnih generatorjev in podamo karakterizacijo grafov, za katere je mešana metrična dimenzija enaka naravnim spodnjim in zgornjim mejam. Podamo tudi rezultate za mešano metrično dimenzijo nekaterih družin grafov in predstavimo zgornjo mejo glede na ožino grafa. Na koncu dokažemo, da je izračun mešane metrične dimenzije povezanih grafov v splošnem NP-težek problem.

Math. Subj. Class. (2010): 05C12, 05C85, 05C90, 05C40, 05C05, 05C70, 05C76.

KLJUČNE BESEDE: Hausdorffova razdalja, razdalja v grafih, algoritmi na grafih, drevesa, podobnost grafov, povezavna metrična dimenzija, povezavni metrični generator, mešana metrična dimenzija, metrična dimenzija.

UDK: 519.173(043.3)

1

Introduction

The distance between two vertices of a given graph is a basic concept that is used in many different invariants and measures in graphs. It is defined as the length of a shortest path between the two vertices. In this doctoral dissertation, we study a recently introduced measure called the Hausdorff distance of graphs and two new graph invariants - the edge metric dimension and the mixed metric dimension of graphs. The basic definitions of all three topics are based on the distance between two vertices in a graph. The motivation to study the topics from metric graph theory is in the applications to other sciences. Many real-life problems can be transformed into a graph theory problem and its solution is applied back to the original problem.

1.1 Hausdorff Distance of Graphs

Comparing the structure of objects is a popular task in several scientific fields, such as chemistry, biology, image processing, robotics, etc. Given two objects, it is very frequently desirable to know if such two objects are identical or similar in some way. For example, in studying the similarity of molecular structures in chemistry, many algorithmic approaches have been developed. The so-called *structure searching* mostly uses a graph isomorphism algorithm to determine whether two molecular compounds are identical; *substructure searching* utilizes the subgraph isomorphism problem and involves determining whether any of the sample structures (usually saved in a database) contain a given structure.

Closely related to the Hausdorff distance of graphs is the problem known in chemistry as *similarity searching*: finding the nearest neighbours of any given molecule of interest within a database - the molecules that are most similar to the given sample - using some measure of inter-molecular similarity [19]. To have a measure of similarity one has to model the compared objects with an appropriate tool. Graphs are often used for this purpose. Determining the distance between two graphs is related to the study of the similarity of molecular structures [50].

A graph can be transformed into another one by a finite sequence of graph edit operations, such as vertex insertion, vertex deletion, vertex substitution, edge insertion, edge deletion and edge substitution. Therefore, the distance between the graphs can be defined by the shortest (or least-cost) edit operation sequence called the *graph edit distance* [24]. The graph edit distance is a general approach of inexact graph matching, and by restricting to some special operations we get special measures. For example, assume that the compared graphs are of the same order and size, the possible operations defined are edge move [5], edge rotation [15] and edge slide [5, 30].

A graph G is said to be a common subgraph of the graphs G_1 and G_2 if it holds that G is a subgraph of G_1 and G is a subgraph of G_2. We say that a common subgraph G of G_1 and G_2 is a maximum common subgraph if a common subgraph H with $|V(H)| > |V(G)|$ does not exist. The problem of determining a maximum common subgraph is also a special case of graph edit distance computation. It was shown [8] that under a particular cost function the graph edit distance computation is equivalent to the maximum common subgraph problem.

In [9], the authors introduced a graph distance metric based on the maximum common subgraph. The metric they define uses only the order of a maximum common subgraph and the order of the graphs compared. A measure of similarity of graphs based on a maximum common subgraph is often used in chemical graph theory to search for molecules that are measured to be close to each other. In [20, 46], the authors describe maximum common subgraph algorithms and their applications to cheminformatics tasks.

The Hausdorff distance of two graphs was introduced in [4]. The Hausdorff distance considers a special kind of common subgraph of the compared graphs and depends on the structural properties outside of the common subgraph. We define and study the Hausdorff distance of graphs in Chapter 3.

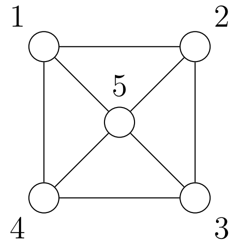
1.2 Edge and Mixed Metric Generators

A *generator* of a metric space is a set S of points in the space with the property that every point of the space is uniquely determined by its distances to the elements of S. Nowadays, several different kinds of metric generators in graphs exist, each one of them studied in theoretical and applied ways, according to their popularity or to their applications. Nevertheless, many other points of view exist which are still not completely taken into account while describing a graph with these metric generators. We introduce and study a new style of metric generators in order to contribute to the knowledge on these distance-related parameters in graphs.
Given a simple and connected graph $G = (V(G), E(G))$, consider the metric $d_G : V(G) \times V(G) \to \mathbb{R}^+$, where $d_G(x, y)$ is the length of a shortest path between x and y. A vertex $v \in V(G)$ is said to *distinguish* two vertices x and y, if $d_G(v, x) \neq d_G(v, y)$. Also, the set $S \subset V(G)$ is said to be a *metric generator* of G if any pair of distinct vertices of G is distinguished by some element of S. A minimum cardinality generator is called a *metric basis*, and its cardinality the *metric dimension* of G, denoted by $\dim(G)$. This is the basic or standard case of metric dimension of graphs and, at this moment, one of the most common in the literature.

This primary concept of metric dimension was introduced by Slater in [47], where the metric generators were called *locating sets* in connection with the problem of uniquely recognizing the location of an intruder in a network. Independently, the concept of metric dimension of a graph was introduced by Harary and Melter in [26], where metric generators were called *resolving sets*. Several applications of this invariant to the navigation of robots in networks are discussed in [32] and applications to chemistry in [13, 14, 31]. Furthermore, this topic has found some applications to problems of pattern recognition and image processing, some of which involve the use of hierarchical data structures [41]. Metric generators are also involved in the theoretical background of certain mind games. In [12], some results are presented that connect metric generators of graphs with the Mastermind game and coin weighing. The metric dimension of infinite graphs was studied in [10], and extremal graphs for metric dimension and diameter were considered in [27]. Moreover, we refer the reader to the work [2], where some historical evolution, nonstandard terminologies and more references on this topic can be found. To determine the metric dimension of a graph is an NP-hard problem in general. The proof is in the paper [32]. The decision problem concerning the metric dimension of a graph was presented in the book [25] as one of the classical NP-complete problems. Because of the problem's difficulty, authors studied metric dimension on several graph families. For example, in [32] authors determined the metric dimension of grids and presented a polynomial algorithm for determining the metric dimension of trees. Later, results for the metric dimension of wheels [7] and fans [11] were also presented. In the paper [12], authors have proven some bounds and some closed formulae for the Cartesian product of several graphs families. It was shown in [18] that determining the metric dimension of a planar graph is an NP-hard problem. They also show that there exists a polynomial-time algorithm to solve the problem on outerplanar graphs.

On the other hand, with respect to the theoretical studies on this topic, different points of view of metric generators have been described in the literature, which have highly contributed to our insight into the mathematical properties of this parameter related to distances in graphs. Several authors have introduced other variations of metric generators. For instance, resolving dominating sets [6], independent resolving sets [16], local metric sets [43], strong resolving sets [42], simultaneous metric generators [45], k-metric generators [23, 52], resolving partitions [17], strong resolving partitions [51], k-antiresolving sets [48], etc. have been presented and studied.

A metric basis S of a connected graph G uniquely identifies all the vertices of G by means of distance vectors. One could think that the edges of the graph are in some way also identified by S with respect to distances to S. However, this is quite far from the truth. For instance, Figure 1.1 shows an example of a graph in which no metric basis uniquely recognizes all the edges of the graph. We observe that graph G in Figure 1.1 satisfies that $\dim(G) = 2$ and the set of all metric bases is the following one: $\{\{1,2\}$, $\{1,4\}$, $\{2,3\}, \{3,4\}\}$. But, for each one of these metric bases, there exists at least a pair of edges that is not distinguished by the corresponding metric basis.



Basis	Edges
$\{1, 2\}$	$14, 15$
$\{1, 4\}$	$12, 15$
$\{2, 3\}$	$12, 25$
$\{3, 4\}$	$15, 25$

Figure 1.1: A graph in which any metric basis does not recognize all edges, and a table with all metric bases and two edges that are not recognized by the corresponding metric basis.

In this sense, a natural question is: Are there some sets of vertices that uniquely identify all the edges of a graph? The answer is positive, and one of our goals in this work is to study such sets. We present a new variant of metric generators of graphs, which distinguishes the edges of a graph. Given a connected graph $G = (V(G), E(G))$ with at least two vertices, a vertex $v \in V(G)$ and an edge $e = uw \in E(G)$, the distance between the vertex v and the edge e is defined as $d_G(e, v) = \min\{d_G(u, v), d_G(w, v)\}$. A vertex $w \in V(G)$ *distinguishes* two edges $e_1, e_2 \in E(G)$ if $d_G(w, e_1) \neq d_G(w, e_2)$. A non-empty set S of vertices in a connected graph G is an *edge metric generator* of G if any two distinct edges of G are distinguished by some vertex of S. The smallest cardinality of an edge metric generator of G is called the *edge metric dimension* and is denoted by $\dim_e(G)$. An *edge metric basis* of G is an edge metric generator of G of cardinality $\dim_e(G)$.

Another useful approach to edge metric generators could be the following one. Given an ordered set of vertices $S = \{s_1, s_2, \ldots, s_d\}$ of a connected graph G, for any edge e in G, we refer to the d-vector $r(e|S) = (d_G(e, s_1), d_G(e, s_2), \ldots, d_G(e, s_d))$ as the *edge metric representation* of e with respect to S. In this sense, S is an edge metric generator of G if and only if for every pair of different edges e_1, e_2 of G, it follows that $r(e_1|S) \neq r(e_2|S)$.

Considering the definition of an edge metric generator, which uniquely determines every edge of the graph, one could think that any edge metric generator S is also a standard metric generator, *i.e.* every vertex of the graph is identified by S. Again, this is far from reality, despite the fact that there are several families in which such a fact occurs. We just have to take, for instance, the hypercube graph Q_4, for which it is known from [13] that $\dim(Q_4) = 4$, and we have computed in [36] that $\dim_{\mathrm{e}}(Q_4) = 3$ (the computation was done by a computer program using an exhaustive search algorithm).

In order to show that computing the metric dimension of the line graph of a bipartite graph is NP-hard, the authors of [21] introduce another edge metric dimension definition related to the line graphs. Their edge metric dimension of a graph G is defined as the metric dimension of the line graph $L(G)$. This definition of the edge metric dimension is clearly different from our definition of the edge metric dimension and these are two completely different things. To avoid confusion about the name, the authors of [40] rename the edge metric dimension from [21] to the *edge version of metric dimension*.

Metric dimension deals with distinguishing the pairs of distinct vertices and edge metric dimension deals with distinguishing the pairs of distinct edges. How about creating a mixed version of these two parameters described above. That is, given a connected graph G, we wish to uniquely identify the elements (edges and vertices) of G by means of vector distances to a fixed set of vertices of G. A vertex v of a connected graph G *distinguishes* two distinct elements (vertices or edges) $x, y \in E(G) \cup V(G)$ of G if $d_G(x, v) \neq d_G(y, v)$. A set S of vertices of G is a *mixed metric generator* if any two elements $x, y \in E(G) \cup V(G)$ of G, where $x \neq y$, are distinguished by some vertex of S. The smallest cardinality of a mixed metric generator of G is called the *mixed metric dimension* and is denoted by $\dim_{\mathrm{m}}(G)$. A *mixed metric basis* of G is a mixed metric generator of G of cardinality $\dim_{\mathrm{m}}(G)$.

We proceed as follows. First, we describe some basic concepts of graph theory that are neccesarry for the remaining part of the dissertation. In Chapter 3, we introduce the Hausdorff distance between grahps and we present original results from [33, 35]. Chapter 4 deals with the edge metric generators and my results from [36] are presented. In Chapter 5 we study the mixed metric dimension and present all the main results from [34]. We conclude the dissertation with some open problems.

2

THE BASIC CONCEPTS

Let $G = (V(G), E(G))$ be a graph with the vertex set $V(G)$ and the edge set $E(G)$, where an edge is an unordered pair of vertices $\{u, v\}$. The short notation uv is used for an edge $\{u, v\}$. Vertices u and v are *endpoints* of the edge uv. A vertex u is *adjacent* to a vertex v if $uv \in E(G)$. A vertex u is *incident* to an edge e if it is an endpoint of the edge e.

Let $G = (V(G), E(G))$ and $H = (V(H), E(H))$ be arbitrary graphs. Graph H is a subgraph of G if $V(H) \subseteq V(G)$ and $E(H) \subseteq E(G)$. Graph H is called a proper subgraph of G if $V(H) \subset V(G)$.

All graphs considered in the dissertation are simple graphs, i.e. the are no multiple edges and no loops ($uu \notin E(G)$, for any $u \in V(G)$).

Let G be a graph and let $S \subseteq V(G)$. By $\langle S \rangle$ we denote the subgraph of G induced by the set S, i.e. for all $u, v \in S$, $uv \in E(\langle S \rangle)$ if and only if $uv \in E(G)$.

Graphs G_1 and G_2 are *isomorphic*, denoted by $G_1 \cong G_2$, if there is a bijective correspondence between their vertex sets, which preserves the adjacency and non-adjacency of the vertices.

A *path* P from a vertex u to a vertex v in a graph G is a sequence $u = v_0 v_1 v_2 \dots v_{k-1} v_k = v$ of pairwise different vertices of G, where $v_i v_{i+1}$ is an edge of G, for each $i \in \{0, \dots, k-1\}$. The vertices u and v are called the *endpoints* of the path. The *length* of a path P, denoted by $\ell(P)$, is the number of edges in P. If we add the edge uv to the path, then we get a *cycle*.

The *girth* $g(G)$ of G is the order of the smallest cycle in G.

If every two different vertices of a graph G are adjacent then we call the graph G a *complete graph*. The notation for a complete graph on n vertices is K_n.

The *distance* between vertices u and v is the length of a shortest path between u and v in G and is denoted by $d_G(u,v)$. The distance between vertex u and subset of vertices $S \subseteq V(G)$ is defined as $d_G(u,S) = \min_{v \in S}\{d_G(u,v)\}$.

A graph G is *connected* if for each pair of vertices $u, v \in V(G)$ there is a path in G from u to v.

A connected subgraph H of a graph G is *convex* in G if for any pair of vertices $u, v \in V(H)$, any shortest path P from u to v in graph G lies entirely in H ($P \subseteq H$).

A graph $T = (V(T), E(T))$ is a *tree* if it is connected and has no cycles. A tree $T = (V(T), E(T))$ is *rooted* if there is a distinguished vertex $r \in V(T)$ called the *root* of the tree. Note, there is a unique path from the root to any other vertex $v \in V(T)$. We can draw a rooted tree in such a way that the root is at the top and the other vertices can be partitioned in the levels according to their distance from the root of the tree. The *depth* of vertex $v \in V(T)$, denoted by $depth[v]$, is the length of the path from the root vertex to vertex v. The depth of the tree T is the maximum among all of the depths of all the vertices. Vertex $v \in V(T)$ is called an *ancestor* of vertex $u \in V(T)$ if vertex v lies on the unique path from u to the root and $u \neq v$. Vertex $v \in V(T)$ is called a *descendant* of vertex $u \in V(T)$ if vertex u lies on the unique path from v to the root and $u \neq v$. The set of all ancestors (descendants) of vertex v is denoted by $ancestors[v]$ ($descendants[v]$), respectively. Vertex $v \in V(T)$ is called the *parent* of vertex $u \in V(T)$, denoted by $parent[u]$, if $vu \in E(T)$ and v is an ancestor of u. The vertex u is then called a *child* of the vertex v. The *children* of vertex v is the set $children[v] = \{u \in V(T) \mid u \text{ is a child of } v\}$. A vertex with no children is called a *leaf*. Non-root vertices $v, u \in V(T)$ are *siblings* if $parent[v] = parent[u]$. The *height* of a vertex $v \in V(T)$, denoted by $height[v]$, is the length of a longest path among all paths from the vertex v to any other vertex in the vertex set $\{v\} \cup descendants[v]$.

Example 2.1. *In Figure 2.1 there is a rooted tree T with the root vertex v_{10}. Tree T is drawn twice. On the left side, T is drawn with regard to the depth of the vertices, and on the right side, T is drawn with regard to the height of the vertices.*

Let G be a graph and v a vertex of G. The *eccentricity* of the vertex v, denoted

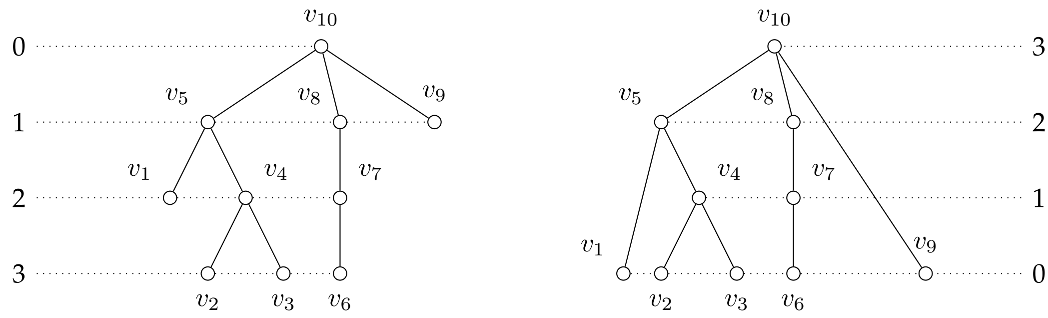


Figure 2.1: A rooted tree T drawn with regard to the depth (left hand-side) and to the height (right hand-side) of vertices.

$\mathrm{e}(v)$ is the maximum distance from v to any vertex of $V(G)$. That is, $\mathrm{e}(v) = \max\{d_G(v,u) \mid u \in V(G)\}$. The *radius* of the graph G, denoted $\mathrm{rad}(G)$, is the minimum eccentricity among the vertices of G, i.e. $\mathrm{rad}(G) = \min\{\mathrm{e}(v) \mid v \in V(G)\}$. The *diameter* of G, denoted $\mathrm{diam}(G)$, is the maximum eccentricity among the vertices of G, i.e. $\mathrm{diam}(G) = \max\{\mathrm{e}(v) \mid v \in V(G)\}$. The *center* of G is the set of vertices with the minimum eccentricity, i.e. $\mathrm{center}(G) = \{v \in V(G) \mid \mathrm{e}(v) = \mathrm{rad}(G)\}$. A vertex $v \in \mathrm{center}(G)$ is called a *central vertex* of G. For an arbitrary graph G it holds that $\mathrm{rad}(G) \leq \mathrm{diam}(G) \leq 2 \cdot \mathrm{rad}(G)$.

A graph $G = (V(G), E(G))$ is *bipartite* if the set of vertices $V(G)$ can be partitioned into two sets, A and B, in such a way that any edge from $E(G)$ has one endpoint in set A and the other in set B. If all possible edges are between partition sets A and B, then graph G is called *a complete bipartite* graph. We denote a complete bipartite graph with $K_{r,t}$, where $r = |A|$ and $t = |B|$.

A *matching* $M \subseteq E(G)$ is a collection of edges in which every vertex of $V(G)$ is incident to at most one edge of M. A vertex is *matched* if it is an endpoint of an edge from the set M. A *maximum matching* is a matching that contains the largest possible number of edges. A matching is called *perfect* if every vertex of a graph G is matched. A maximum matching in bipartite graph $G = (V(G), E(G))$ is called *a maximum bipartite matching*. The problem of finding a maximum bipartite matching can be solved in polynomial time. The Hopcroft-Karp algorithm [28] finds a maximum bipartite matching in $\mathcal{O}(\sqrt{|V(G)|}|E(G)|)$ time.

A line graph of graph G is defined as the graph $L(G)$, where the vertex set $V(L(G)) = E(G)$ and the edge set $E(L(G)) = \{e_ie_j \mid e_i, e_j \in E(G) \wedge$

e_i has a common endpoint with $e_j\}$.

The *open neighbourhood* $N(v)$ of vertex v in graph G is given by all the vertices that are adjacent to v and the *closed neighbourhood* of v is $N[v] = N(v) \cup \{v\}$. The vertex v is called a *simplicial vertex* if $N(v)$ induces a complete graph. Two vertices u, v of G are called *false twins* if they are have the same open neighbourhoods, *i.e.* $N(u) = N(v)$. Similarly, the vertices u, v are called *true twins* if $N[u] = N[v]$. A vertex v is a *true twin* or a *false twin* in G, if such a $u \neq v$ exists that u and v are true twins or false twins, respectively.

The *Cartesian product* of two graphs G and H is the graph $G \Box H$, in which $V(G \Box H) = \{(a,b) \mid a \in V(G),\ b \in V(H)\}$ and two vertices (a,b) and (c,d) are adjacent in $G \Box H$ if and only if, either

- $a = c$ and $bd \in E(H)$, or
- $b = d$ and $ac \in E(G)$.

Let $h \in V(H)$ be an arbitrary vertex of graph H. In the Cartesian product, a set $V(G) \times \{h\}$ is a G-layer. Similarly, $\{g\} \times V(H)$, $g \in V(G)$ is an H-layer. When referring to a specific G or H layer, we denote them by G^h or ${}^gH$, respectively. Obviously, the subgraph induced by a G-layer or by an H-layer is isomorphic to G or H, respectively.

The *join graph* $G \vee H$ of graphs G and H is the graph obtained from G and H by adding all the possible edges between vertices of G and vertices of H.

3

Hausdorff Distance of Graphs

In this chapter, we present results on the Hausdorff distance of graphs. It is a measure of the similarity of graphs that depends on the structural properties outside of the common subgraph of the compared graphs. First, we define the Hausdorff distance of graphs and show results for general graphs. We study the Hausdorff distance between certain families of graphs from chemical graph theory, namely path, cycles and trees. We determine formulae for the distance between paths and cycles. We present a polynomial-time algorithm for the problem of determining the Hausdorff distance between two trees.

The Hausdorff distance of two graphs was introduced in 2014 by Banič and Taranenko. To define it we need the following definitions from [4].

Definition 3.1. *[4] Let G be an arbitrary graph. The Hausdorff graph of graph G, denoted by 2^G, has for the vertex set $V(2^G)$ the set of all non-empty subgraphs of G. The adjacency of vertices in 2^G is defined as follows: for all $H_1, H_2 \in V(2^G)$, $H_1 \neq H_2$, it holds that $H_1H_2 \in E(2^G)$ if and only if*

1. *for each $v \in V(H_1)$ there exists $v' \in V(H_2)$ such that $d_G(v, v') \leq 1$ and*
2. *for each $v' \in V(H_2)$ there exists $v \in V(H_1)$ such that $d_G(v', v) \leq 1$.*

The Hausdorff metric h_G between two subgraphs of graph G is described in the following definition. It will tell us how much two subgraphs of G coincide.

Definition 3.2. *[4] Let G be an arbitrary graph. The* distance between two subgraphs H_1 and H_2 of G, *denoted by $h_G(H_1, H_2)$, is the distance between their corresponding vertices in 2^G. In other words,*

$$h_G(H_1, H_2) := d_{2^G}(H_1, H_2).$$

We call h_G the Hausdorff metric on 2^G.

Note that for two different isomorphic subgraphs H_1 and H_2 of a graph G, the value $h_G(H_1, H_2)$ may be arbitrarily large. Also, the following corollary is proven in [4].

Corollary 3.3. *[4] If G is connected, then h_G is a metric on $V(2^G)$.*

in order to define the Hausdorff distance on the class of all connected simple graphs as a measure of the similarity of two such graphs we have to introduce amalgams (cf. [3, 37]).

Definition 3.4. *Let H_1 be a (convex) subgraph of G_1 and H_2 a (convex) subgraph of G_2. If H_1 and H_2 are isomorphic graphs, then a (convex) amalgam of G_1 and G_2 is any graph A obtained from G_1 and G_2 by identifying their subgraphs H_1 and H_2. We call the isomorphic copies of G_1 and G_2 in A the covers of the amalgam A and denote them with G_1^A and G_2^A, respectively. See Figure 3.1 for reference.*

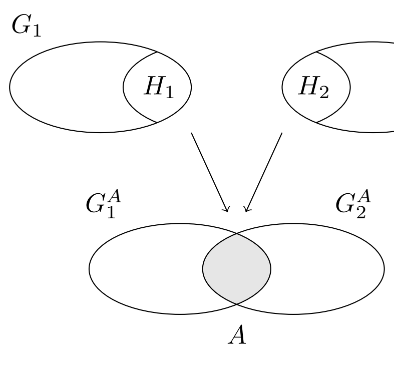


Figure 3.1: An amalgam A of G_1 and G_2.

Denote by $\mathcal{A}(G_1, G_2)$ and $\mathcal{X}(G_1, G_2)$ the sets of all amalgams and all convex amalgams of the graphs G_1 and G_2, respectively.

Remark 3.5. *Let A be an amalgam of G_1 and G_2, obtained from G_1 and G_2 by identifying their subgraphs H_1 and H_2. Then $G_1^A \cap G_2^A = H_1^A = H_2^A$ is isomorphic to H_1 and H_2.*

Let $\mathcal{G}$ be the class of all simple connected graphs.

Theorem 3.6. *[4, Theorem 4.10] Let $G_1, G_2 \in \mathcal{G}$. Let d be a non-negative integer and A an amalgam of G_1 and G_2. Then $h_A(G_1^A, G_2^A) = d$ if and only if*

(i) for each $u \in V(G_1^A)$ there is a vertex $v \in V(G_2^A)$ such that $d_A(u, v) \leq d$,

(ii) for each $u \in V(G_2^A)$ there is a vertex $v \in V(G_1^A)$ such that $d_A(u, v) \leq d$, and

(iii) there is $u \in V(G_1^A)$ such that for each vertex $v \in V(G_1^A \cap G_2^A)$ the distance $d_A(u, v) \geq d$ or

there is $u \in V(G_2^A)$ such that for each vertex $v \in V(G_1^A \cap G_2^A)$ the distance $d_A(u, v) \geq d$.

From Theorem 3.6 we get the following Corollary.

Corollary 3.7. *Let $G_1, G_2 \in \mathcal{G}$. Let A be an amalgam of G_1 and G_2. Then*

$$h_A(G_1^A, G_2^A) = \max_{u \in V(A)} \left\{ d_A(u, G_1^A \cap G_2^A) \right\}.$$

Proof. Let $d := \max_{u \in V(A)} \{ d_A(u, G_1^A \cap G_2^A\}$ and $u \in V(G_i^A)$, for some $i \in \{1, 2\}$, such that $d_A(u, G_1^A \cap G_2^A) = d$. Thus, for every vertex $v \in V(G_1^A \cap G_2^A)$ it holds that $d_A(u, v) \geq d$. Therefore, the condition (iii) of Theorem 3.6 holds true.

Choose a vertex $u_1 \in V(G_1^A)$. Let $v_1 \in V(G_1^A \cap G_2^A)$ be such that $d_A(u_1, v_1) = d_A(u_1, G_1^A \cap G_2^A)$. Then $d_A(u_1, v_1) \leq \max_{u \in V(G_1^A)} \{ d_A(u, G_1^A \cap G_2^A\} \leq d$. It follows that the condition (i) of Theorem 3.6 holds true.

Following the same line of thought one can prove that the condition (ii) of Theorem 3.6 is also fulfilled.

Since all of the conditions of Theorem 3.6 hold true, the assertion follows immediately. $\square$

Given $G_1, G_2 \in \mathcal{G}$ and an amalgam A of G_1 and G_2, Corollary 3.7 states that to determine $h_A(G_1^A, G_2^A)$ it suffices to find a vertex $v \in V(A)$ with the maximum distance to $G_1^A \cap G_2^A$, since $h_A(G_1^A, G_2^A) = d_A(v, G_1^A \cap G_2^A)$. This idea is often used in our proofs.

Finally, the Hausdorff distance $\mathcal{H} : \mathcal{G} \times \mathcal{G} \to \mathbb{R}$ on $\mathcal{G}$ can be defined as follows:

Definition 3.8. *[4] For any graphs $G_1, G_2 \in \mathcal{G}$, we define*

$$\mathcal{H}(G_1, G_2) = \begin{cases} \min\left\{h_A(G_1^A, G_2^A) \mid A \in \mathcal{X}(G_1, G_2)\right\}, & \textit{if } G_1 \ncong G_2 \\ 0, & \textit{if } G_1 \cong G2 \end{cases}.$$

We call $\mathcal{H}$ the Hausdorff distance *on $\mathcal{G}$.*

Note, Definition 3.8 is equivalent to the definition of the Hausdorff distance in [4, Definition 4.18]. Moreover, it is proven in [4] that $\mathcal{H}$ is a metric on the class of all simple connected pairwise non-isomorphic graphs. A convex amalgam A of two simple connected graphs G_1 and G_2, for which $h_A(G_1^A, G_2^A) = \mathcal{H}(G_1, G_2)$ is called an *optimal amalgam*.

To determine the Hausdorff distance between graphs G_1 and G_2 from $\mathcal{G}$ one has to find an optimal amalgam. Having a convex common subgraph of G_1 and G_2 makes it possible to construct an amalgam of graphs G_1 and G_2. Therefore, the task is to find a convex common subgraph of G_1 and G_2 such that the distance between the covers G_1^A and G_2^A of the corresponding amalgam A is minimized.

As noted in [4], for fixed isomorphic subgraphs H_1 and H_2 of G_1 and G_2, respectively, there may be many isomorphisms from H_1 onto H_2. Therefore, there may be more than just one amalgam A of G_1 and G_2, which is obtained by identifying H_1 and H_2 (see Example 3.9).

Example 3.9. *Let G_1 and G_2 be the graphs depicted in Figure 3.2, and H_1 and H_2 their subgraphs, respectively, both isomorphic to P_2. Let f_1 and f_2 be two isomorphisms from H_1 onto H_2. In Figure 3.2, they are depicted with dotted and dashed arrows, respectively. Next, let A_i be the amalgam of G_1 and G_2 obtained by identifying H_1 and H_2 according to the isomorphism f_i, $i \in \{1, 2\}$. Obviously, A_1 and A_2 are not isomorphic, although they were both obtained by identifying the same subgraphs.*

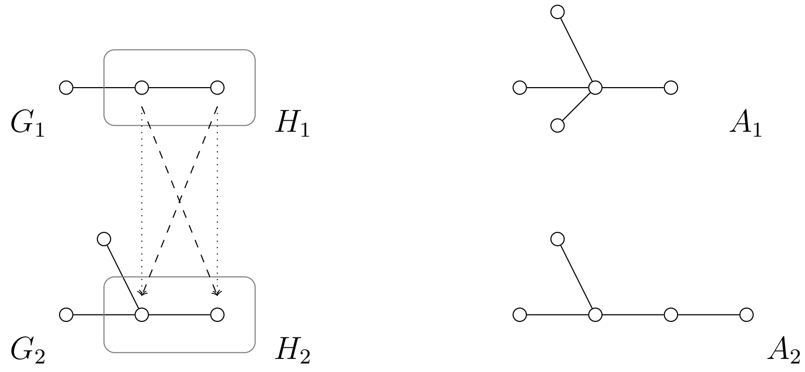


Figure 3.2: The amalgams A_1 and A_2 from Example 3.9.

In the next theorem, we prove that distance between the covers of a convex amalgam (therefore also the Hausdorff distance between two graphs) is not dependant on the choice of the isomorphism between the subgraphs.

Theorem 3.10. *Let $G_1, G_2 \in \mathcal{G}$ and let H_1 and H_2 be fixed isomorphic convex subgraphs of G_1 and G_2, respectively. Also, let f_1 and f_2 be any two isomorphisms between H_1 and H_2, and A_1 and A_2, the two convex amalgams of G_1 and G_2 obtained by identifying H_1 and H_2 with respect to isomorphisms f_1 and f_2, respectively. Then $h_{A_1}(G_1^{A_1}, G_2^{A_1}) = h_{A_2}(G_1^{A_2}, G_2^{A_2})$.*

Proof. Let $h_k = h_{A_k}(G_1^{A_k}, G_2^{A_k})$, for each $k \in \{1,2\}$. Towards contradiction, suppose $h_{A_1}(G_1^{A_1}, G_2^{A_1}) < h_{A_2}(G_1^{A_2}, G_2^{A_2})$. Then, according to Corollary 3.7 a vertex $u \in G_i^{A_1}$ exists for some $i \in \{1,2\}$, with $d_{A_1}(u, G_1^{A_1} \cap G_2^{A_1}) = h_1$. Let $v \in V(G_1^{A_1} \cap G_2^{A_1})$ be such that $d_{A_1}(u,v) = h_1$. Similarly, a vertex $x \in G_i^{A_2}$ exists for some $i \in \{1,2\}$, with $d_{A_2}(x, G_1^{A_2} \cap G_2^{A_2}) = h_2$. Let $y \in V(G_1^{A_2} \cap G_2^{A_2})$ be such that $d_{A_2}(x,y) = h_2$. Use x' to denote the vertex in a cover of A_1 corresponding to x, and $y' \in V(G_1^{A_1} \cap G_2^{A_1})$ to denote the vertex corresponding to y.

Obviously, $d_{A_1}(x', G_1^{A_1} \cap G_2^{A_1}) \leq d_{A_1}(x', y')$. To show the other inequality, suppose that $d_{A_1}(x', G_1^{A_1} \cap G_2^{A_1}) < d_{A_1}(x', y')$. A vertex $z' \in V(G_1^{A_1} \cap G_2^{A_1})$ exists such that $d_{A_1}(x', z') = d_{A_1}(x', G_1^{A_1} \cap G_2^{A_1})$. Denote as $z \in V(G_1^{A_2} \cap G_2^{A_2})$ the vertex in a cover of A_2 corresponding to z'. It follows that $d_{A_2}(x,z) < d_{A_2}(x,y) = d_{A_2}(x, G_1^{A_2} \cap G_2^{A_2})$, a contradiction. Therefore, $d_{A_1}(x', G_1^{A_1} \cap G_2^{A_1}) \geq d_{A_1}(x', y')$ and taking both inequalities we get $d_{A_1}(x', G_1^{A_1} \cap G_2^{A_1}) = d_{A_1}(x', y')$. Moreover, $d_{A_1}(x', y') = d_{A_2}(x,y) = h_2$. For all $w \in V(A_1)$ it holds that $d_{A_1}(u,v) \geq d_{A_1}(w, G_1^{A_1} \cap G_2^{A_1})$ by Corollary 3.7.

Therefore $h_1 = d_{A_1}(u,v) \geq d_{A_1}(x', G_1^{A_1} \cap G_2^{A_1}) = d_{A_1}(x', y') = h_2$, so $h_1 \geq h_2$, a contradiction to our assumption.

Similarly, one can disprove the case that $h_{A_2}(G_1^{A_2}, G_2^{A_2}) < h_{A_1}(G_1^{A_1}, G_2^{A_1})$. Therefore, the assertion follows. □

Let G be a graph and H its convex subgraph. The distance between H and G is defined as $\max_{v\in V(G)}\{d_G(v,H)\}$. Note that G can be viewed as an amalgam of G and H', where H' is isomorphic to H, and the amalgam of G and H' is obtained by identifying H and H'. Therefore, by Corollary 3.7, $\max_{v\in V(G)}\{d_G(v,H)\} = h_G(G^G, H^G)$.

Proposition 3.11. *Let $G_1, G_2 \in \mathcal{G}$. Let H_1 and H_2 be two isomorphic convex subgraphs of G_1 with $d_1 \leq d_2$, where d_1 and d_2 are the distances between H_1 and G_1, and H_2 and G_1, respectively. Let H_3 be a convex subgraph of G_2, isomorphic to H_1 (and H_2). Let A_1 be a convex amalgam of G_1 and G_2 obtained by identifying H_1 and H_3, and A_2 be a convex amalgam of G_1 and G_2 obtained by identifying H_2 and H_3. Then the inequality $h_{A_1}(G_1^{A_1}, G_2^{A_1}) \leq h_{A_2}(G_1^{A_2}, G_2^{A_2})$ holds true.*

Proof. Let d_3 be the distance between H_3 and G_2. From Corollary 3.7 it follows that $h_{A_1}(G_1^{A_1}, G_2^{A_1}) = \max\{d_1, d_3\}$ and $h_{A_2}(G_1^{A_2}, G_2^{A_2}) = \max\{d_2, d_3\}$. Since $d_1 \leq d_2$, it follows that $\max\{d_1, d_3\} \leq \max\{d_2, d_3\}$, which implies that $h_{A_1}(G_1^{A_1}, G_2^{A_1}) \leq h_{A_2}(G_1^{A_2}, G_2^{A_2})$. □

For two arbitrary simple connected graphs, the upper bound for the Hausdorff distance can be expressed using the radius of the graphs.

Theorem 3.12. *Let G_1 and G_2 be two arbitrary simple, connected graphs. Then*

$$\mathcal{H}(G_1, G_2) \leq \max\left\{\mathrm{rad}(G_1), \mathrm{rad}(G_2)\right\}.$$

Proof. Let c_1 be a central vertex of G_1 and c_2 be a central vertex of G_2. Let A be an amalgam, which is created by identifying c_1 and c_2. Since there is exactly one vertex in $G_1^A \cap G_2^A$, A is a convex amalgam. In G_1 it holds that for each $v \in V(G_1)$ the distance $d_{G_1}(v, c_1) \leq \mathrm{rad}(G_1)$. Similarly, in G_2 it holds that for each $v \in V(G_2)$ the distance $d_{G_2}(v, c_2) \leq \mathrm{rad}(G_2)$. Since A is a convex amalgam, the same holds for the corresponding vertices of G_1^A and G_2^A in A. Using Corollary 3.7, it follows that $h_A(G_1^A, G_2^A) = \max\left\{\mathrm{rad}(G_1), \mathrm{rad}(G_2)\right\}$ and $\mathcal{H}(T_1, T_2) \leq \max\left\{\mathrm{rad}(G_1), \mathrm{rad}(G_2)\right\}$. □

Note, this bound is sharp if one of the graphs is trivial (a one vertex graph).

3.1 Results on Some Simple Families of Graphs

In this section, we present some results on the Hausdorff distance between two graphs of some simple families of graphs that often appear in chemical graph theory.

First, consider the following Remarks which can be easily verified.

Remark 3.13. *We will often use the following implication. If a and b are two arbitrary positive integers with $a < b$, then $2a < 2b - 1$. Clearly, if $b \geq a + 1$, then $2b \geq 2a + 2 > 2a + 1$.*

Remark 3.14. *For an arbitrary positive integer m the following equality holds:*

$$\left\lceil \frac{\left\lfloor \frac{m}{2} \right\rfloor}{2} \right\rceil = \left\lceil \frac{m-1}{4} \right\rceil .$$

Note that for a path every connected subgraph is also a convex subgraph. Now we give formulae for the Hausdorff distance between some simple families of graphs. In all cases we construct a convex amalgam and thus obtain an upper bound. Then we show there can be no amalgam that would give a smaller result.

Proposition 3.15. *If P_n and P_m are two paths on n and m vertices, respectively, with $n \geq m \geq 1$, then $\mathcal{H}(P_n, P_m) = \left\lceil \frac{n-m}{2} \right\rceil$.*

Proof. Denote the vertices of P_n with $u_1, \ldots, u_n$, where $u_i u_{i+1} \in E(P_n)$, for each $i \in \{1, \ldots, n-1\}$, and the vertices of P_m with $v_1, \ldots, v_m$, where $v_i v_{i+1} \in E(P_m)$, for each $i \in \{1, \ldots, m-1\}$.

Let A be an amalgam that is created by identifying pairs of vertices $u_{\lceil \frac{n-m}{2} \rceil + i}$ and v_i for each $1 \leq i \leq m$. A is clearly a convex amalgam. Using Corollary 3.7 we immediately deduce that $h_A(P_n^A, P_m^A) = \left\lceil \frac{n-m}{2} \right\rceil$ and therefore $\mathcal{H}(P_n, P_m) \leq \left\lceil \frac{n-m}{2} \right\rceil$.

Suppose now, that there exists an amalgam $A' \in \mathcal{X}(P_n, P_m)$ such that $k := h_{A'}(P_n^{A'}, P_m^{A'}) < \left\lceil \frac{n-m}{2} \right\rceil$. Due to Corollary 3.7, for each $w \in V(A')$ it holds that

$k \geq d_{A'}(w, P_n^{A'} \cap P_m^{A'})$. The graph $P_n^{A'} \cap P_m^{A'}$ is isomorphic to a path with at most m vertices. Thus, for every path P in A' it follows that the length $\ell(P) \leq m-1+2k$. It holds that $\ell(P) \leq m-1+2k < m-1+2\left\lceil \frac{n-m}{2} \right\rceil -1 \leq m-1+2\frac{n-m+1}{2}-1 = n-1$. So for every path P in A' it holds that $\ell(P) < n-1$. But $P_n^{A'} \subseteq A'$ and $\ell(P_n^{A'}) = n-1$; this is a contradiction with the assumption that such an amalgam A' exists. $\square$

If C_n is a cycle on n vertices, with $n \geq 3$, then the largest convex subgraph of C_n is a path on $\left\lceil \frac{n}{2} \right\rceil$ vertices.

Proposition 3.16. *If P_n and C_m are a path and a cycle on n and m vertices, respectively, with $n \geq 1$ and $m \geq 3$, then*

$$\mathcal{H}(P_n, C_m) = \begin{cases} \left\lceil \frac{m-n}{2} \right\rceil, & \text{if } n \leq \frac{m}{2} \\ \left\lceil \frac{m-1}{4} \right\rceil, & \text{if } \frac{m}{2} < n \leq m \\ \left\lceil \frac{n-\left\lceil \frac{m}{2} \right\rceil}{2} \right\rceil, & \text{if } n > m. \end{cases}$$

Proof. Denote vertices of P_n with $u_1, \dots, u_n$, where $u_i u_{i+1} \in E(P_n)$, for each $i \in \{1, \dots, n-1\}$, and vertices of C_m with $v_0, v_1, v_2, \dots, v_{m-1}$, where $v_i v_{i+1} \in E(C_m)$, for each $i \in \{0, \dots, m-1\}$. All indices in C_m are computed modulo m.

Let $n \leq \frac{m}{2}$. Let A be an amalgam, which is created by identifying pairs of vertices u_i and v_i, for each $1 \leq i \leq n$. Since every subgraph of C_m isomorphic to a path on n vertices is a convex subgraph of C_m, A is a convex amalgam. Clearly, $\max_{u \in V(A)}\{d_A(u, P_n^A \cap C_m^A)\} = \left\lceil \frac{m-n}{2} \right\rceil$. Using Corollary 3.7, it follows that $h_A(P_n^A, C_m^A) = \left\lceil \frac{m-n}{2} \right\rceil$ and $\mathcal{H}(P_n, C_m) \leq \left\lceil \frac{m-n}{2} \right\rceil$.

Suppose a convex amalgam A' with $h_{A'}(P_n^{A'}, C_m^{A'}) < \left\lceil \frac{m-n}{2} \right\rceil$ exists. Define $h := \left\lceil \frac{m-n}{2} \right\rceil$. Due to convexity, $P_n^{A'} \cap C_m^{A'}$ is isomorphic to a path on k vertices, $1 \leq k \leq n$. Say the vertices in $P_n^{A'} \cap C_m^{A'}$ are $v_i^{A'}, v_{i+1}^{A'}, \dots, v_{i+k-1}^{A'}$, with an edge between two consecutive vertices. We now consider the vertex $v_{i-h}^{A'}$. Clearly, $d_{A'}(v_i^{A'}, v_{i-h}^{A'}) = h =$

$\left\lceil \frac{m-n}{2} \right\rceil$. On the other hand,

$$
\begin{aligned}
d_{A'}(v_{i+k-1}^{A'}, v_{i-h}^{A'}) &= \\
\ell(C_m) - h - (k-1) &= \\
m - h - k + 1 &\geq \\
m - h - n + 1 &= \\
2\frac{m-n+1}{2} - h &\geq \\
2\left\lceil \frac{m-n}{2} \right\rceil - h &= h.
\end{aligned}
$$

It follows that $d_{A'}(v_{i-h}^{A'}, P_n^{A'} \cap C_m^{A'}) = \left\lceil \frac{m-n}{2} \right\rceil > h_{A'}(P_n^{A'}, C_m^{A'})$. A contradiction to Corollary 3.7.

Let $\frac{m}{2} < n \leq m$. Set $l := \left\lceil \frac{n - \lceil \frac{m}{2} \rceil}{2} \right\rceil$. Let A be an amalgam, which is created by identifying pairs of vertices u_{i+l+1} and v_i for each $0 \leq i < \left\lceil \frac{m}{2} \right\rceil$, see Figure 3.3 for reference. It is easy to verify that A is a convex amalgam. Due to Corollary 3.7, to determine the value of $h_A(P_n^A, C_m^A)$ it suffices to find the vertex in A with the maximum distance to $P_n^A \cap C_m^A$. Clearly, the candidates are the two endpoints of the path P_n^A that are outside of $P_n^A \cap C_m^A$ (vertices u_1^A and u_n^A) and a vertex of $V(C_m^A) \setminus V(P_n^A \cap C_m^A)$ with the maximum distance to $P_n^A \cap C_m^A$ (the vertex $v^A_{0-\lceil \frac{\lfloor m/2 \rfloor}{2} \rceil}$).

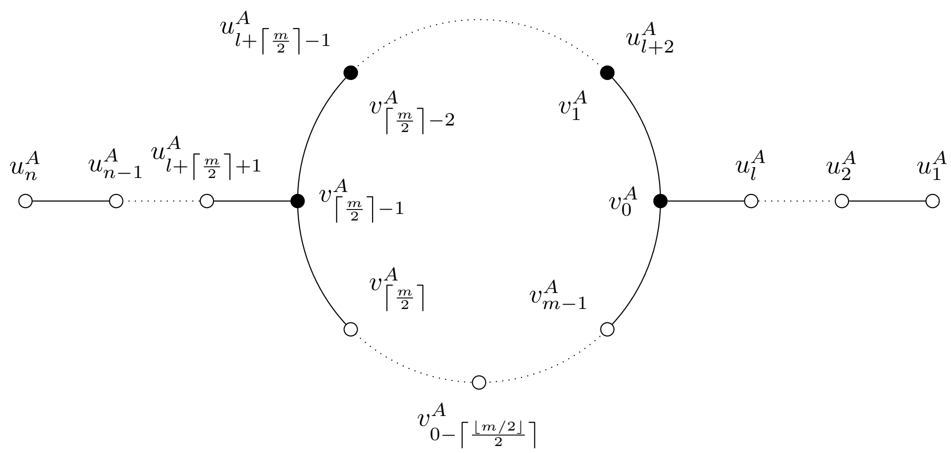


Figure 3.3: An amalgam A of path P_n (vertices u_i) and cycle C_m (vertices v_j).

Note that $d_A(u_1^A, P_n^A \cap C_m^A) = d_A(u_1^A, u_{l+1}^A) = l$ and $d_A(u_n^A, P_n^A \cap C_m^A) = d_A(u_n^A, u^A_{l+\lceil \frac{m}{2} \rceil})$.

The distance between the vertices u_n^A and $u_{l+\lceil \frac{m}{2} \rceil}^A$ can be expressed as the difference between the length of the path P_n and the length of the path between u_1^A and $u_{l+\lceil \frac{m}{2} \rceil}^A$. Therefore,

$$\begin{aligned} d_A(u_n^A, u_{l+\lceil \frac{m}{2} \rceil}^A) &= \\ n-1-(l+\left\lceil \frac{m}{2} \right\rceil - 1) &= \\ n-1-l-\left\lceil \frac{m}{2} \right\rceil + 1 &= \\ 2\frac{n-\lceil \frac{m}{2} \rceil}{2} - l &\leq \\ 2\left\lceil \frac{n-\lceil \frac{m}{2} \rceil}{2} \right\rceil - l &= \\ 2l-l = l. \end{aligned}$$

The distance $d_A(v_{0-\lceil \frac{\lfloor m/2 \rfloor}{2} \rceil}^A, P_n^A \cap C_m^A) = \min\{\left\lceil \frac{\lfloor m/2 \rfloor}{2} \right\rceil, d_A(v_{0-\lceil \frac{\lfloor m/2 \rfloor}{2} \rceil}^A, v_{\lceil \frac{m}{2} \rceil - 1}^A)\}$. It holds that

$$\begin{aligned} d_A(v_{0-\lceil \frac{\lfloor m/2 \rfloor}{2} \rceil}^A, v_{\lceil \frac{m}{2} \rceil - 1}^A) &= \\ m - d_A(v_{0-\lceil \frac{\lfloor m/2 \rfloor}{2} \rceil}^A, v_0^A) - d_A(v_0^A, v_{\lceil \frac{m}{2} \rceil - 1}^A) &= \\ m - \left\lceil \frac{m}{2} \right\rceil + 1 - \left\lceil \frac{\lfloor \frac{m}{2} \rfloor}{2} \right\rceil &= \\ \left\lfloor \frac{m}{2} \right\rfloor - \left\lceil \frac{\lfloor \frac{m}{2} \rfloor}{2} \right\rceil + 1 &= \\ \left\lfloor \frac{\lfloor \frac{m}{2} \rfloor}{2} \right\rfloor + 1 \geq \left\lceil \frac{\lfloor \frac{m}{2} \rfloor}{2} \right\rceil. \end{aligned}$$

Therefore, $d_A(v_{0-\lceil \frac{\lfloor m/2 \rfloor}{2} \rceil}^A, P_n^A \cap C_m^A) = \left\lceil \frac{\lfloor m/2 \rfloor}{2} \right\rceil$. Since $l \leq \left\lceil \frac{\lfloor m/2 \rfloor}{2} \right\rceil$, by Corollary 3.7, $h_A(P_n^A, C_m^A) = \left\lceil \frac{\lfloor m/2 \rfloor}{2} \right\rceil$. It follows that $\mathcal{H}(P_n, C_m) \leq \left\lceil \frac{\lfloor m/2 \rfloor}{2} \right\rceil$. See Figure 3.3 for reference.

Suppose a convex amalgam A' with $h_{A'}(P_n^{A'}, C_m^{A'}) < \left\lceil \frac{\lfloor \frac{m}{2} \rfloor}{2} \right\rceil$ exists. Define $h := \left\lceil \frac{\lfloor \frac{m}{2} \rfloor}{2} \right\rceil$. Again, due to convexity, $P_n^{A'} \cap C_m^{A'}$ is isomorphic to a path on k vertices, $1 \leq k \leq \lceil \frac{m}{2} \rceil$. Say the vertices in $P_n^{A'} \cap C_m^{A'}$ are $v_i^{A'}, v_{i+1}^{A'}, \ldots, v_{i+k-1}^{A'}$. We consider the vertex $v_{i-h}^{A'}$. Since $d_{A'}(v_i^{A'}, v_{i-h}^{A'}) = h$ and

$$\begin{aligned}
d_{A'}(v_{i+k-1}^{A'}, v_{i-h}^{A'}) &= \\
m - d_{A'}(v_i^{A'}, v_{i+k-1}^{A'}) - d_{A'}(v_i^{A'}, v_{i-h}^{A'}) &= \\
m - (k-1) - h &= \\
m - k + 1 - h &\geq \\
m - \left\lceil \frac{m}{2} \right\rceil + 1 - h &= \\
\left\lfloor \frac{m}{2} \right\rfloor + 1 - \left\lceil \frac{\lfloor \frac{m}{2} \rfloor}{2} \right\rceil &= \\
\left\lfloor \frac{\lfloor \frac{m}{2} \rfloor}{2} \right\rfloor + 1 &\geq \\
\left\lceil \frac{\lfloor \frac{m}{2} \rfloor}{2} \right\rceil &= h,
\end{aligned}$$

it follows that $d_{A'}(v_{i-h}^{A'}, P_n^{A'} \cap C_m^{A'}) = h = \left\lceil \frac{\lfloor \frac{m}{2} \rfloor}{2} \right\rceil > h_{A'}(P_n^{A'}, C_m^{A'})$. A contradiction to Corollary 3.7. Using Remark 3.14, the assertion follows.

Let $n > m$. Set $l := \left\lceil \frac{n - \lceil \frac{m}{2} \rceil}{2} \right\rceil$. Let A be an amalgam, which is created by identifying pairs of vertices u_{i+l+1} and v_i for each $0 \leq i < \lceil \frac{m}{2} \rceil$. It is easy to verify that A is a convex amalgam. As in the previous case, the value of $h_A(P_n^A, C_m^A)$ can be determined by finding a vertex of A with the maximum distance to $P_n^A \cap C_m^A$; the same candidate vertices have to be considered (vertices u_1^A, u_n^A and $v^A_{0-\lceil \frac{\lfloor m/2 \rfloor}{2} \rceil}$). Following the same line of thought as in the previous case and taking into account that $l \geq \left\lceil \frac{\lfloor \frac{m}{2} \rfloor}{2} \right\rceil$, it follows that $h_A(P_n^A, C_m^A) = \left\lceil \frac{n - \lceil \frac{m}{2} \rceil}{2} \right\rceil$ and $\mathcal{H}(P_n, C_m) \leq \left\lceil \frac{n - \lceil \frac{m}{2} \rceil}{2} \right\rceil$.

Suppose a convex amalgam A' with $h_{A'}(P_n^{A'}, C_m^{A'}) < \left\lceil \frac{n - \lceil \frac{m}{2} \rceil}{2} \right\rceil$ exists. Due to convexity, $P_n^{A'} \cap C_m^{A'}$ is isomorphic to a path on k vertices, $1 \leq k \leq \lceil \frac{m}{2} \rceil$. Say the vertices in $P_n^{A'} \cap C_m^{A'}$ are $v_i^{A'}, v_{i+1}^{A'}, \ldots, v_{i+k-1}^{A'}$. The length of path $P_n^{A'}$ is clearly $n-1$ and equals $d_{A'}(u_1^{A'}, v_i^{A'}) + d_{A'}(v_i^{A'}, v_{i+k-1}^{A'}) + d_{A'}(v_{i+k-1}^{A'}, u_n^{A'})$. On the other hand, by Corollary 3.7, it holds that $d_{A'}(u_1^{A'}, v_i^{A'}) \leq h_{A'}(P_n^{A'}, C_m^{A'})$ and $d_{A'}(v_{i+k-1}^{A'}, u_n^{A'}) \leq h_{A'}(P_n^{A'}, C_m^{A'})$.

Putting this together, we determine that

$$\begin{aligned}
\ell(P_n^{A'}) = d_{A'}(u_1^{A'}, v_i^{A'}) + d_{A'}(v_i^{A'}, v_{i+k-1}^{A'}) + d_{A'}(v_{i+k-1}^{A'}, u_n^{A'}) &\leq \\
h_{A'}(P_n^{A'}, C_m^{A'}) + k - 1 + h_{A'}(P_n^{A'}, C_m^{A'}) &< \\
2\left\lceil \frac{n - \lceil \frac{m}{2} \rceil}{2} \right\rceil - 1 + k - 1 &\leq \\
2\frac{n - \lceil \frac{m}{2} \rceil + 1}{2} - 1 + \left\lceil \frac{m}{2} \right\rceil - 1 &= n - 1.
\end{aligned}$$

So, $n - 1 = \ell(P_n^{A'}) < n - 1$, a contradiction. □

Now, we derive a formula for the Hausdorff distance between two cycles. If the cycles are isomorphic, the Hausdorff distance equals 0 by definition. For non-isomorphic cycles we get the following proposition.

Proposition 3.17. *If C_n and C_m are two cycles of length n and m, respectively, with $n > m \geq 3$, then $\mathcal{H}(C_n, C_m) = \left\lceil \frac{n - \lceil \frac{m}{2} \rceil}{2} \right\rceil$.*

Proof. Denote vertices of C_n with $u_0, \ldots, u_{n-1}$, where $u_i u_{i+1} \in E(C_n)$, for each $i \in \{0, \ldots, n-1\}$, and vertices of C_m with $v_0, \ldots, v_{m-1}$, where $v_i v_{i+1} \in E(C_m)$, for each $i \in \{0, \ldots, m-1\}$. All indices are computed modulo of the length of the corresponding cycle.

Let A be an amalgam, which is created by identifying pairs of vertices u_i and v_i for each $1 \leq i \leq \lceil \frac{m}{2} \rceil$. Since every subgraph of C_m (C_n) isomorphic to a path on $\lceil \frac{m}{2} \rceil$ vertices is a convex subgraph of C_m (and also C_n), A is a convex amalgam. Thus, by Corollary 3.7 $h_A(C_n^A, C_m^A) = \left\lceil \frac{n - \lceil \frac{m}{2} \rceil}{2} \right\rceil$ and $\mathcal{H}(C_n, C_m) \leq \left\lceil \frac{n - \lceil \frac{m}{2} \rceil}{2} \right\rceil$.

Suppose a convex amalgam A' with $h_{A'}(C_n^{A'}, C_m^{A'}) < \left\lceil \frac{n - \lceil \frac{m}{2} \rceil}{2} \right\rceil$ exists. Therefore, $C_n^{A'} \cap C_m^{A'}$ is isomorphic to a path on k vertices, $1 \leq k \leq \lceil \frac{m}{2} \rceil$. Say the vertices in $C_n^{A'} \cap C_m^{A'}$ are $u_i^{A'}, u_{i+1}^{A'}, \ldots, u_{i+k-1}^{A'}$. We now choose the vertex $u^{A'}_{i - \left\lceil \frac{n - \lceil \frac{m}{2} \rceil}{2} \right\rceil}$. Since $d_{A'}(u_i^{A'}, u^{A'}_{i - \left\lceil \frac{n - \lceil \frac{m}{2} \rceil}{2} \right\rceil}) = \left\lceil \frac{n - \lceil \frac{m}{2} \rceil}{2} \right\rceil$ and $d_{A'}(u_{i+k-1}^{A'}, u^{A'}_{i - \left\lceil \frac{n - \lceil \frac{m}{2} \rceil}{2} \right\rceil}) \geq \left\lceil \frac{n - \lceil \frac{m}{2} \rceil}{2} \right\rceil$, it follows that $d_{A'}(u^{A'}_{i - \left\lceil \frac{n - \lceil \frac{m}{2} \rceil}{2} \right\rceil}, C_n^{A'} \cap C_m^{A'}) = \left\lceil \frac{n - \lceil \frac{m}{2} \rceil}{2} \right\rceil > h_{A'}(C_n^{A'}, C_m^{A'})$. A contradiction to Corollary 3.7. □

3.2 Trees and the Hausdorff Distance

Trees often appear in chemical graph theory, since many organic molecules have a graph representation that is a tree (e.g. saturated hydrocarbons). Isomers, for example, have the same chemical formula but different molecular structures. One of the problems that arises with respect to chemical structure is to determine whether two chemical structures are the same or how similar they are. Say that chemical structures can be presented as trees. This means we have to determine whether two trees are isomorphic; this is a simple problem and can be done in linear time [1]. Also, as a measure of the similarity of two non-isomorphic trees one can use a maximum common subtree of the two trees compared. The problem of finding a maximum common subtree of two arbitrary trees can be done in non-linear polynomial time [49].

On the other hand, to determine the Hausdorff distance between two trees, using a maximum common subtree to form a convex amalgam of two arbitrary trees may not produce an optimal amalgam (see Example 3.18). Therefore, the mentioned algorithms may not suffice in determining the Hausdorff distance of two arbitrary trees.

Example 3.18. *In Figure 3.4, we have two non-isomorphic trees T_1 (left hand side) and T_2 (right hand side) with central vertices c_1 and c_2, respectively. A maximum common subtree of T_1 and T_2 is clearly isomorphic to T_2.*

Let A_1 be a convex amalgam obtained from T_1 and T_2 by identifying the subgraphs induced by the sets of black vertices (using the maximum common subtree). In this case $h_{A_1}(T_1^{A_1}, T_2^{A_1}) = 4$. On the other hand, one can form a convex amalgam A_2 by identifying the central vertices of the two trees for which $h_{A_2}(T_1^{A_2}, T_2^{A_2}) = 3$ and, therefore, $\mathcal{H}(T_1, T_2) \leq 3$. It follows that A_1 is not an optimal amalgam.

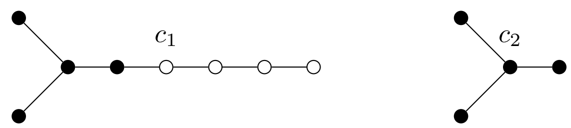


Figure 3.4: Maximum common subtree does not suffice.

In the following subsections, we present certain bounds for the Hausdorff distance between two trees, some formulae for special cases, and an exact polynomial-time algorithm for computing the Hausdorff distance between two trees.

3.2.1 Some General Results for Trees

It is well known that any tree has either exactly one central vertex or exactly two central vertices that are adjacent. We say that a tree T is *central*, if $|\mathrm{center}(T)| = 1$, otherwise it is *bicentral*. Also, for an arbitrary tree T it holds that $\mathrm{diam}(T) = 2\mathrm{rad}(T) - 1$, if T is bicentral, and $\mathrm{diam}(T) = 2\mathrm{rad}(T)$, if T is central. This fact, together with Theorem 3.12, immediately implies the following corollary.

Corollary 3.19. *Let T_1 and T_2 be two arbitrary trees. Then*

$$\mathcal{H}(T_1, T_2) \leq \max\left\{\left\lceil\frac{\mathrm{diam}(T_1)}{2}\right\rceil, \left\lceil\frac{\mathrm{diam}(T_2)}{2}\right\rceil\right\}.$$

Clearly, if one of the trees is trivial, one obtains an optimal amalgam of the two trees by identifying the only vertex of the trivial tree with a central vertex of the other tree and the bound is sharp. For this reason, in the following results we restrict ourselves to non-trivial trees.

Proposition 3.20. *Let T_1 and T_2 be two non-trivial trees with $\mathrm{diam}(T_1) \geq \mathrm{diam}(T_2)$. If T_1 is bicentral, then $\mathcal{H}(T_1, T_2) < \mathrm{rad}(T_1)$.*

Proof. Let $\mathrm{center}(T_1) = \{c_1, c_2\}$. Let c be a central vertex of T_2 and c' its arbitrary neighbour, if T_2 is central, otherwise let c' be the other central vertex of T_2. Let H_1 be the subgraph of T_1 induced on the set $\mathrm{center}(T_1)$, and H_2 the subgraph of T_2 induced on the set $\{c, c'\}$. Let A be a convex amalgam of T_1 and T_2 obtained by identifying the graphs H_1 and H_2.

For any vertex $u \in V(T_1^A)$ it holds that $d_A(u, T_1^A \cap T_2^A) < \mathrm{rad}(T_1)$, since both central vertices are in $T_1^A \cap T_2^A$. Let $v \in V(T_2^A)$. If T_2 is bicentral (both its central vertices are also in $T_1^A \cap T_2^A$), then $d_A(v, T_1^A \cap T_2^A) < \mathrm{rad}(T_2) \leq \mathrm{rad}(T_1)$. If T_2 is central, then $\mathrm{rad}(T_2) < \mathrm{rad}(T_1)$. Since $c^A \in V(T_1^A \cap T_2^A)$, it holds that $d_A(v, T_1^A \cap T_2^A) \leq \mathrm{rad}(T_2) < \mathrm{rad}(T_1)$. Using Corollary 3.7 the assertion follows immediately. □

Next, we will study some properties of the optimal amalgams of trees. Remember, a convex amalgam of two graphs is called optimal if it gives rise to the Hausdorff distance between the two graphs.

Theorem 3.21. *Let T_1 and T_2 be two arbitrary non-trivial trees, with* $\mathrm{diam}(T_1) \geq \mathrm{diam}(T_2)$. *Let* $c \in \mathrm{center}(T_1)$. *Then for every optimal amalgam $A \in \mathcal{X}(T_1, T_2)$ it holds that* $\{c^A\} \subseteq V(T_1^A \cap T_2^A)$.

Proof. Assume that $A \in \mathcal{X}(T_1, T_2)$ with $h_A(T_1^A, T_2^A) = \mathcal{H}(T_1, T_2)$ exists, such that at least one central vertex of T_1, say v, is not in $T_1^A \cap T_2^A$. Then it holds that $d_A(v, T_1^A \cap T_2^A) \geq 1$.

Suppose T_1 is central. Since $T_1^A \cap T_2^A$ is convex in A, a vertex $u \in V(T_1^A) \backslash V(T_1^A \cap T_2^A)$ with $d_A(v, u) = \left\lceil \frac{\mathrm{diam}(T_1^A)}{2} \right\rceil$ exists. But then $d_A(u, T_1^A \cap T_2^A) \geq \left\lceil \frac{\mathrm{diam}(T_1^A)}{2} \right\rceil + 1$. This is a contradiction to Corollary 3.7 and Corollary 3.19 together with the assumption $\mathrm{diam}(T_1) \geq \mathrm{diam}(T_2)$.

Suppose T_1 is bicentral. Since $T_1^A \cap T_2^A$ is convex in A, it follows that a vertex $u \in V(T_1^A) \backslash V(T_1^A \cap T_2^A)$ with $d_A(v, u) = \mathrm{rad}(T_1) - 1$ exists. But then $d_A(u, T_1^A \cap T_2^A) \geq \mathrm{rad}(T_1)$. This is a contradiction to Corollary 3.7 and Proposition 3.20. □

Let G be a graph and H its subgraph with a property P. We say H is *a minimal subgraph* with the property P if there is no preexisting proper subgraph of H with the property P.

Theorem 3.22. *Let T_1 and T_2 be two arbitrary non-trivial trees, with* $\mathrm{diam}(T_1) \geq \mathrm{diam}(T_2)$. *Let $0 \leq k \leq rad(T_1)$ be a fixed integer. Let H be a minimal subtree of T_1, containing a central vertex of T_1, such that* $\max_{u \in V(T_1) \backslash V(H)}\{d_{T_1}(u, H)\} \leq k$. *If T_2 does not contain a subgraph isomorphic to H, then $\mathcal{H}(T_1, T_2) > k$.*

Proof. Suppose, $\mathcal{H}(T_1, T_2) \leq k$, for a fixed integer $0 \leq k \leq rad(T_1)$. In that case, a convex amalgam A of T_1 and T_2 exists, such that $h_A(T_1^A, T_2^A) = k$. Let H' be the subgraph of T_1 corresponding to $T_1^A \cap T_2^A$. By Theorem 3.21 the graph H' contains a central vertex of T_1. By Corollary 3.7 it holds true that $\max_{u \in V(T_1) \backslash V(H')}\{d_{T_1}(u, H')\} \leq k$. Now let H be a minimal subtree of H', such that $\max_{u \in V(T_1) \backslash V(H)}\{d_{T_1}(u, H)\} \leq k$ is still true. Clearly, H is a (convex) subgraph of H', therefore H^A is a convex subgraph of $T_1^A \cap T_2^A$. Thus, T_2 clearly contains a subgraph isomorphic to H. □

The minimal subgraph H of a tree T, with the properties as required by Theorem 3.22, can be easily found as follows. Set $S := \mathrm{center}(T)$. Say k is a fixed integer as in Theorem 3.22. Choose a central vertex c of the tree T. Now, for each leaf u of the tree consider the path P_u from the leaf to the central vertex c. If $\ell(P_u) \leq k$, then do nothing. Otherwise, let $v_u \in V(P_u)$ be the vertex with $d_T(u, v_u) = k$. Let R_u be the path from v_u to c. Add the vertices of R_u to S. Clearly, the graph induced on the vertices in S is the subgraph we are constructing, i. e. $H = \langle S \rangle$.

Theorem 3.21 states that the center of the tree with the larger diameter is always in the intersection of an optimal amalgam. On the other hand, trees T_1 and T_2 with $\mathrm{diam}(T_1) \geq \mathrm{diam}(T_2)$ also exist, such that no central vertex of T_2 is in $T_1^A \cap T_2^A$ for any optimal amalgam A of T_1 and T_2, as Example 3.23 demonstrates.

Example 3.23. *In Figure 3.5, we have two non-isomorphic trees T_1 and T_2. The (connected) subgraphs induced on the sets of blacks vertices in each tree are clearly isomorphic. Moreover, since they are connected, they are also convex in the corresponding graphs. Therefore, by identifying these two subgraphs we obtain a convex amalgam A such that, by Corollary 3.7, $h_A(T_1^A, T_2^A) = 4$. Therefore, $\mathcal{H}(T_1, T_2) \leq 4$.*

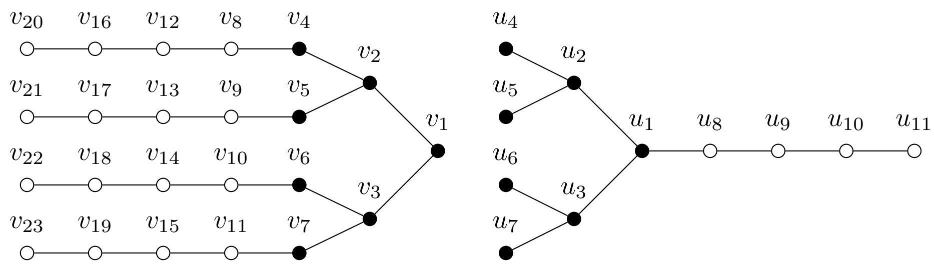


Figure 3.5: Trees T_1 (left) and T_2 (right).

To see that $\mathcal{H}(T_1, T_2) \geq 4$, suppose that an amalgam $A' \in \mathcal{X}(T_1, T_2)$ exists, for which it holds that $h_{A'}(T_1^{A'}, T_2^{A'}) \leq 3$. Using Theorem 3.22, a minimal subtree H of T_1 containig the center of T_1 and satisfying the condition $\max_{u \in V(T_1) \setminus V(H)} \{d_{T_1}(u, H)\} \leq 3$ is the subgraph induced on the set of vertices $\{v_1, v_2, \ldots, v_{11}\}$. Clearly, T_2 contains no subgraph isomorphic to H, therefore $\mathcal{H}(T_1, T_2) > 3$. It follows that $\mathcal{H}(T_1, T_2) = 4$.

Now, we show that no central vertex of T_2 is in some optimal amalgam of T_1 and T_2. Note that u_8 is the only central vertex of T_2. Suppose an amalgam A' exists such that $h_{A'}(T_1^{A'}, T_2^{A'}) = 4$ and $u_8^{A'} \in V(T_1^{A'} \cap T_2^{A'})$. For the same reason as above the set of *vertices* $\{v_1^{A'}, v_2^{A'}, \ldots, v_7^{A'}\}$ *is a subset of* $V(T_1^{A'} \cap T_2^{A'})$. *Since the subgraph of* T_2 *induced on the set of black vertices in Figure 3.5 is the only subgraph of* T_2*, which is isomophic to the subgraph of* T_1 *induced on the set of (black) vertices* $\{v_1, v_2, \ldots, v_7\}$ *and it does not contain* u_8*, it follows that no such amalgam* A' *exists.*

Proposition 3.24. *Let* T_1 *and* T_2 *be two arbitrary non-trivial trees, with* $\mathrm{diam}(T_1) \geq \mathrm{diam}(T_2)$. *Let* $A \in \mathcal{X}(T_1, T_2)$ *be an optimal amalgam of* T_1 *and* T_2. *Then there exist* $c_1 \in \mathrm{center}(T_1)$ *and* $c_2 \in \mathrm{center}(T_2)$ *such that* $d_A(c_1^A, c_2^A) \leq \mathcal{H}(T_1, T_2)$.

Proof. Choose vertices $c_1 \in \mathrm{center}(T_1)$ and $c_2 \in \mathrm{center}(T_2)$ such that $d_A(c_1^A, c_2^A)$ is the smallest possible distance. Choose the vertex $u \in V(T_1^A)$ for which it holds that $\mathrm{rad}(T_1) = d_A(c_1^A, u) \leq d_A(c_2^A, u)$. Such a vertex u exists because $c_1 \in \mathrm{center}(T_1)$. Note that if T_1 is bicentral, the second central vertex is on the shortest path between c_1 and u. Choose a vertex v for which it holds that $v \in V(T_1^A \cap T_2^A)$ and $d_A(u, v)$ is the smallest possible. Then

$$\begin{aligned}
\mathcal{H}(T_1, T_2) \geq& \\
d_A(u, v) =& \\
d_A(c_1^A, u) - d_A(c_1^A, v) =& \\
d_A(c_1^A, u) - \big(d_A(c_2^A, v) - d_A(c_1^A, c_2^A)\big) \geq& \\
\mathrm{rad}(T_1) - \big(\mathrm{rad}(T_2) - d_A(c_1^A, c_2^A)\big) =& \\
d_A(c_1^A, c_2^A) + (\mathrm{rad}(T_1) - \mathrm{rad}(T_2)) \geq& \quad d_A(c_1^A, c_2^A).
\end{aligned}$$

□

The following proposition shows that the bound from Proposition 3.24 is sharp.

Proposition 3.25. *For an arbitrary non-negative integer* k *there exist trees* T_1 *and* T_2*, with* $\mathrm{diam}(T_1) \geq \mathrm{diam}(T_2)$ *and* $\mathcal{H}(T_1, T_2) = k$*, such that for every optimal amalgam* A *of* T_1 *and* T_2 *it holds that* $d_A(c_1^A, c_2^A) = \mathcal{H}(T_1, T_2)$*, where* $c_1 \in \mathrm{center}(T_1)$ *and* $c_2 \in \mathrm{center}(T_2)$.

Proof. Let k be a fixed non-negative integer. We will construct two non-isomorphic trees T_1 and T_2, such that the Hausdorff distance between T_1 and T_2 is $\mathcal{H}(T_1, T_2) = k$

and the distance between the vertices c_1^A and c_2^A corresponding to the centers of T_1 and T_2 in every optimal convex amalgam A is $d_A(c_1^A, c_2^A) = k$.

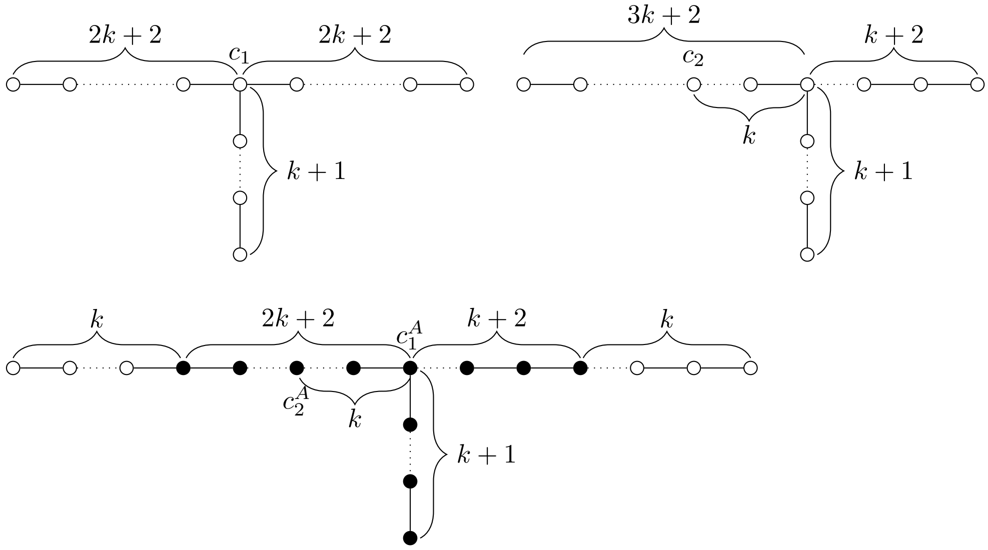


Figure 3.6: Trees T_1, T_2 and an optimal amalgam A of T_1 and T_2.

Let T_1 be the tree constructed from a path of length $4k+4$ and a path of length $k+1$, where we identify one end-vertex of the shorter path with the central vertex of the longer path; see the top left-hand tree in Figure 3.6 for reference. T_1 is a star-like tree with three rays, two of length $2k+2$ and one of length $k+1$. Clearly, c_1 is the only central vertex of T_1.

Next, let T_2 be the tree constructed from a path of length $4k+4$ and a path of length $k+1$, where we identify one end-vertex of the shorter path with a vertex at distance k from the central vertex of the longer path; see the top right-hand tree in Figure 3.6 for reference. By construction, T_2 is also a star-like tree with three rays, one of length $3k+2$, one of length $k+2$, and one of length $k+1$, with exactly one central vertex, namely c_2.

Now we construct an amalgam A of T_1 and T_2 as shown in the bottom tree in Figure 3.6. Clearly, A is a convex amalgam of T_1 and T_2, the distance between vertices corresponding to the centers of the trees is $d_A(c_1^A, c_2^A) = k$. From the construction and Corollary 3.7 it is also obvious that $\mathcal{H}(T_1, T_2) \leq h_A(T_1^A, T_2^A) = k$. Using Theorem 3.22, it can easily be confirmed that $\mathcal{H}(T_1, T_2) > k - 1$.

All that is left is to show that in every optimal amalgam of trees T_1 and T_2 the distance between vertices corresponding to the central vertices of covers is k. Let A be an arbitrary optimal amalgam of T_1 and T_2. Note, $\operatorname{diam}(T_1) = \operatorname{diam}(T_2) = 4k + 4$. From Theorem 3.21 it follows that $c_1^A \in V(T_1^A \cap T_2^A)$. Moreover, we claim that the vertices corresponding to all neighbours of c_1 are also in $T_1^A \cap T_2^A$. Towards contradiction, let $v \in V(T_1)$ be a neighbour of c_1 such that $v^A \notin V(T_1^A \cap T_2^A)$. Also, let w denote the leaf of T_1 such that the path $P_{v,w}$ from v to w does not contain c_1. Since $T_1^A \cap T_2^A$ is convex (and therefore connected) no vertex of $P_{v,w}$ can be in $T_1^A \cap T_2^A$. But then $d_A(w^A, T_1^A \cap T_2^A) = k + 1 > k$, a contradiction with Theorem 3.6 and the fact that $\mathcal{H}(T_1, T_2) = k$. It follows that c_1 and all its three neighbours are in $T_1^A \cap T_2^A$. Since T_2 contains exactly one vertex, say u, of degree three and A is a convex amalgam of T_1 and T_2, this vertex and its neighbours must also be in $T_1^A \cap T_2^A$. Moreover c_1 is mapped with an isomorphism to u. Since A was chosen arbitrarily, the distance between vertices c_1^A and c_2^A is the same in all optimal amalgams. Clearly, $d_A(c_1^A, c_2^A) = k$. $\square$

3.2.2 The Algorithm for Trees

The algorithm for the Hausdorff distance of two trees described in this subsection runs in polynomial time. The algorithm is recursive and it utilizes the divide and conquer technique. As a subtask it also uses the procedure that is based on the well-known graph algorithm for finding the maximum bipartite matching. The main procedure of the algorithm is working with the so called *top-down common subtrees* and, therefore, we need the following definitions summarized in [49].

Definition 3.26. *Let $T = (V(T), E(T))$ be a rooted tree. A* subtree *of T is a connected subgraph of T. A* top-down subtree *$S = (V(S), E(S))$ is a rooted subtree of T, where $parent[v] \in V(S)$, for all non-root vertices $v \in V(S)$. The root vertex of a top-down subtree S is the same vertex as the root vertex of the rooted tree T. Let $u \in V(T)$. A subtree of T is called* a subtree rooted at u *if it is induced on the vertex set $\{u\} \cup descendants[u]$.*

Definition 3.27. *Two rooted trees $T_1 = (V(T_1), E(T_1))$ and $T_2 = (V(T_2), E(T_2))$ are* isomorphic *if there is a bijection $M \subseteq V(T_1) \times V(T_2)$ such that $(root[T_1], root[T_2]) \in M$ and $(parent[v], parent[u]) \in M$, for all non-root vertices $v \in V(T_1), u \in V(T_2)$ with $(v, u) \in M$. The set M is called* a rooted tree isomorphism.

Definition 3.28. *A* top-down common subtree *of the rooted tree $T_1 = (V(T_1), E(T_1))$ and the rooted tree $T_2 = (V(T_2), E(T_2))$ is the structure (S_1, S_2, M), where $S_1 = (V(S_1), E(S_1))$ is a top-down subtree of T_1, $S_2 = (V(S_2), E(S_2))$ is a top-down subtree of T_2, and $M \subseteq V(S_1) \times V(S_2)$ is a rooted tree isomorphism of S_1 and S_2.*

Example 3.29. *In Figure 3.7, there are two trees T_1 and T_2. A subtree S_1 induced on the vertex set $\{v_2, v_6, v_7, v_8, v_9, v_{11}\}$ is a top-down subtree of T_1. Similarly, a subtree S_2 induced on the vertex set $\{u_3, u_4, u_5, u_6, u_7, u_8\}$ is a top-down subtree of T_2.*

A subtree of T_1, induced with grey vertices, is a subtree rooted at vertex v_5 and it is not a top-down subtree since, for example, v_5 is not the root and $parent[v_5]$ is not in the subtree.

Let $M = \{(v_2, u_3), (v_6, u_4), (v_7, u_5), (v_8, u_6), (v_9, u_7), (v_{11}, u_8)\}$ be a rooted tree isomorphism of S_1 and S_2. The structure (S_1, S_2, M) is a top-down common subtree of rooted trees T_1 and T_2.

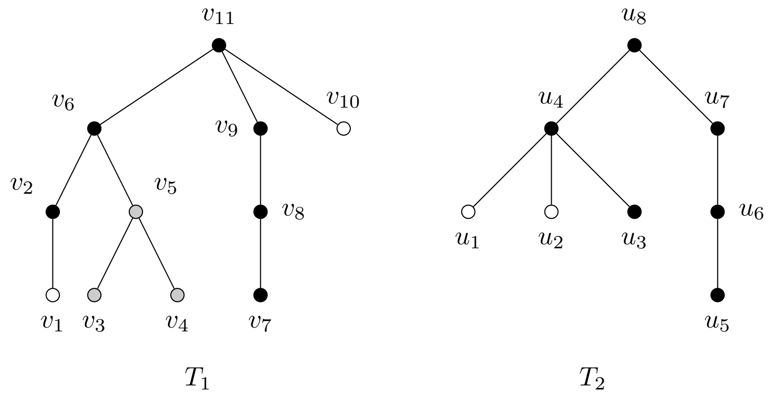


Figure 3.7: Illustration of the concepts defined above.

Recall that in order to determine the Hausdorff distance between two trees, one has to find a convex common subgraph (a subtree) of the input trees such that the distance between the covers of the corresponding amalgam is minimized (an optimal amalgam). Note, a subtree of a tree is always a convex subgraph.

A convex amalgam of trees T_1 and T_2 is a tree. If we root an amalgam A at a vertex from the intersection of the amalgam $v^A \in V(T_1^A \cap T_2^A)$, then the intersection of the amalgam is a top-down subtree of the amalgam A. The subtrees of T_1 and T_2 that give rise to the amalgam A are top-down subtrees of the trees T_1 and T_2 rooted in the vertices corresponding to the vertex v^A.

Any optimal amalgam can be obtained by finding the appropriate top-down subtrees of the input trees. For this reason, the algorithm works on top-down common subtrees and, therefore, we have to root both input trees.

An optimal top-down amalgam is an amalgam optimal with respect to the rooted structure (meaning that the corresponding isomorphism is a rooted tree isomorphism). We call a top-down common subtree optimal if the corresponding amalgam is an optimal top-down amalgam. Note that since the corresponding isomorphism is a rooted tree isomorphism, both root vertices of an optimal top-down common subtree have to be in the intersection of the corresponding amalgam.

Example 3.30. *We can see that in Figure 3.8 there are two non-isomorphic rooted trees T_1 and T_2. Since the top-down common subtree drawn with black vertices gives rise to an amalgam in which the distance between the covers is equal to one, it follows that this is an optimal top-down common subtree.*

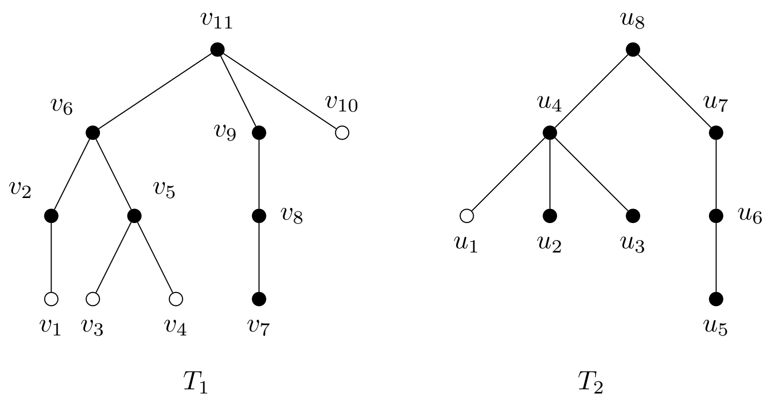


Figure 3.8: An optimal top-down common subtree of trees T_1 (rooted at v_{11}) and T_2 (rooted at u_8). It is drawn with black vertices in both trees.

As the input of the algorithm we get two non-rooted trees $T_1 = (V(T_1), E(T_1))$ and $T_2 = (V(T_2), E(T_2))$, where $\mathrm{diam}(T_1) \geq \mathrm{diam}(T_2)$. Since a central vertex of T_1 is in the intersection of any optimal amalgam (Theorem 3.21), we can root T_1 in a central vertex. For T_2 we have no such property.

In Example 3.31 we can see that an optimal top-down amalgam is not necessarily an optimal amalgam (non-rooted). This depends on the choice of the root vertices of the input trees T_1 and T_2. If we root the tree T_2 in each vertex $v \in V(T_2)$ and run the procedure for each case, then we are guaranteed that the algorithm is able to find a common subtree of the input trees such that the distance between the covers of the corresponding amalgam is minimized. In other words, the algorithm finds an optimal top-down amalgam that is also an optimal amalgam.

Example 3.31. *Figure 3.9 shows an optimal top-down common subtree of the non isomorphic rooted trees T_1 and T_2. Trees T_1 and T_2 are similar to those in Figure 3.8, with the difference that the tree T_2 here is rooted in the vertex u_7. An optimal top-down common subtree is induced on black vertices and it gives rise to an amalgam in which the distance between the covers is equal to two. Therefore, this common subtree does not minimize the distance between the covers of the corresponding amalgam of non-rooted trees. The minimum distance is one, see Figure 3.8.*

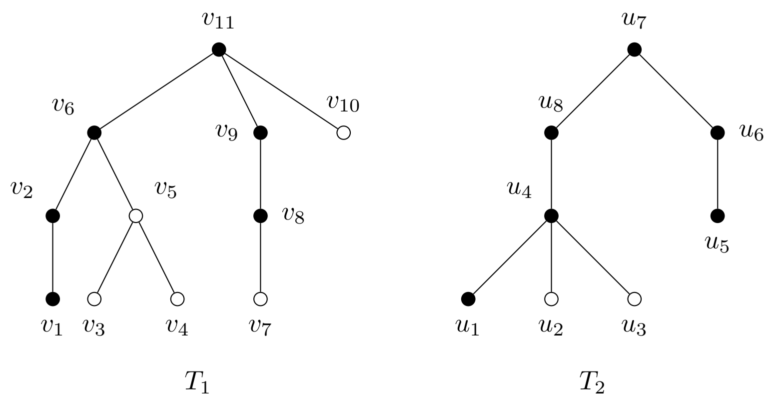


Figure 3.9: An optimal top-down common subtree of trees T_1 (rooted at v_{11}) and T_2 (rooted at u_7), induced on black vetrices in both trees.

Now we are ready to present Algorithm 1 that determines the Hausdorff distance between two arbitrary trees T_1 and T_2 in polynomial time. The corresponding common subtree structure is also determined by the algorithm.

The algorithm uses two procedures. With respect to Definition 3.28, an optimal top-down common subtree is a structure (S_1, S_2, M) and, therefore, we have to find subtrees S_1, S_2 and a mapping M between them. The procedure `OptimalTopDownCommonSubtree` is used to determine the distance between the covers of the optimal top-down amalgam of two rooted trees, and the procedure `ReconstructionOfMapping` is for the reconstruction of the subtree isomorphism that corresponds to the optimal amalgam. Notice that the first proce-

Algorithm 1: HausdorffDistanceBetweenTrees

input : An arbitrary trees T_1 and T_2, where $\mathrm{diam}(T_1) \geq \mathrm{diam}(T_2)$.
output: The Hausdorff distance between T_1 and T_2 stored in hd, and the corresponding common subtree structure stored in M.

$hd \leftarrow \infty$
$O \leftarrow \emptyset$
$r_1 \in center(T_1)$
Compute heights of vertices of tree T_1 rooted in r_1
foreach $u \in V(T_2)$ **do**
    $M' \leftarrow \emptyset$
    Compute heights of vertices of tree T_2 rooted in u
    $distance \leftarrow$ `OptimalTopDownCommonSubtree(`T_1,r_1,T_2,u,M'`)`
    if $distance < hd$ **then**
        $hd \leftarrow distance$
        $r_2 \leftarrow u$
        $O \leftarrow M'$
$M \leftarrow \emptyset$
`ReconstructionOfMapping(`T_1,r_1,r_2,O,M`)`

dure is called many times with different rooted trees as the input, while the second one (for the reconstruction of solution) is called just once, at the end of the algorithm.

First, let us describe the procedure `OptimalTopDownCommonSubtree`. The result of the procedure is the distance between the covers of the optimal top-down amalgam of the input rooted trees. Remember, an optimal top-down common subtree gives rise to an optimal top-down amalgam. An optimal top-down common subtree of the rooted input trees T_1 and T_2 can be constructed by breaking down the original rooted trees to rooted subtrees and finding optimal top-down common subtrees of those smaller rooted trees. We start with the root vertices r_1 and r_2, and traverse both trees recursively.

At each step we are in the vertices $v \in V(T_1)$ and $u \in V(T_2)$. We break down each rooted tree into rooted subtrees, such that the rooted subtrees of T_1 are rooted in the children of v and the rooted subtrees of T_2 are rooted in the children of u. We consider optimal top-down common subtrees for all possible pairs of those smaller subtrees. After we obtain all optimal top-down common subtrees for the children of v and the children of u, we can combine some of them and determine an optimal top-down common subtree of the subtree rooted at v and the subtree rooted at u. When we combine the optimal top-down common subtrees of the children of v and the children of u, we have to be careful that we do not combine one subtree with more than one other subtree.

An optimal top-down common subtree can easily be determined if one of the root vertices is a leaf of the original input tree (subtree rooted at this leaf is a trivial graph). If a vertex $v \in V(T_1)$ is a leaf (or a vertex $u \in V(T_2)$ is a leaf), then mapping v to u gives an optimal top-down common subtree. The distance between the covers of the corresponding amalgam is determined by the farthest vertex from the root in the other subtree. The farthest vertex from the root is always at the distance equal to $height[u]$ (or $height[v]$), respectively. Therefore, the case where one of the root vertices is a leaf is our stopping condition for the recursion.

Otherwise, none of the root vertices u and v is a leaf. Let $p = |children[v]|$, $q = |children[u]|$ and without loss of generality assume $p \geq q$. Denote with $v_1, \ldots, v_p$ and $u_1, \ldots, u_q$ the children of v and u, respectively. If $p > q$ then we add to the set $children[u]$ some *dummy* vertices $D = \{d_1, \ldots, d_{p-q}\}$, otherwise $D = \emptyset$. Build the complete bipartite graph

$$G_{vu} = (\{v_1, \ldots, v_p\} \cup (\{u_1, \ldots, u_q\} \cup D), E)$$

on $p + (q + |D|) = 2p$ vertices with partition sets $\{v_1, \ldots, v_p\}$ and $(\{u_1, \ldots, u_q\} \cup D)$. For technical reasons related to the reconstruction of an optimal top-down common subtree, the edges $(v_i, u_j) \in E$ of graph G_{vu} are ordered pairs of vertices. The first vertex is from T_1 and the second is from T_2. Each edge of G_{vu} is assigned a non-negative weight. We want to be able to determine the distance between the covers of an optimal top-down amalgam of a subtree rooted at v and a subtree rooted at u from the weights of the edges of the graph G_{vu}. The weight of an edge $(v_i, u_j) \in E$ is equal to the distance between the covers in an optimal top-down amalgam of a subtree (of T_1) rooted at v_i and a subtree (of T_2) rooted at u_j. Therefore, we will recursively call the same procedure with different root vertices. If $v_i \in V(T_1)$ is a leaf (or $u_j \in V(T_2)$ is a leaf), then the recursive call hits the stopping condition and returns the distance $height[u]$ (or $height[v]$), respectively. A dummy vertex d_k represents an empty subtree and no such top-down common subtree exists. If we want the weight of the edge $(v_i, d_k) \in E$ to possibly give rise to the distance between the covers of an optimal top-down amalgam of a subtree rooted at v and a subtree rooted at u, then the edge (v_i, d_k) must get the weight that is equal to the distance of the farthest vertex from the v_i plus 1 ($height[v] + 1$), i.e. vertices v and u are in the intersection of such optimal top-down amalgam while the whole subtree rooted at v_i is not in the intersection of such optimal top-down amalgam.

When all the weights of the graph G_{vu} are determined, we need to get the best possible combination of the corresponding optimal top-down amalgams to combine them into an optimal top-down amalgam A of a subtree rooted at v and a subtree rooted at u. We have to minimize the distance between the covers of an optimal top-down amalgam A. To do this we need the following concept.

Let M_{vu} be a perfect matching of the complete bipartite graph G_{vu} that minimizes the value of the largest weight (we will call it *an optimal perfect matching*).

Lemma 3.32. *The distance between the covers of an optimal top-down amalgam of a subtree (of T_1) rooted at v and a subtree (of T_2) rooted at u is equal to the largest weight in an optimal perfect matching M_{vu}.*

Proof. Every perfect matching of the graph G_{vu} corresponds to a bijective mapping between the partitions of the graph G_{vu}. Therefore, a perfect matching of the graph G_{vu} tell us how the subtrees rooted at $children[v]$ and subtrees rooted at $children[u]$, together with possible dummy vertices, are matched when building an optimal top-down amalgam of a subtree rooted at v and a subtree rooted at u. Every subtree rooted at some vertex from the set $children[v]$ is matched either with exactly one subtree rooted at some vertex from the set $children[u]$ or exactly one dummy vertex. Such a matching of optimal top-down amalgams induces an amalgam A of a subtree rooted at v and a subtree rooted at u. Since the weights of edges in the graph G_{vu} are the distances between the covers of the corresponding optimal top-down amalgams, the distance between the covers of the amalgam A is equal to the largest weight in a perfect matching.

Let M_{vu} be an optimal perfect matching of the graph G_{vu}. From the construction of the graph G_{vu} and facts stated above it follows that the distance between the covers of an optimal top-down amalgam is at most the largest weight in an optimal perfect matching M_{vu}. To prove that the equality holds true, suppose that the distance between the covers of an optimal top-down amalgam is smaller than the largest weight in an optimal perfect matching M_{vu}. Using the corresponding subtree isomorphism M of the optimal top-down common subtree we can construct the complete bipartite graph G'_{vu}, which has an optimal perfect matching with the largest weight that is smaller than the largest weight in M_{vu}, a contradiction with the construction of G_{vu}. □

Using Lemma 3.32, the distance between the covers of an optimal top-down amalgam is equal to

$$\min_{M \subset E(G_{vu})} \left(\max_{e \in M} w(e) \right),$$

where M is a perfect matching of the complete bipartite graph G_{vu} and $w(e)$ represents the weight of the edge e.

When all the recursive calls are completed, we return to the root vertices, and the largest weight of the optimal perfect matching $M_{r_1 u}$ is the distance between the covers of an optimal top-down amalgam of the rooted trees T_1 and T_2.

The procedure described uses the sub-procedure named `SolveOptimalPerfectMatching`, which finds a perfect matching of the complete bipartite graph G_{vu} that minimizes the value of the largest weight (an optimal perfect matching) and returns the value of that largest weight. For the sake of clarity, we briefly describe this sub-procedure.

Given a complete bipartite graph $G_{vu} = (V(G_{vu}), E(G_{vu}))$ with $|V(G_{vu})| = 2p$, we first sort the edges in the ascending order of the edge weights. Then take the induced subgraph G'_{vu} of the graph G_{vu} with the smallest p edges with respect to the weights. Find a maximum bipartite matching M_{vu} of the graph G'_{vu}, using the Hopcroft-Karp algorithm. If $|M_{vu}| = p$, then M_{vu} is the solution. Otherwise, add to G'_{vu} all the edges with the smallest weight that have not yet been added and repeat the search for a maximum bipartite matching. Since the graph G_{vu} is a finite

Procedure OptimalTopDownCommonSubtree(T_1,v,T_2,u,M')

input : Rooted tree T_1 and its root vertex v, rooted tree T_2 and its root vertex u, and the union set of solutions to the optimal perfect matching problems M'.

output: Distance between the subtree of T_1 rooted at v and subtree of T_2 rooted at u, and the union set of solutions to all optimal perfect matchings solved during the procedure saved in M'.

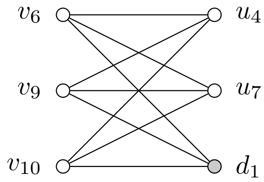
if `isLeaf(`T_1,v`)` **or** `isLeaf(`T_2,u`)` **then**
    return `max(height(`T_1,v`),height(`T_2,u`))`
Create the complete bipartite graph G_{vu} without edge weights.
foreach $e = xy \in G_{vu}$ **do**
    if x *is dummy vertex* **then**
        `weight(`e`)` $\leftarrow$ `height(`T_2,y`)`$+1$
    else if y *is dummy vertex* **then**
        `weight(`e`)` $\leftarrow$ `height(`T_1,x`)`$+1$
    else
        `weight(`e`)` $\leftarrow$ `OptimalTopDownCommonSubtree(`T_1,x,T_2,y,M'`)`
$distance \leftarrow$ `SolveOptimalPerfectMatching(`G_{vu},M_{vu}`)`
Remove edges incident with dummy vertices from M_{vu}.
$M' = M' \cup M_{vu}$
return $distance$

complete bipartite graph, sooner or later the found maximum bipartite matching will have cardinality p. In the end, return the largest weight of M_{vu}.

Let us take a look at an example of executing the procedure `OptimalTopDownCommonSubtree` on the input rooted trees T_1 (rooted at v_{11}) and T_2 (rooted at u_8), both depicted in Figure 3.8.

Example 3.33. *We start with tree T_1 rooted at v_{11} and tree T_2 rooted at u_8. Since none of the root vertices is a leaf, we build the complete bipartite graph $G_{v_{11}u_8}$ with the edge weights table shown on the right-hand side:*

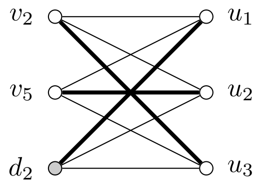
$G_{v_{11}u_8}$:



	u_4	u_7	d_1
v_6			3
v_9			3
v_{10}	1	2	1

We know the weights of the edges if one of the endpoints is a leaf or a dummy vertex. To determine the missing weights we have to proceed recursively down the trees.

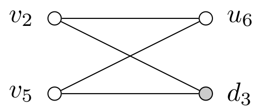
First, we want to determine the weight of the edge v_6u_4*. In order to find the optimal top-down common subtree of the subtree of* T_1 *rooted at* v_6 *and subtree of* T_2 *rooted at* u_4*, we construct the complete bipartite graph* $G_{v_6u_4}$*:*



	u_1	u_2	u_3
v_2	1	1	1
v_5	1	1	1
d_2	1	1	1

Since the vertices u_1, u_2 *and* u_3 *are leaves, all the weights are known. Therefore, we obtain an optimal perfect matching* $M_{v_6u_4} = \{(v_2, u_3), (v_5, u_2), (d_2, u_1)\}$ *of the complete bipartite graph* $G_{v_6u_4}$ *(drawn with bold edges and encircled weights). The largest weight in* $M_{v_6u_4}$ *is* 1*, therefore the weight of the edge* v_6u_4 *from graph* $G_{v_{11}u_8}$ *is* 1.

Next, we want to determine the weight of the edge v_6u_7 *of* $G_{v_{11}u_8}$*. In order to find the optimal top-down common subtree of the subtree of* T_1 *rooted at* v_6 *and the subtree of* T_2 *rooted at* u_7*, we construct the complete bipartite graph* $G_{v_6u_7}$*:*



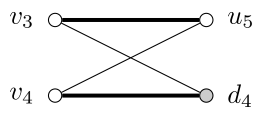
	u_6	d_3
v_2		2
v_5		2

For the weights of edges v_2u_6 *and* v_5u_6 *we have to find optimal top-down common subtrees of the following two pairs of rooted subtrees. The first pair with the subtree of* T_1 *rooted at* v_2 *and subtree of* T_2 *rooted at* u_6 *yields the trivial weighted complete bipartite graph* $G_{v_2u_6}$

with the optimal perfect matching $M_{v_2u_6} = \{(v_1, u_5)\}$*:*

	u_5
v_1	0

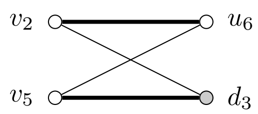
The second one with the subtree of T_1 *rooted at* v_5 *and subtree of* T_2 *rooted at* u_6 *yields the complete bipartite graph* $G_{v_5u_6}$ *with an optimal perfect matching* $M_{v_5u_6} = \{(v_3, u_5), (v_4, d_4)\}$*:*



	u_5	d_4
v_3	0	1
v_4	0	1

Therefore, the weights of edges v_2u_6 *and* v_5u_6 *from graph* $G_{v_6u_7}$ *are* 0 *and* 1*, respectively.*

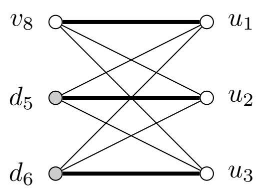
Now we have all the weights of the graph $G_{v_6u_7}$ *to find the optimal top-down common subtree of the subtree rooted at* v_6 *and the subtree rooted at* u_7*:*



	u_6	d_3
v_2	0	2
v_5	1	2

From the largest weight of an optimal perfect matching $M_{v_6u_7} = \{(v_2, u_6), (v_5, d_3)\}$ *it follows that the weight of the edge* v_6u_7 *from graph* $G_{v_{11}u_8}$ *is equal to* 2*.*

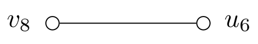
Proceeding in the same way, we have to determine the weight of the edge v_9u_4 *of the graph* $G_{v_{11}u_8}$*. In order to find the optimal top-down common subtree of the subtree of* T_1 *rooted at* v_9 *and the subtree of* T_2 *rooted at* u_4*, we construct the complete bipartite graph* $G_{v_9u_4}$ *with an optimal perfect matching* $M_{v_9u_4} = \{(v_8, u_1), (d_5, u_2), (d_6, u_3)\}$*:*



	u_1	u_2	u_3
v_8	1	1	1
d_5	1	1	1
d_6	1	1	1

The largest weight of the optimal perfect matching $M_{v_9u_4}$ is equal to 1, so the weight of the edge v_9u_4 from graph $G_{v_{11}u_8}$ is 1.

To get the last missing weight of the graph $G_{v_{11}u_8}$, namely the weight of the edge v_9u_7, we have to find an optimal top-down common subtree of the subtree of T_1 rooted at v_9 and the subtree of T_2 rooted at u_7. We get the trivial weighted complete bipartite graph $G_{v_9u_7}$:

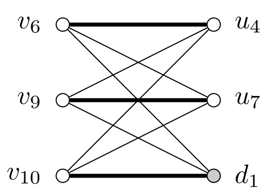


	u_6
v_8	

The perfect matching is trivial but we still need the weight of the edge v_8u_6. To determine the weight of the edge v_8u_6 we create another trivial complete bipartite graph $G_{v_8u_6}$ with optimal perfect matching $M_{v_8u_6} = \{(v_7, u_5)\}$:

	u_5
v_7	0

Since the largest weight of the matching $M_{v_8u_6}$ is equal to 0, so too is the largest weight of the previous trivial matching $M_{v_9u_7} = \{(v_8, u_6)\}$. Therefore, the weight of the edge v_9u_7 is equal to 0, and now we have all the weights to find the perfect matching of the complete bipartite graph $G_{v_{11}u_8}$:



	u_4	u_7	d_1
v_6	1	2	3
v_9	1	0	3
v_{10}	1	2	1

After finding an optimal perfect matching $M_{v_{11}u_8} = \{(v_6, u_4), (v_9, u_7), (v_{10}, d_1)\}$, we obtain the optimal top-down common subtree of the input rooted trees T_1 (rooted at v_{11}) and T_2 (rooted at u_8). The largest weight of an optimal perfect matching $M_{v_{11}u_8}$ is equal to 1, so the distance between the covers of the corresponding amalgam is equal to 1.

The procedure `ReconstructionOfMapping` is used to construct an actual optimal top-down common subtree isomorphism mapping M of the input rooted trees. The construction is based on Lemma 3.34. First, let us recall some properties of optimal perfect matchings.

At a fixed step during the procedure `OptimalTopDownCommonSubtree` we are in the vertices $v \in V(T_1)$ and $u \in V(T_2)$. Let $S_1 = (V(S_1), E(S_1))$ be the subtree of T_1 rooted at v, and $S_2 = (V(S_2), E(S_2))$ be the subtree of T_2 rooted at u. The solution to an optimal perfect matching M_{vu} of the complete bipartite graph G_{vu} is a set of weighted edges. Notice that the endpoints of those edges are from the vertex sets $V(S_1)$, $V(S_2)$ or dummy vertices D. If we remove from set M_{vu} all the edges with a dummy vertex as an endpoint, then we get a set of ordered pairs of vertices $M'_{vu} \subseteq V(S_1) \times V(S_2)$. Since $V(S_1) \subseteq V(T_1)$ and $V(S_2) \subseteq V(T_2)$ it follows that $M'_{vu} \subseteq V(T_1) \times V(T_2)$.

Lemma 3.34. *Let $T_1 = (V(T_1), E(T_1))$ and $T_2 = (V(T_2), E(T_2))$ be input rooted trees for the procedure* `OptimalTopDownCommonSubtree`*, and let $M' \subseteq V(T_1) \times V(T_2)$ be the union set of solutions to all optimal perfect matching problems solved during the procedure without the edges incident with dummy vertices. There is a unique optimal top-down common subtree isomorphism $M \subseteq V(T_1) \times V(T_2)$, such that $M \subseteq M'$.*

Proof. Let $T_1 = (V(T_1), E(T_1))$ and $T_2 = (V(T_2), E(T_2))$ be the input rooted trees for the procedure `OptimalTopDownCommonSubtree`, and let M' be the corresponding union set of solutions to optimal perfect matching problems without the edges incident with dummy vertices.

If we prove that for each non-root vertex $v \in V(T_1)$ with $(parent(v), z) \in M'$, for some vertex $z \in V(T_2)$, there is at most one pair $(v, w) \in M'$ such that $parent(w) = z$, then we can reconstruct the unique optimal top-down common subtree isomorphism $M \subseteq M'$ of T_1 and T_2 by following the order of non-decreasing depth of the vertices in the tree T_1. Namely, we start with adding the pair of the root vertices (r_1, r_2) to the isomorphism mapping M, and then on each step the parent of the vertex v is already mapped to a fixed vertex from $V(T_2)$. It follows that the mapping of the vertex v is determined.

Let $(v, w_1), (v, w_2) \in M'$ with $w_1 \neq w_2$. Suppose that vertices w_1 and w_2 are siblings. Both of them appear in the bipartite graph G_{pz} in the same partition set, where $p = parent(v)$. No two edges in a matching can share a common vertex. Therefore, only one pair, either (v, w_1) or (v, w_2), can be part of an optimal perfect matching of G_{pz}, a contradiction. It follows that the vertices w_1 and w_2 are not siblings. Therefore, $parent(w_1) \neq parent(w_2)$. $\square$

We will reconstruct an optimal top-down common subtree isomorphism mapping $M \subseteq V(T_1) \times V(T_2)$ from the set $M' \subseteq V(T_1) \times V(T_2)$ as follows. Begin with $M = \{(r_1, r_2)\}$, and for all the remaining vertices $v \in V(T_1)$ in pre-order traversal[1] of the tree T_1 add the pair (v, w) to the set M if it holds that $(v, w) \in M'$ and $(parent(v), parent(w)) \in M$.

Procedure ReconstructionOfMapping(T_1,r_1,r_2,M',M)

input : Rooted tree T_1 and its root vertex r_1, root vertex r_2 of T_2, the union set of solutions to the optimal perfect matching problems M' and mapping M.

output: Optimal top-down common subtree isomorphism mapping M from the subtree of T_1 rooted at r_1 to subtree of T_2 rooted at r_2 reconstructed from the union set of solutions to all optimal perfect matchings saved in M'.

$M \leftarrow M \cup (r_1, r_2)$
Let $P(T_1) = (v_1, \ldots, v_n)$ be the pre-order set of the vertex set $V(T_1)$.
for $i \leftarrow 1$ *to* n **do**
    foreach $(v_i, w) \in M'$ **do**
        if $((parent(v_i), parent(w)) \in M$ **then**
            $M \leftarrow M \cup (v_i, w)$

In Example 3.35, we continue Example 3.33 with the reconstruction of an optimal top-down common subtree isomorphism mapping M.

Example 3.35. *All the solutions to optimal perfect matching problems solved during the procedure are listed below.*

[1] In this case, pre-order traversal means that we start in the root vertex, and the parent vertices have to be visited before their child vertices. The visiting order of the children of a vertex is not important.

$$
\begin{aligned}
M_{v_6u_4} &= \{(v_2, u_3), (v_5, u_2), (d_2, u_1)\} \\
M_{v_2u_6} &= \{(v_1, u_5)\} \\
M_{v_5u_6} &= \{(v_3, u_5), (v_4, d_4)\} \\
M_{v_6u_7} &= \{(v_2, u_6), (v_5, d_3)\} \\
M_{v_9u_4} &= \{(v_8, u_1), (d_5, u_2), (d_6, u_3)\} \\
M_{v_8u_6} &= \{(v_7, u_5)\} \\
M_{v_9u_7} &= \{(v_8, u_6)\} \\
M_{v_{11}u_8} &= \{(v_6, u_4), (v_9, u_7), (v_{10}, d_1)\}
\end{aligned}
$$

The set $M' \subseteq V(T_1) \times V(T_2)$ is equal to the union of the above sets without the edges incident with dummy vertices. Therefore,

$$
\begin{aligned}
M' =\{&(v_1, u_5), \\
&(v_2, u_3), (v_2, u_6), \\
&(v_3, u_5), \\
&(v_5, u_2), \\
&(v_6, u_4), \\
&(v_7, u_5), \\
&(v_8, u_1), (v_8, u_6), \\
&(v_9, u_7)\}.
\end{aligned}
$$

We start with the mapping set $M = \{(v_{11}, u_8)\}$. Following the preorder traversal of T_1 rooted at v_{11}, we add $(v_6, u_4), (v_2, u_3), (v_5, u_2), (v_9, u_7), (v_8, u_6)$ and (v_7, u_5) to the set M. In Figure 3.10, there is the mapping of the optimal top-down common subtree of trees T_1 (rooted at v_{11}) and T_2 (rooted at u_8). The number above the arc represents the order of adding the pair to the mapping.

Finally, we have the following Theorem.

Theorem 3.36. *Algorithm 1 determines the Hausdorff distance between the input trees and finds the corresponding common subtree isomorphism M.*

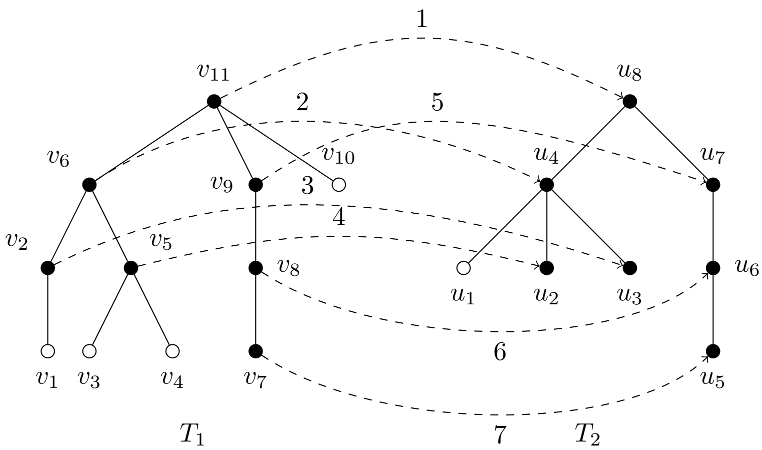


Figure 3.10: The mapping of the optimal top-down common subtree.

Proof. In the optimal top-down amalgam, the root vertices are always in the intersection of the amalgam. Therefore, we can root T_1 in a central vertex because of Theorem 3.21. For the root of T_2 we choose each vertex of the vertex set of T_2, making sure that one of the optimal top-down amalgams will coincide with an optimal amalgam of the input trees. The correctness of the Procedure `OptimalTopDownCommonSubtree` follows from Lemma 3.32, and the correctness of the Procedure `ReconstructionOfMapping` follows from Lemma 3.34.

□

In order to bound the time complexity of Algorithm 1, we need the time complexities of the procedures and sub-procedures used in the algorithm.

Lemma 3.37. *Let $T_1 = (V(T_1), E(T_1))$ and $T_2 = (V(T_2), E(T_2))$ be rooted input trees of the procedure* `OptimalTopDownCommonSubtree`*, and let G_{vu} be the complete bipartite graph on $2p$ vertices considered during the procedure. The sub-procedure of finding an optimal perfect matching of the graph G_{vu} runs in $\mathcal{O}\left(|V(T_1)| \cdot p^{\frac{5}{2}}\right)$.*

Proof. Graph G_{vu} has p^2 edges. First we sort all of the edges in $\mathcal{O}(p^2 \log(p^2))$ time. Then we take the first p edges with the smallest weights and run the Hopcroft-Karp algorithm for maximum bipartite matching. For a graph G, the Hopcroft-Karp algorithm runs in $\mathcal{O}(\sqrt{|V(G)|}|E(G)|)$ time [28]. In the worst case we have to repeat the Hopcroft-Karp algorithm $\mathcal{O}(|V(T_1)|)$ times, since there are at most

$|V(T_1)|$ different edge weights in the graph G_{vu} . This gives us the $\mathcal{O}\left(|V(T_1)| \cdot p^{\frac{5}{2}}\right)$ overall time complexity.

□

Lemma 3.38. *Let* $T_1 = (V(T_1), E(T_1))$ *and* $T_2 = (V(T_2), E(T_2))$ *be rooted input trees of the procedure* `OptimalTopDownCommonSubtree`*. The time complexity of the procedure* `OptimalTopDownCommonSubtree` *is bounded by* $\mathcal{O}\left(|V(T_1)|^2 \cdot |V(T_2)| \cdot \left(|V(T_1)|^{\frac{3}{2}} + |V(T_2)|^{\frac{3}{2}}\right)\right)$.

Proof. If one of the root vertices is a leaf, then the complexity of the procedure is constant. Therefore, the total effort spent on leaves is bounded by $\mathcal{O}\left(|V(T_1)| + |V(T_2)|\right)$.

If both of the root vertices are non-leaves, then the most (time) consuming part of the procedure is the sub-procedure `SolveOptimalPerfectMatching` and is bounded by time complexity $\mathcal{O}\left(|V(T_1)| \cdot p^{\frac{5}{2}}\right)$ due to Lemma 3.37, where $p = \max\{|children[v]|, |children[u]|\}$. In the remainder of the proof, we will denote $|children[v]|$ with $c(v)$. If we sum the time complexities of all the possible pairs of vertices such that one is from $V(T_1)$ and the other is from $V(T_2)$, we obtain an upper bound for the time complexity. Therefore, using the following equalities and inequalities

$$\begin{aligned}
&\sum_{v\in V(T_1),u\in V(T_2)} \max\left\{|V(T_1)|\cdot c(v)^{\frac{5}{2}}, |V(T_1)|\cdot c(u)^{\frac{5}{2}}\right\} \leq \\
&\leq \sum_{v\in V(T_1),u\in V(T_2)} \left(|V(T_1)|\cdot c(v)^{\frac{5}{2}} + |V(T_1)|\cdot c(u)^{\frac{5}{2}}\right) = \\
&= |V(T_1)|\cdot \sum_{v\in V(T_1)} \left(\sum_{u\in V(T_2)} c(v)^{\frac{5}{2}} + c(u)^{\frac{5}{2}}\right) = \\
&= |V(T_1)|\cdot \sum_{v\in V(T_1)} \left(\left(c(v)^{\frac{5}{2}} + c(u_1)^{\frac{5}{2}}\right) + \cdots + \left(c(v)^{\frac{5}{2}} + c(u_{|V(T_2)|})^{\frac{5}{2}}\right)\right) = \\
&= |V(T_1)|\cdot \sum_{v\in V(T_1)} \left(\left(|V(T_2)|\cdot c(v)^{\frac{5}{2}}\right) + \left(c(u_1)^{\frac{5}{2}} + \cdots + c(u_{|V(T_2)|})^{\frac{5}{2}}\right)\right) \leq \\
&\leq |V(T_1)|\cdot \sum_{v\in V(T_1)} \left(\left(|V(T_2)|\cdot c(v)^{\frac{5}{2}}\right) + \left(c(u_1) + \cdots + c(u_{|V(T_2)|})\right)^{\frac{5}{2}}\right) \leq \\
&\leq |V(T_1)|\cdot \sum_{v\in V(T_1)} \left(\left(|V(T_2)|\cdot c(v)^{\frac{5}{2}}\right) + |V(T_2)|^{\frac{5}{2}}\right) = \\
&= |V(T_1)|\cdot \Bigg(\left(|V(T_2)|\cdot c(v_1)^{\frac{5}{2}} + |V(T_2)|^{\frac{5}{2}}\right) + \cdots \\
&\qquad\qquad \cdots + \left(|V(T_2)|\cdot c(v_{|V(T_1)|})^{\frac{5}{2}} + |V(T_2)|^{\frac{5}{2}}\right)\Bigg) = \\
&= |V(T_1)|\cdot \Bigg(\left(|V(T_1)|\cdot |V(T_2)|^{\frac{5}{2}}\right) + \\
&\qquad\qquad + \left(\left(|V(T_2)|\cdot c(v_1)^{\frac{5}{2}}\right) + \cdots + \left(|V(T_2)|\cdot c(v_{|V(T_1)|})^{\frac{5}{2}}\right)\right)\Bigg) = \\
&= |V(T_1)|\cdot \left(\left(|V(T_1)|\cdot |V(T_2)|^{\frac{5}{2}}\right) + |V(T_2)|\cdot \left(c(v_1)^{\frac{5}{2}} + \cdots + c(v_{|V(T_1)|})^{\frac{5}{2}}\right)\right) \leq \\
&\leq |V(T_1)|\cdot \left(\left(|V(T_1)|\cdot |V(T_2)|^{\frac{5}{2}}\right) + |V(T_2)|\cdot \left(c(v_1) + \cdots + c(v_{|V(T_1)|})\right)^{\frac{5}{2}}\right) \leq \\
&\leq |V(T_1)|\cdot \left(\left(|V(T_1)|\cdot |V(T_2)|^{\frac{5}{2}}\right) + \left(|V(T_2)|\cdot |V(T_1)|^{\frac{5}{2}}\right)\right)
\end{aligned}$$

we discover that the total effort spent on non-leaves is bounded by

$$\mathcal{O}\left(|V(T_1)|^2\cdot |V(T_2)|\cdot \left(|V(T_1)|^{\frac{3}{2}} + |V(T_2)|^{\frac{3}{2}}\right)\right).$$

□

Theorem 3.39. *Let* $T_1 = (V(T_1), E(T_1))$ *and* $T_2 = (V(T_2), E(T_2))$ *be input trees of the Algorithm 1, where* $\mathrm{diam}(T_1) \geq \mathrm{diam}(T_2)$. *The time complexity of the Algorithm 1 is* *bounded by*

$$\mathcal{O}\left(|V(T_1)|^2 \cdot |V(T_2)|^2 \cdot \left(|V(T_1)|^{\frac{3}{2}} + |V(T_2)|^{\frac{3}{2}}\right)\right).$$

Proof. Since the procedure `ReconstructionOfMapping` runs in $\mathcal{O}(|V(T_1)| \cdot |V(T_2)|)$, it follows that the most expensive part of the Algorithm 1 is the for loop, which iterates through all the vertices of $V(T_2)$. At every iteration, the procedure `OptimalTopDownCommonSubtree` is called. Therefore, the time complexity of the Algorithm 1 is bounded by

$$\mathcal{O}\left(|V(T_2)| \cdot \left(|V(T_1)|^2 \cdot |V(T_2)| \cdot \left(|V(T_1)|^{\frac{3}{2}} + |V(T_2)|^{\frac{3}{2}}\right)\right)\right).$$

□

4

EDGE METRIC DIMENSION

Given a connected graph $G = (V(G), E(G))$ with at least two vertices, a vertex $v \in V(G)$, and an edge $e = uw \in E(G)$, the distance between the vertex v and the edge e is defined as $d_G(e, v) = \min\{d_G(u, v), d_G(w, v)\}$. A vertex $w \in V(G)$ *distinguishes* two edges $e_1, e_2 \in E(G)$ if $d_G(w, e_1) \neq d_G(w, e_2)$. A non-empty set S of vertices of a connected graph G is an *edge metric generator* of G if every two distinct edges of G are distinguished by some vertex of S. The smallest cardinality of an edge metric generator of G is called the *edge metric dimension* and is denoted with $\dim_{\mathrm{e}}(G)$. An *edge metric basis* of G is an edge metric generator of G of cardinality $\dim_{\mathrm{e}}(G)$.

The edge metric dimension was first introduced in 2015. The topic became popular and it was further investigated by several authors. In [54] a characterization of graphs achieving the upper bound for the edge metric dimension is done, and this piece of work showed that $\frac{\dim_{\mathrm{e}}(G)}{\dim(G)}$ is not bounded from above. The edge metric dimension of the Erdős-Rényi random graph G(n, p) is given in [55]. The authors of [53] independently characterize the graphs achieving the upper bound for the edge metric dimension using the graph complement. The exact formulae for the edge metric dimension of some generalized Petersen graphs are given in [38]. The edge metric dimension of the join, lexicographic and corona product of graphs is considered in [44].

In this Chapter, my original results from [36] are presented. We give some bounds for the edge metric dimension and determine formulae for the edge metric dimension of different families of graphs. We make a comparison between the edge metric dimension and the standard metric dimension of graphs. We present some realization results concerning the edge metric dimension and the standard metric dimension of graphs. We prove that computing the edge metric dimension of connected graphs is NP-hard and give an approximation algorithm for computing the edge metric dimension.

If we look at the problem of determining the edge metric dimension from different perspective, we see that it can be represented as a mathematical programming model. The model can be used to solve the problem of computing the edge metric dimension or finding an edge metric basis of a graph G. A similar model for the metric dimension was described in [13].

Let G be a graph of order n and size m with the vertex set $V(G) = \{v_1, v_2, \ldots, v_n\}$ and the edge set $E(G) = \{e_1, e_2, \ldots, e_m\}$. We consider the $n \times m$ dimensional matrix $D = [d_{ij}]$ such that $d_{ij} = d_G(v_i, e_j)$, where $v_i \in V(G)$ and $e_j \in E(G)$. Given the variables $y_i \in \{0, 1\}$, for $i \in \{1, 2, \ldots, n\}$, we define the following function:

$$\mathcal{F}(y_1, y_2, \ldots, y_n) = y_1 + y_2 + \cdots + y_n.$$

Minimizing the function $\mathcal{F}$ subject to the constraints

$$\sum_{i=1}^{n} |d_{ij} - d_{il}| y_i \geq 1, \quad \text{for every } 1 \leq j < l \leq m,$$

is equivalent to finding an edge metric basis of G. The solution for $y_1, y_2, \ldots, y_n$ represents a set of values for which the function $\mathcal{F}$ achieves the minimum value possible. This is equivalent to saying that the set $W = \{v_i \in V \mid y_i = 1\}$ is an edge metric basis of G. On the other hand, let W' be an edge metric basis of G and let $(y'_1, y'_2, \ldots, y'_n)$ be a vector, such that $y'_i = 0$ if $v_i \notin W'$ and $y'_i = 1$ if $v_i \in W'$, for any $i \in \{1, 2, \ldots, n\}$. The function $\mathcal{F}(y'_1, y'_2, \ldots, y'_n)$ gives a minimum subject to the constraints given before, otherwise there is a contradiction with W' being an edge metric basis.

4.1 Some Bounds and Closed Formulae

For any vertex v of a connected graph G, the set $V(G) \setminus \{v\}$ is an edge metric generator, since all the vertices from $V(G) \setminus \{v\}$ have a distance of 0 only to themselves while the vertex v does not have a distance of 0 to any of the vertices from $V(G) \setminus \{v\}$ and, therefore, any pair of edges is distinguished by at least one endpoint. Also, it is necessary to have at least one vertex in any edge metric generator. Thus, the natural bounds on the edge metric dimension of a graph are as follows.

Proposition 4.1. *For any connected graph G of order n,*

$$1 \leq \dim_{\mathrm{e}}(G) \leq n - 1. \tag{4.1}$$

The graphs achieving the equality in the lower bound above are relatively easy to deal with, being the same as for the standard metric dimension, namely paths P_n. The proof is given in Proposition 4.11. However, for the upper bound, characterizing all of the graphs that satisfy the equality is not as easy as for the standard metric dimension, where it is known that $\dim(G) = n - 1$ if and only if G is a complete graph.

First, let us present some partial results.

Proposition 4.2. *If G is a connected graph of order n and $\dim_{\mathrm{e}}(G) = n-1$, then for every $u, v \in V(G)$, $u \neq v$ it holds $N(u) \cap N(v) \neq \emptyset$.*

Proof. If there are two distinct vertices u and v, such that $N(u) \cap N(v) = \emptyset$, then we will show that $S = V(G) \setminus \{u, v\}$ is an edge metric generator. If e is an edge of G, we have the following options:

- If $e = xy$, where $x, y \in S$, then e has a distance of 0 to exactly two vertices is S, i.e. x and y.
- If $e = xu$ or $e = xv$, where $x \in S$, then e has a distance of 0 to just one vertex in S, namely x.
- If $e = uv$, then e has a distance of more than 0 to every vertex in S.

It is obvious that the two edges e and f can have the same distance to every vertex in S only when $e = xu$ and $f = xv$ for some vertex $x \in S$. But we assumed that $N(u) \cap N(v) = \emptyset$ and, therefore, this case cannot happen. Hence, S is an edge metric generator and $\dim_e(G) \leq n-2$, a contradiction. $\square$

Proposition 4.3. *Let G be a connected graph of order n. If there is a vertex $v \in V(G)$ of degree $n-1$, then either $\dim_e(G) = n-1$ or $\dim_e(G) = n-2$.*

Proof. Let x and y be distinct vertices, different from v and $S \subseteq V(G) \setminus \{x, y\}$. If $e = xv$ and $f = yv$, then $d(e, v) = d(f, v) = 0$ and $d(e, z) = d(f, z) = 1$, for every $z \in S \setminus \{v\}$. Therefore, S can not be an edge metric generator. It follows that any edge metric generator contains all vertices of G, except maybe v and one other vertex. Hence, $\dim_e(G) \geq n-2$. $\square$

Proposition 4.4. *Let G be a connected graph of order n. If there are two distinct vertices $u, v \in V(G)$ of degree $n-1$, then $\dim_e(G) = n-1$.*

Proof. We will show that every $S \subseteq V(G)$, which does not contain exactly two vertices of G, is not an edge metric generator. We consider two cases:

1. u and v are not in S: if $e = xu$ and $f = xv$, where $x \in S$, then e and f both have a distance of 0 to x and a distance of 1 to every other vertex in S.

2. At least one of the vertices u and v is in S: without loss of generality, assume that $v \in S$ and $S = V(G) \setminus \{x, y\}$, where $x, y \in V(G) \setminus \{v\}$. If $e = vy$ and $f = vx$, then e and f both have distances of 0 to v and distances of 1 to every other vertex in S.

In both cases, we can find two distinct edges with the same distance to every vertex in S. Therefore, S is not an edge metric generator. With this we have proved that $\dim_e(G) = n-1$. $\square$

We observe that there are graphs G of order n and maximum degree strictly less than $n-1$ for which $\dim_e(G) = n-1$. The circulant graph[1] $CR(6, 2)$ is a simple

[1]A circulant graph $CR(n, r)$ is a graph of order n with vertex set $V = \{v_0, v_1, \ldots, v_{n-1}\}$, such that v_i is adjacent to v_{i+j} with $j \in \{1, \ldots, r\}$, $i \in \{0, \ldots, n-1\}$ and the operation $i + j$ is done modulo n.

example of this, see Figure 4.1 for reference. Therefore, it is not only graph G of order n and maximum degree $n-1$ that satisfy that $\dim_e(G) = n-1$.

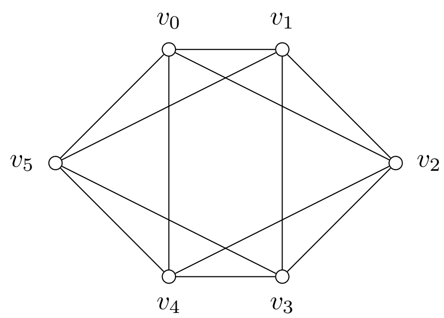


Figure 4.1: The circulant graph $CR(6,2)$.

Some other authors studied the graphs G achieving the upper bound $n-1$ for the edge metric dimension and they managed to characterize them. Zubrilina characterizes those graphs in [54]. Independently, Zhu et al. [53] make a characterization of graphs achieving the upper bound for the edge metric dimension using the graph complement. Using this characterization, they design an $\mathcal{O}(n^3)$ time algorithm, which determines whether a graph of order n has the edge metric dimension equal to $n-1$.

We now continue with several bounds on the edge metric dimension of connected graphs. Some of these general bounds are obtained by using the approach of the edge metric representation of edges with respect to an edge metric basis.

Proposition 4.5. *If G is a connected graph and $\Delta(G)$ is the maximum degree of G, then*

$$\dim_e(G) \geq \lceil \log_2 \Delta(G) \rceil .$$

Proof. From an arbitrary vertex $v \in V(G)$ there can be only two different distances to some set of incident edges. Therefore, to distinguish all edges that have the vertex u with $\deg u = \Delta(G)$ as one endpoint, it must hold that $2^{\dim_e(G)} \geq \Delta(G)$ and the assertion follows. □

Proposition 4.6. *If G is a connected graph and S is an edge metric basis with $|S| = k$, then S does not contain a vertex with a degree greater than 2^{k-1}.*

Proof. Suppose that an edge metric basis of cardinality k with a vertex v of a degree greater than 2^{k-1} exists. The edges incident with the vertex v all have an equal distance to v. So, there remain $k-1$ vertices to distinguish all those incident edges. Since from an arbitrary vertex $u \in V(G)$ there can be only two different distances to the set of incident edges, it follows that this is not an edge metric generator. We get a contradiction with our assumption, so all vertices in an edge metric basis are of a degree smaller or equal to 2^{k-1}. □

Proposition 4.7. *Let G be a connected graph. If $\dim_{\mathrm{e}}(G) = k$ and G has a diameter D, then $|E(G)| \leq (D+1)^k$.*

Proof. Since the diameter of the graph G equals D, the distance from an arbitrary vertex to an arbitrary edge in the graph G can have values from 0 to D. Therefore, an edge metric basis can distinguish at most $(D+1)^k$ edges, and therefore the graph G cannot have more edges. □

The next object of our study is the edge metric dimension of the hypercube graphs Q_n. To this end, we use a binary representation of Q_n. That is, the vertex set of Q_n consists of the 2^n-dimensional boolean vectors, *i.e.*, vectors with binary coordinates 0 or 1, and two vertices are adjacent whenever they differ in exactly one coordinate. It is known (see [22]) that in any n-dimensional hypercube, the set of vertices $B_n = \{11\ldots11, 01\ldots11, 10\ldots11, \ldots, 11\ldots01\}$ is a metric generator. We will prove that this set is also an edge metric generator for Q_n.

Theorem 4.8. *If n is a positive integer and Q_n is the n-dimensional hypercube, then $\dim_{\mathrm{e}}(Q_n) \leq n$.*

Proof. We will demonstrate that the set of n vertices $B_n = \{11\ldots11, 01\ldots11, 10\ldots11, \ldots, 11\ldots01\}$ is an edge metric generator. If $n = 1$, this result follows immediately. Therefore, we assume that $n > 1$. Let $e = uv$ and $f = xy$ be two different edges of Q_n. It suffices to prove that there exists $z \in B_n$ such that $d(e,z) \neq d(f,z)$. Suppose that this is not true. Thus, for every $z \in B_n$ it holds $d(e,z) = d(f,z)$. Of course, there is exactly one coordinate, let us say i, such that $u_i \neq v_i$, and there is exactly one coordinate, let us say j, such that $x_j \neq y_j$. Consider the following two cases.

1. $i \neq j$, and without loss of generality, let $i < j$:
   Let E be the number of coordinates $k \in \{1, 2, \ldots, n\} \setminus \{i, j\}$, such that $u_k = v_k = 0$. Furthermore, let F be the number of coordinates $k \in \{1, 2, \ldots, n\} \setminus \{i, j\}$, such that $x_k = y_k = 0$.

   - If $x_i = y_i = u_j = v_j = 0$ or $x_i = y_i = u_j = v_j = 1$, then since $d(e, 11\ldots11) = d(f, 11\ldots11)$, it follows that $E = F$. Let $z \in B_n$ be a vertex with $z_i = 0$. Thus, ($d(e, z) = E + 1$ and $d(f, z) = F$) or ($d(e, z) = E$ and $d(f, z) = F + 1$). Therefore, $d(e, z) \neq d(f, z)$, a contradiction.
   - If ($x_i = y_i = 0$ and $u_j = v_j = 1$) or ($x_i = y_i = 1$ and $u_j = v_j = 0$), then let $z \in B_n$ be a vertex with $z_i = 0$. Since $d(e, z) = d(f, z)$, it follows that $E = F$. Thus, ($d(e, 11\ldots11) = E + 1$ and $d(f, 11\ldots11) = F$) or ($d(e, 11\ldots11) = E$ and $d(f, 11\ldots11) = F+1$). Therefore, $d(e, 11\ldots11) \neq d(f, 11\ldots11)$, a contradiction.

2. $i = j$:
   In this case, let B_{n-1} be a metric generator for the hypercube Q_{n-1} as proved in [22]. Let u' be a vertex in Q_{n-1}, obtained by deleting i-th coordinate in the vertex u, and let x' be a vertex in Q_{n-1}, obtained by deleting i-th coordinate in the vertex x. Since the edges e and f are different, it follows that $u' \neq x'$. Also, for every $w \in B_{n-1}$ there is a $z_w \in B_n$, such that w is obtained from z_w by deleting the i-th coordinate. Since $d(u', w) = d(e, z_w) = d(f, z_w) = d(x', w)$ for every $w \in B_{n-1}$, we have $d(u', w) = d(x', w)$ for every $w \in B_{n-1}$. Since B_{n-1} is a metric generator in Q_{n-1}, this is a contradiction.

We have proved that for every two distinct edges e and f of the hypercube Q_n, it holds that there is $z \in B_n$, such that $d(e, z) \neq d(f, z)$. Therefore, B_n is an edge metric generator and the bound is obtained. □

4.2 Edge Metric Generators and Metric Generators

We have shown that the edge metric dimension of a graph G with order n is bounded by $1 \leq \dim_e(G) \leq n - 1$. Now, let discuss the existence of graphs with predetermined values for the edge metric dimension.

Proposition 4.9. *For two integers n, r, with $1 \leq r \leq n-1$, there exists a connected graph G of order n, such that $\dim_{\mathrm{e}}(G) = r$.*

Proof. If $r = n-1$ or $r = 1$, then we take the complete graph K_n or the path P_n, respectively. Otherwise ($2 \leq r \leq n-2$), we can easily check the positive answer by constructing a tree $T_{r,n}$ as follows. We begin with the star graph $S_{1,r}$. Then, we add a path with $n-r-1$ vertices and add an edge between exactly one leaf of the path and the center of the star $S_{1,r}$. It is straightforward to observe that such a tree $T_{r,n}$ has order n and edge metric dimension r. The r leaves of the star form an edge metric generator of $T_{r,n}$. On the other hand, the center of the star graph has degree $r+1$. Therefore, one has to take at least r vertices to an edge metric generator, otherwise there exist two distinct edges incident with the center of the star that are not distinguished. $\square$

Since the metric dimension and the edge metric dimension are closely related, another realization result is connected with considering them together. Therefore, we have the following question.

Question 4.10. *Given three integers r, t, n with $1 \leq r, t \leq n-1$: Is there a connected graph G of order n, such that $\dim(G) = r$ and $\dim_{\mathrm{e}}(G) = t$?*

In contrast to the first realizability question, the answer to this second question seems to be more difficult to find. One reason is based on the fact that for a graph G there is no clear relationship between $\dim(G)$ and $\dim_{\mathrm{e}}(G)$. Namely, it is possible to find graphs for which the metric dimension equals the edge metric dimension, as well as other graphs G for which $\dim(G) < \dim_{\mathrm{e}}(G)$ or $\dim_{\mathrm{e}}(G) < \dim(G)$. It is now our goal to explore such situations by comparing the values of $\dim(G)$ and $\dim_{\mathrm{e}}(G)$ for several families of connected graphs and to focus further on the realization question (Question 4.10).

4.2.1 Graphs for Which $\dim(G) = \dim_{\mathrm{e}}(G)$

The equality $\dim(G) = \dim_{\mathrm{e}}(G)$ holds true for several basic families of graphs. In some cases, obtaining the value of the edge metric dimension of a graph G is quite similar to computing the metric dimension of G. We begin this section with such classes of graphs, namely paths P_n, cycles C_n and complete graphs K_n.

Proposition 4.11. *For any integer* $n \geq 2$, $\dim_{\mathrm{e}}(P_n) = \dim(P_n) = 1$, $\dim_{\mathrm{e}}(C_n) = \dim(C_n) = 2$ *and* $\dim_{\mathrm{e}}(K_n) = \dim(K_n) = n-1$. *Moreover,* $\dim_{\mathrm{e}}(G) = 1$ *if and only if* G *is a path* P_n.

Proof. The metric dimension of paths, cycles and complete graphs is well-known, see [27] for reference.

First, to determine the edge metric dimension of a path take one endpoint of the path as an edge metric generator. The distances to all the edges are unique and therefore $\dim_{\mathrm{e}}(P_n) = 1$.

Second, taking only one vertex $v \in V(C_n)$ as an edge metric generator of a cycle is not enough. For example, the edges that are incident with the vertex v both have a distance of 0 to the vertex v. But if we take two vertices $S = \{u, v\}$ that are neighbours, then it is easy to verify that all the edges of C_n have different edge metric representations with respect to the set S.

Next, for complete graphs, let $S = V(K_n) \setminus \{u, v\}$ for any two distinct vertices $u, v \in V(K_n)$. Moving towards contradiction, suppose that S is an edge metric generator. Take an arbitrary vertex $w \in S$. The edges uw and vw have the same edge metric representations with respect to the set S, a contradiction. Therefore, an edge metric generator has a cardinality of at least $n-1$. Due to Proposition 4.1, this is an upper bound for the edge metric dimension, so the equality $\dim_{\mathrm{e}}(K_n) = n-1$ holds true.

To finish the proof we have to prove that if $\dim_{\mathrm{e}}(G) = 1$, then G is a path. Let $S = \{s\}$ be an edge metric basis of graph G. Since S is an edge metric generator, the vertex s has a degree of 1. If graph G has a cycle, then a vertex $u \in V(G)$ exists with a degree of at least 3, such that at least two edges incident with vertex u have a distance to the vertex s equal to $d_G(s, u)$. Therefore, G does not have a cycle. In other words, G is a tree. Suppose that there is a vertex $v \in V(G)$ with a degree of at least 3. We know that $v \neq s$. There is only one edge incident with v having distance d to the vertex s where all the other edges incident with v have distance $d+1$ to the vertex s. It follows that at least two edges with equal distances to the vertex s exist, a contradiction. Therefore, all the vertices in graph G have a degree of at most 2. Since G is a tree, we conclude that G is a path. $\square$

If $K_{r,t}$ is a complete bipartite graph different from $K_{1,1}$, then it is known that $\dim(K_{r,t}) = r + t - 2$ [11]. Next, we show that the same is true for the edge metric dimension.

Proposition 4.12. *For any complete bipartite graph $K_{r,t}$ different from $K_{1,1}$, $\dim_{\mathrm{e}}(K_{r,t}) = \dim(K_{r,t}) = r + t - 2$.*

Proof. Let V and U be the bipartition sets of $K_{r,t}$. To show that $\dim_{\mathrm{e}}(K_{r,t}) \geq r+t-2$, suppose that S is an edge metric generator without two elements of V, *i.e.* there are two distinct vertices $x, y \in V$, such that x and y are not in S. Let $u \in U$ and consider the edges $e = ux$ and $f = uy$. It follows that e and f have a distance of 0 to u and a distance of 1 to every other element in S. Therefore, S is not an edge metric generator, a contradiction. We proceed with the set U in a similar manner, and it follows that any edge metric generator must contain all but (maybe) one element of every partition set. Hence, $\dim_{\mathrm{e}}(K_{r,t}) \geq r + t - 2$.

On the contrary, take $v \in V$, $u \in U$ and let $S = V(K_{r,t}) \setminus \{v, u\}$. Edge uv is the only edge that has a distance of 1 to all vertices in S. Now, take any two distinct edges e, f of $K_{r,t}$ different from uv. There are at least two different endpoints of the edges e and f that are in the set S. Edges e and f are distinguished by at least one of those endpoints. Therefore, $\dim_{\mathrm{e}}(K_{r,t}) \leq r + t - 2$ and the equality follows. $\square$

Another family of graphs with equality on the values for metric dimension and edge metric dimension are trees. Since we already know that the edge metric dimension of a path is 1, we only consider trees that are not paths and compute the value of their edge metric dimensions. To this end, we need the following terminology from [32].

Let $T = (V(T), E(T))$ be a tree and let $v \in V(T)$. Define the equivalence relation R_v in the following way: for every two edges e, f, let eR_vf if and only if there is a path in T including e and f that does not have v as an internal vertex. The subgraphs induced by the edges of the equivalence classes of $E(T)$ are called the bridges of T

relative to v. Furthermore, for each vertex $v \in V$, the legs at v are the bridges that are paths. We use l_v to denote the number of legs at v.

We remark that the edge metric dimension of a tree can be computed in linear time. The algorithm to obtain an edge metric basis is the same as for the metric dimension (see [32]). For the sake of completeness, in the following proof we briefly describe the procedure.

Proposition 4.13. *Let $T = (V(T), E(T))$ be a tree. If T is not a path, then*

$$\dim_{\mathrm{e}}(T) = \dim(T) = \sum_{v \in V,\ l_v > 1} (l_v - 1).$$

Proof. Let v be a vertex of T, such that $l_v > 1$, and let S be an edge metric generator. Suppose that at least two of the v's legs do not contain an element of S. Then, the edges incident to v in those legs without an element of S have the same distance to every element of S, a contradiction. Therefore, at least $l_v - 1$ legs of v must contain an element of S. Since T is not a path, the vertices with a degree of 2 cannot have more than one leg. The legs corresponding to vertices with a degree of more than 2 are disjoint and, therefore, $\dim_{\mathrm{e}}(T) \geq \sum_{v \in V,\ l_v > 1}(l_v - 1)$.

On the contrary, we shall construct an edge metric generator S' for an arbitrary tree (which is not a path) in the following way:

- Compute l_v for each vertex v,
- For every vertex v with $l_v > 1$, put in the set S' all but one of the leaves associated with the legs of v.

As in [32], we will show that S' is an edge metric generator.

Root tree T at an arbitrary leaf r from the set S'. Let e and f be two arbitrary distinct edges from T. We will show that a vertex $s \in S'$ exists that distinguishes these two edges. With $lca(e, f)$ (the least common ancestor of the edges e and f) denote the vertex that lies on the path from r to e and on the path from r to f, and the distance $d_T(r, lca(e, f))$ is maximized.

Case 1: $d(r, e) \neq d(r, f)$. Vertex r distinguishes e and f.

Case 2: $d(r, e) = d(r, f)$ and at least one of the edges e or f has a descendant w with a degree greater than 2. Vertex w has a descendant from S', which distinguishes e and f.

Case 3: $d(r, e) = d(r, f)$ and none of the edges e or f has a descendant with a degree greater than 2. If the path from e to f has only one vertex w of a degree greater than 2 ($w = lca(e, f)$), then e and f are on different legs of w, and at least one of those two legs has a leaf in the set S', which distinguishes e and f. Otherwise, there is a vertex x on the path from e to f with a degree greater than 2 and is different from $w = lca(e, f)$. Notice that any vertex v of degree greater than 2 has a descendant from the set S' (a vertex with at least two legs exists in the subtree rooted at v). Vertex x has a descendant from the set S', which distinguishes e and f.

Therefore, S' is an edge metric generator. Inequality $\dim_{\mathrm{e}}(T) \leq \sum_{v \in V,\, l_v > 1}(l_v - 1)$ holds and the equality $\dim_{\mathrm{e}}(T) = \sum_{v \in V,\, l_v > 1}(l_v - 1)$ follows. Finally, since the same formula is used to calculate the metric dimension of a tree that is not a path (see [32]), the proof is completed. □

Next, we give the value of the edge metric dimension of the grid graph, which is the Cartesian product of two paths P_r and P_t with r and t vertices, respectively.

Proposition 4.14. *If G is the grid graph $G = P_r \square P_t$ with $r \geq t \geq 2$, then* $\dim_{\mathrm{e}}(G) = \dim(G) = 2$.

Proof. Since G is not a path, as stated in Proposition 4.11, it follows that $\dim_{\mathrm{e}}(G) \geq 2$. For easier computation of distances, let us embed G into $\mathbb{Z}^2$. Hence, each vertex can be represented as an ordered pair of its coordinates (x, y). We embed G into $\mathbb{Z}^2$ so that $(0, 0)$, $(r-1, 0)$, $(0, t-1)$, $(r-1, t-1)$ are the corner vertices of G. See Figure 4.2 for reference.

Let S be the set containing the two vertices $a = (0, 0)$ and $b = (r - 1, 0)$. We shall prove that S is an edge metric generator for the graph G. To this end, we notice that the distance between any two vertices in such a representation of G is $d((x_1, y_1), (x_2, y_2)) = |x_1 - x_2| + |y_1 - y_2|$. We assume that each edge is an unordered pair of its endpoints $e = (x_1, y_1)(x_2, y_2)$ and always transcribe such an edge in a

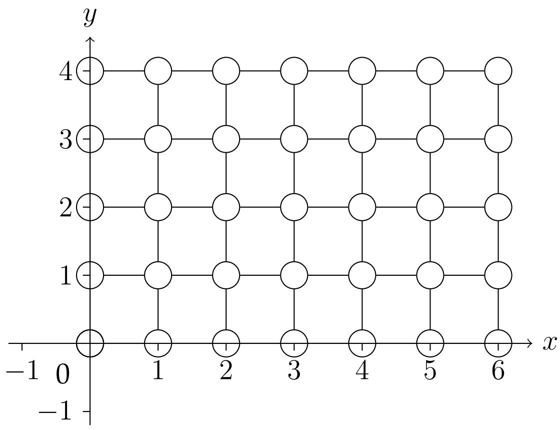


Figure 4.2: Embedding of a grid graph $G = P_7 \square P_5$ into $\mathbb{Z}^2$.

way that $x_1 \leq x_2$ and $y_1 \leq y_2$. This implies that the distances from the edge $e = (x_1, y_1)(x_2, y_2)$ to the vertices a and b are $d(a, e) = x_1 + y_1$ and $d(b, e) = r - 1 - x_2 + y_1$, respectively.

Toward contradiction, suppose that two different edges $e = (x_1, y_1)(x_2, y_2)$ and $f = (w_1, z_1)(w_2, z_2)$ exist with the same distances to the vertices a and b. This implies two equalities:

$$x_1 + y_1 = w_1 + z_1$$

$$r - 1 - x_2 + y_1 = r - 1 - w_2 + z_1 \iff y_1 - z_1 = x_2 - w_2.$$

Thus, it follows that $x_1 + x_2 = w_1 + w_2$. In both cases $x_1 = x_2$ or $x_1 = x_2 - 1$, we get $x_1 = w_1$ and $x_2 = w_2$. The equality $x_1 = w_1$ together with $x_1 + y_1 = w_1 + z_1$ implies that $y_1 = z_1$. Therefore, e and f have a common endpoint (x_1, y_1) and the second vertices have the same first coordinate. Both coordinates y_2 and z_2 can have values of either y_1 or $y_1 + 1$. Since they cannot have different values, it follows that $e = f$, which is a contradiction.

We already know from [32] that the metric dimension of grid graphs equals two. Thus, we finally get $\dim(G) = \dim_e(G)$ and the proof is completed. $\square$

4.2.2 Graphs for Which $\dim(G) < \dim_e(G)$

The *wheel graph* $W_{1,n}$ is isomorphic to $C_n \vee K_1$, where the operator $(\vee)$ represents the join graph. It is known (see [7]) that

$$\dim(W_{1,n}) = \begin{cases} 3, & \text{if } n = 3, 6, \\ 2, & \text{if } n = 4, 5, \\ \left\lfloor \frac{2n+2}{5} \right\rfloor, & \text{if } n \geq 6. \end{cases}$$

In the next proposition, we consider the edge metric dimension of wheel graphs and observe that it is strictly larger than the metric dimension, except in the case of $W_{1,3}$.

Proposition 4.15. *If* $W_{1,n}$ *is a wheel graph, then*

$$\dim_e(W_{1,n}) = \begin{cases} n, & \textit{if } n = 3, 4, \\ n-1, & \textit{if } n \geq 5. \end{cases}$$

Proof. If $n = 3$ or $n = 4$, then the proof is straightforward. Let $n \geq 5$ and $V(W_{1,n}) = \{x, g_1, g_2, \ldots, g_n\}$, where the vertex x has a degree of n and the vertices $g_1, \ldots, g_n$ induce a cycle C_n. Set $S = \{g_1, g_2, \ldots, g_{n-1}\}$. We prove that S is an edge metric generator. Let e be an edge of $W_{1,n}$. Consider the following cases:

- If $e = g_i g_{i+1}$, for some $i \in \{1, \ldots, n-2\}$, then e has a distance of 0 to g_i and g_{i+1}, and a distance of 1 or 2 to every other vertex in S.
- If $e = g_{n-1} g_n$, then e has a distance of 0 to g_{n-1}, a distance of 1 to g_1 and g_{n-2}, and a distance of 2 to every other vertex in S (and since $n \geq 5$ there is at least one such vertex).
- If $e = g_n g_1$, then e has a distance of 0 to g_1, a distance of 1 to g_{n-1} and g_2, and a distance of 2 to every other vertex in S (and since $n \geq 5$ there is at least one such vertex).
- If $e = x g_i$, for some $i \in \{1, \ldots, n-1\}$, then e has a distance of 0 to g_i and a distance of 1 to every other vertex in S.

- If $e = xg_n$, then e has a distance of 1 to every vertex in S.

Now we count the repetitions of the digits $0, 1, 2$ and their positions in the edge metric representations for the items above in order to check that the edge metric representations of any two distinct edges of $W_{1,n}$ are different. Thus, S is an edge metric generator and, therefore, $\dim_e(W_{1,n}) \leq n - 1$.

On the other hand, assume that S is a set of vertices without at least two distinct vertices g_i, g_j of the set $\{g_1, \ldots, g_n\}$ and that S is an edge metric generator of graph $W_{1,n}$. Consider the edges $e = xg_i$ and $f = xg_j$. Notice that e and f have the same distance to every vertex in S and so, S is not an edge metric generator. Therefore, $\dim_e(W_{1,n}) \geq n - 1$ and we are done. □

Similarly to the wheel graph, the *fan graph* $F_{1,n}$ is isomorphic to $P_n \vee K_1$. For the case of fan graphs, it is known (see [11]) that

$$\dim(F_{1,n}) = \begin{cases} 1, & \text{if } n = 1, \\ 2, & \text{if } n = 2, 3, \\ 3, & \text{if } n = 6, \\ \left\lfloor \frac{2n+2}{5} \right\rfloor, & \text{otherwise.} \end{cases}$$

Using an analogous procedure, as in the case of the wheel graphs, we can determine the edge metric dimension of fan graphs, which is again strictly larger than the metric dimension, with the exception of $F_{1,n}$ with $n \in \{1, 2\}$.

Proposition 4.16. *If $F_{1,n}$ is a fan graph, then*

$$\dim_e(F_{1,n}) = \begin{cases} n, & \textit{if } n = 1, 2, 3, \\ n - 1, & \textit{if } n \geq 4. \end{cases}$$

Proof. If $n \in \{1, 2, 3\}$, then the proof is straightforward. Let $n \geq 4$ and $V(F_{1,n}) = \{x, g_1, g_2, \ldots, g_n\}$, where vertex x has a degree of n and vertices $g_1 \ldots g_n$ induce a path P_n. Set $S = \{g_1, g_2, \ldots, g_{n-1}\}$. We shall show that S is an edge metric generator of $F_{1,n}$. Let e be an edge of $F_{1,n}$. Consider the following cases:

- If $e = g_i g_{i+1}$, for some $i \in \{1, \ldots, n - 2\}$, then e has a distance of 0 to g_i and g_{i+1}, and a distance of 1 or 2 to every other vertex in S.

- If $e = g_{n-1}g_n$, then e has a distance of 0 to g_{n-1}, a distance of 1 to g_{n-2}, and a distance of 2 to every other vertex in S (and since $n \geq 4$ there is at least one such vertex).
- If $e = xg_i$, for some $i \in \{1, \ldots, n-1\}$, then e has a distance of 0 to g_i and a distance of 1 to every other vertex in S.
- If $e = xg_n$, then e has a distance of 1 to every vertex in S.

We count the repetitions of the digits $0, 1, 2$ and their positions in the edge metric representations for the items above in order to check that the edge metric representations of any two distinct edges of $F_{1,n}$ are different. Thus, S is an edge metric generator and so, $\dim_e(F_{1,n}) \leq n - 1$.

For the opposite, assume that S is a set of vertices without at least two distinct vertices of the set $\{g_1, \ldots, g_n\}$, say g_i, g_j and that S is an edge metric generator of the graph $F_{1,n}$. Consider the edges $e = xg_i$ and $f = xg_j$. Clearly, e and f have a distance of 1 to every vertex in S. Thus, S is not an edge metric generator, which is a contradiction. Therefore, $\dim_e(W_{1,n}) \geq n - 1$ and the equality $\dim_e(F_{1,n}) = n - 1$ holds for $n \geq 4$. $\square$

4.2.3 Graphs for Which $\dim_e(G) < \dim(G)$

According to the definition of layers in the Cartesian product of two graphs given in Chapter 2, we say that an edge $e \in E(G \Box H)$ is *vertical*, if e lies in a ${}^gH$-layer for some $g \in V(G)$. Similarly, $e \in E(G \Box H)$ is *horizontal*, if e lies in an ${}^hG$-layer for some $h \in V(H)$.

The value of the metric dimension of several families of Cartesian product graphs was obtained in [12]. For instance, they proved that

$$\dim(C_r \Box C_t) = \begin{cases} 4, & \text{if } r \text{ and } t \text{ are even,} \\ 3, & \text{otherwise.} \end{cases}$$

Next we determine that for some particular cases of the torus graphs $C_r \Box C_t$, it follows that $\dim_e(C_r \Box C_t) < \dim(C_r \Box C_t)$.

Theorem 4.17. *For any pair of positive integers r, t, $\dim_e(C_{4r}\square C_{4t}) = 3$.*

Proof. We assume that $V(C_{4r}) = \{a_0, a_1, \ldots, a_{4r-1}\}$ and $V(C_{4t}) = \{b_0, b_1, \ldots, b_{4t-1}\}$, and for short let $G = C_{4r}\square C_{4t}$. From now on, in this proof, all the operations with the indices of the vertices of C_{4r} and C_{4t} are done modulo $4r$ and $4t$, respectively. Moreover, we assume that $a_ia_{i+1} \in E(C_{4r})$ and $b_jb_{j+1} \in E(C_{4t})$ for every $i \in \{0, \ldots, 4r-1\}$ and $j \in \{0, \ldots, 4t-1\}$, respectively. We shall prove that the set $S = \{(a_0, b_0), (a_0, b_{2t}), (a_r, b_t)\}$ is an edge metric generator of G. Let e, f be distinct edges of G. We consider the following cases.

Case 1: e is a horizontal edge and f is a vertical edge.
Without loss of generality, assume that the edges $e = (g_1, h)(g_2, h)$ and $f = (g, h_1)(g, h_2)$ satisfy that g_1 is closer to a_0 than g_2 and that h_1 is closer to b_0 than h_2. Thus, we have the following:

$$d_G(e, (a_0, b_0)) = d_{C_{4t}}(h, b_0) + d_{C_{4r}}(g_1, a_0),$$

$$d_G(e, (a_0, b_{2t})) = d_{C_{4t}}(h, b_{2t}) + d_{C_{4r}}(g_1, a_0),$$

$$d_G(f, (a_0, b_0)) = d_{C_{4t}}(h_1, b_0) + d_{C_{4r}}(g, a_0),$$

$$d_G(f, (a_0, b_{2t})) = d_{C_{4t}}(h_2, b_{2t}) + d_{C_{4r}}(g, a_0).$$

Suppose, $d_G(e, (a_0, b_0)) = d_G(f, (a_0, b_0))$ and $d_G(e, (a_0, b_{2t})) = d_G(f, (a_0, b_{2t}))$. From the equalities above we obtain

$$d_{C_{4t}}(h, b_0) + d_{C_{4r}}(g_1, a_0) = d_{C_{4t}}(h_1, b_0) + d_{C_{4r}}(g, a_0)$$

and

$$d_{C_{4t}}(h, b_{2t}) + d_{C_{4r}}(g_1, a_0) = d_{C_{4t}}(h_2, b_{2t}) + d_{C_{4r}}(g, a_0).$$

Since $d_{C_{4t}}(h, b_0)+d_{C_{4t}}(h, b_{2t}) = 2t$ and $d_{C_{4t}}(h_1, b_0)+d_{C_{4t}}(h_2, b_{2t}) = 2t-1$, by summing the last two equalities we deduce that

$$2d_{C_{4r}}(g_1, a_0) = 2d_{C_{4r}}(g, a_0) - 1,$$

which is not possible, since the left side of the equality is an even number and the right side is odd. Thus, we have that $d_G(e,(a_0,b_0)) \neq d_G(f,(a_0,b_0))$ or $d_G(e,(a_0,b_{2t})) \neq d_G(f,(a_0,b_{2t}))$. Equivalently, e, f are distinguished by (a_0,b_0) or by (a_0,b_{2t}).

Case 2: e, f are vertical edges.
Similarly to the case above, without loss of generality, we assume that the edges $e = (x,h_1)(x,h_2)$ and $f = (y,h_3)(y,h_4)$ satisfy that h_1 is closer to b_0 than h_2 and that h_3 is closer to b_0 than h_4. Thus, we have the following:

$$d_G(e,(a_0,b_0)) = d_{C_{4t}}(h_1,b_0) + d_{C_{4r}}(x,a_0),$$

$$d_G(e,(a_0,b_{2t})) = d_{C_{4t}}(h_2,b_{2t}) + d_{C_{4r}}(x,a_0),$$

$$d_G(f,(a_0,b_0)) = d_{C_{4t}}(h_3,b_0) + d_{C_{4r}}(y,a_0),$$

$$d_G(f,(a_0,b_{2t})) = d_{C_{4t}}(h_4,b_{2t}) + d_{C_{4r}}(y,a_0).$$

Now, assume that $d_G(e,(a_0,b_0)) = d_G(f,(a_0,b_0))$ and $d_G(e,(a_0,b_{2t})) = d_G(f,(a_0,b_{2t}))$. Thus, the four equalities above lead to

$$d_{C_{4t}}(h_1,b_0) + d_{C_{4r}}(x,a_0) = d_{C_{4t}}(h_3,b_0) + d_{C_{4r}}(y,a_0), \tag{4.2}$$

$$d_{C_{4t}}(h_2,b_{2t}) + d_{C_{4r}}(x,a_0) = d_{C_{4t}}(h_4,b_{2t}) + d_{C_{4r}}(y,a_0). \tag{4.3}$$

By summing these two equalities and by utilizing the fact that $d_{C_{4t}}(h_1,b_0) + d_{C_{4t}}(h_2,b_{2t}) = 2t-1$ and $d_{C_{4t}}(h_3,b_0) + d_{C_{4t}}(h_4,b_{2t}) = 2t-1$, we deduce that

$$d_{C_{4r}}(x,a_0) = d_{C_{4r}}(y,a_0).$$

Moreover, by using this equality in the equalities (4.2) and (4.3), it follows that

$$d_{C_{4t}}(h_1,b_0) = d_{C_{4t}}(h_3,b_0),$$

$$d_{C_{4t}}(h_2,b_{2t}) = d_{C_{4t}}(h_4,b_{2t}).$$

As a consequence of these last three relationships, we notice that any two edges e and f that have the same distance to the vertices (a_0,b_0) and (a_0,b_{2t}) satisfy one of the following situations:

- e, f are symmetrical with respect to the ${}^{a_0}C_{4t}$-layer (see pairs of edges (e_i, f_i), with $i \in \{1, \ldots, 4\}$, drawn in Figure 4.3),
- e, f are symmetrical with respect to the $C_{4r}{}^{b_0}$-layer or equivalently to the $C_{4r}{}^{b_{2t}}$-layer (see pairs of edges (e_1, e_4), (e_2, e_3), (f_1, f_4), (f_2, f_3), drawn in Figure 4.3),
- e, f are symmetrical with respect to the vertex (a_0, b_0) or equivalently to the vertex (a_0, b_{2t}) (see pairs of edges (e_1, f_4), (e_2, f_3), (f_1, e_4), (f_2, e_3) drawn in Figure 4.3).

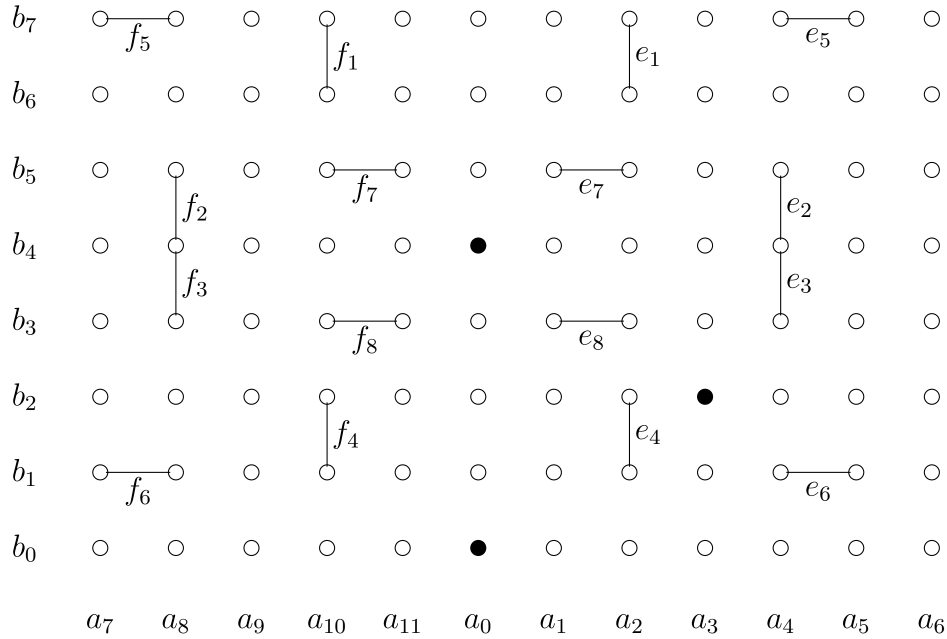


Figure 4.3: A sketch of the graph $C_{12}\Box C_8$. Only some of the edges are drawn. Vertices in bold represent an edge metric generator.

According to these items above and because of the fact that the cycles used to generate the graph G have an order of $4r$ and $4t$, it is not difficult to notice that if two vertical edges are not distinguished by the vertices (a_0, b_0) and (a_0, b_{2t}), then they are distinguished by the vertex (a_r, b_t). For instance, assume that e, f are symmetrical with respect to the ${}^{a_0}C_{4t}$-layer. Without loss of generality, assume that e lies in a ${}^{a_i}C_{4t}$-layer with $i \in \{1, \ldots, 2r-1\}$. Thus, f lies in a ${}^{a_j}C_{4t}$-layer with $j \in \{2r+1, \ldots, 4r-1\}$ (notice that neither e nor f lie in the ${}^{a_{2r}}C_{4t}$-layer, since in such a case $e = f$, which is not possible). Hence, it follows that

$$d_G(e,(a_r,b_t)) = d_G(e,(a_i,b_t)) + d_{C_{4r}}(a_i,a_r) \tag{4.4}$$

and

$$d_G(f,(a_r,b_t)) = d_G(f,(a_j,b_t)) + d_{C_{4r}}(a_j,a_r). \tag{4.5}$$

Note that $d_G(e,(a_i,b_t)) = d_G(f,(a_j,b_t))$, since e, f are symmetrical with respect to the ${}^{a_0}C_{4t}$-layer. Moreover, it is clear that $d_{C_{4r}}(a_i,a_r) < d_{C_{4r}}(a_j,a_r)$ happens, since $d_{C_{4r}}(a_i,a_0) = d_{C_{4r}}(a_j,a_0)$. Thus, the equalities given in (4.4) and (4.5) lead to $d_G(e,(a_r,b_t)) \neq d_G(f,(a_r,b_t))$.

Case 3: e, f are horizontal edges.
The procedure in this case is relatively similar to that in Case 2. As such, we assume that the edges $e = (g_1,y)(g_2,y)$ and $f = (g_3,z)(g_4,z)$ satisfy that g_1 is closer to a_0 than g_2 and that g_3 is closer to a_0 than g_4. Thus,

$$d_G(e,(a_0,b_0)) = d_{C_{4t}}(y,b_0) + d_{C_{4r}}(g_1,a_0),$$

$$d_G(e,(a_0,b_{2t})) = d_{C_{4t}}(y,b_{2t}) + d_{C_{4r}}(g_1,a_0),$$

$$d_G(f,(a_0,b_0)) = d_{C_{4t}}(z,b_0) + d_{C_{4r}}(g_3,a_0),$$

$$d_G(f,(a_0,b_{2t})) = d_{C_{4t}}(z,b_{2t}) + d_{C_{4r}}(g_3,a_0).$$

As before, we assume that $d_G(e,(a_0,b_0)) = d_G(f,(a_0,b_0))$ and $d_G(e,(a_0,b_{2t})) = d_G(f,(a_0,b_{2t}))$. Thus, the four equalities above lead to

$$d_{C_{4t}}(y,b_0) + d_{C_{4r}}(g_1,a_0) = d_{C_{4t}}(z,b_0) + d_{C_{4r}}(g_3,a_0), \tag{4.6}$$

$$d_{C_{4t}}(y,b_{2t}) + d_{C_{4r}}(g_1,a_0) = d_{C_{4t}}(z,b_{2t}) + d_{C_{4r}}(g_3,a_0). \tag{4.7}$$

By summing these two equalities and by using the fact that $d_{C_{4t}}(y,b_0)+d_{C_{4t}}(y,b_{2t}) = 2t$ and $d_{C_{4t}}(z,b_0) + d_{C_{4t}}(z,b_{2t}) = 2t$, we deduce that

$$d_{C_{4r}}(g_1,a_0) = d_{C_{4r}}(g_3,a_0).$$

Also, by using the equality above in the equalities (4.6) and (4.7), we find that

$$d_{C_{4t}}(y, b_0) = d_{C_{4t}}(z, b_0),$$

$$d_{C_{4t}}(y, b_{2t}) = d_{C_{4t}}(z, b_{2t}).$$

Thus, we deduce that for any two edges e, f that have the same distance to the vertices (a_0, b_0) and (a_0, b_{2t}) one of the following situations is satisfied:

- e, f are symmetrical with respect to the ${}^{a_0}C_{4t}$-layer (see pairs of edges (e_i, f_i), with $i \in \{5, \dots, 8\}$, drawn in Figure 4.3),
- e, f are symmetrical with respect to the $C_{4r}{}^{b_0}$-layer or equivalently to the $C_{4r}{}^{b_{2t}}$-layer (see pairs of edges (e_5, e_6), (e_7, e_8), (f_5, f_6) and (f_7, f_8) drawn in Figure 4.3),
- e, f are symmetrical with respect to the vertex (a_0, b_0) or equivalently to the vertex (a_0, b_{2t}) (see pairs of edges (e_5, f_6), (e_6, f_5), (e_7, f_8) and (e_8, f_7) drawn in Figure 4.3).

By using a similar reasoning as in Case 2, we deduce that if two horizontal edges are not distinguished by the vertices (a_0, b_0) and (a_0, b_{2t}), they are distinguished by the vertex (a_r, b_t).

As a consequence of the three cases above, we gather that S is an edge metric generator, which leads to $\dim_e(C_{4r}\square C_{4t}) \leq 3$. Now, consider two distinct vertices $(a, b), (c, d) \in V(C_{4r}\square C_{4t})$. Notice that there are always two incident edges with (a, b) (or with (c, d)), such that they are not distinguished by (a, b) nor by (c, d). Therefore, $\dim_e(C_{4r}\square C_{4t}) > 2$, which completes the proof. $\square$

An infinite family of graphs exists, all with the edge metric dimension smaller than the metric dimension. A natural question that arises is the following one.

Problem 4.18. *Are there any other families of graph (different from the torus graph $C_{4r}\square C_{4t}$) such that $\dim_e(G) < \dim(G)$?*

4.2.4 Realization of the Edge Metric Dimension Versus the Metric Dimension

Since it is possible to find classes of graphs G such that $\dim(G) = \dim_e(G)$, $\dim(G) < \dim_e(G)$ or $\dim_e(G) < \dim(G)$, Question 4.10 stated at the beginning of this chapter (concerning the triplet r, t, n: metric dimension, edge metric dimension and order, respectively) must be dealt with by separating these three possibilities.

The case $\dim(G) = \dim_e(G)$ can be realized through complete or tree graphs, for instance. That is, the triplet $n-1, n-1, n$ can be realized with a complete graph K_n and the triplet r, r, n with $1 \leq r \leq n-2$ can be realized with a tree T with $r+1$ leaves obtained from a star $S_{1,n-1}$ by removing $n-1-r$ edges of $S_{1,n-1}$ and subdividing one of the remaining edges with $n-1-r$ vertices. Clearly, the order of T is n, and by Proposition 4.13 $\dim(T) = \dim_e(T) = r$. Notice that the particular case $r = 1$ is given by the path graph P_n.

Next, we continue with the case $\dim(G) < \dim_e(G)$. To this end, we need the following family $\mathcal{F}$ of graphs. Let $a \geq 1$, $b \geq 2$ and $c \geq 0$ be arbitrary integers. We begin with a star graph $S_{1,b}$, where $b \geq 2$, and the graph $G_1 = K_1 \vee (\bigcup_{i=1}^{a} K_2)$, $a \geq 1$. To obtain a graph $G_{a,b,c} \in \mathcal{F}$, we choose a path P_c of order c and join with an edge one leaf of P_c with the center of G_1, and the other leaf with the center of the star $S_{1,b}$. If $c = 1$, then P_c is a trivial graph with only one vertex x. In this case, use an edge to join vertex x with the center of G_1 and vertex x with the center of the star $S_{1,b}$. We shall make the assumption that c could be equal to zero, and in this case the action above (adding the path P_c) is understood as adding an edge between the centers of G_1 and $S_{1,b}$. See Figure 4.4 for an example.

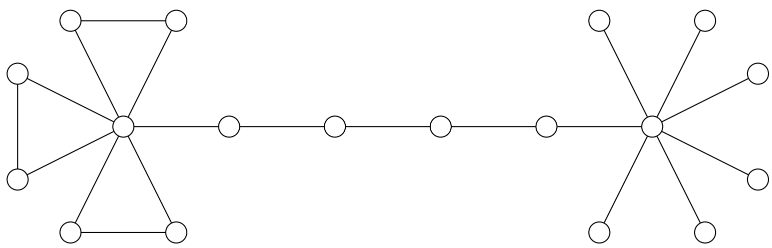


Figure 4.4: The graph $G_{3,6,4}$.

Observe that a graph $G_{a,b,c} \in \mathcal{F}$ has an order of $2a+b+c+2$. Next we compute $\dim(G_{a,b,c})$ and $\dim_e(G_{a,b,c})$ for any $G_{a,b,c} \in \mathcal{F}$.

Remark 4.19. *If $G_{a,b,c} \in \mathcal{F}$, then* $\dim(G_{a,b,c}) = a + b - 1$ *and* $\dim_e(G_{a,b,c}) = 2a + b - 2$.

Proof. Let S be a metric basis of $G_{a,b,c}$. Notice that any two distinct leaves of the star $S_{1,b}$ have the same distance to any other vertex of $G_{a,b,c}$. Moreover, any two adjacent vertices of G_1 different from the center have the same distance to any other vertex of $G_{a,b,c}$. As a consequence of these two observations, we deduce that S must contain at least $b - 1$ vertices of the star $S_{1,b}$ and at least a vertices of G_1. Thus, $\dim(G_{a,b,c}) \geq a + b - 1$. On the other hand, it is straightforward to observe that a set composed by $b - 1$ leaves of the star $S_{1,b}$ and one vertex of each graph K_2 used to generate G_1 is a metric generator of $G_{a,b,c}$. Therefore, $\dim(G_{a,b,c}) \leq a + b - 1$ and the first equality follows.

Now, let S' be an edge metric basis of $G_{a,b,c}$. We observe that any two edges from G_1 incident with the center of G_1 have the same distance to every other vertex of $G_{a,b,c}$. Also, any two edges of the star $S_{1,b}$ have the same distance to every other vertex of $G_{a,b,c}$. Thus, we deduce that S' must contain at least $b - 1$ vertices of the star $S_{1,b}$ and $2a - 1$ vertices of G_1. Thus, $\dim_e(G_{a,b,c}) \geq 2a + b - 2$. It is again straightforward to observe that a set composed by $b-1$ leaves of the star $S_{1,b}$ and all but two vertices of G_1 (the center and one other vertex) is an edge metric generator of $G_{a,b,c}$. Therefore, $\dim_e(G_{a,b,c}) \leq 2a + b - 2$ and the second equality follows. □

By using the family above we partially solve the realization question regarding the triplet $\dim(G)$, $\dim_e(G)$ and the order of G, whenever $\dim(G) < \dim_e(G)$. We first observe that the triplet $1, t, n$, with $t \geq 2$, is not realizable for any graph G, since $\dim(G) = 1$ if and only if G is a path P_n and $\dim_e(P_n) = 1$. In the next theorem, we consider that $2r \leq n - 2$.

Theorem 4.20. *For any r, t, n such that $2 \leq r < t \leq 2r \leq n - 2$, there exists a connected graph G of order n such that* $\dim(G) = r$ *and* $\dim_e(G) = t$.

Proof. We first deal with the case $t = 2r$. Let $G_{r,n}$ be the graph obtained as follows. We begin with the join graph $G' = K_1 \vee (K_1 \cup (\bigcup_{i=1}^{r} K_2))$. Then, we add a path of order $n - 2r - 2$ and add an edge between one of its leaves with the unique vertex of G' of degree one. See Figure 4.5 for an example. Clearly, $G_{r,n}$ has an order of $n - 2r - 2 + 2r + 2 = n$.

Let S be a metric basis of $G_{r,n}$. Any two adjacent vertices of G' different from the center and the vertex with a degree of one have the same distance to any other vertex of $G_{r,n}$. Thus, S must contain at least r vertices of G' and $\dim(G_{r,n}) \geq r$. On the other hand, a set composed by one vertex of each graph K_2 used to generate G' is a metric generator of $G_{r,n}$. Therefore, $\dim(G_{r,n}) \leq r$, and using both inequalities it follows that $\dim(G_{r,n}) = r$.

Now, let S' be an edge metric basis of $G_{r,n}$. We observe that any two edges from G' incident with the center of G', except the edge connecting the center of G' and the vertex with a degree of one, have the same distance to every other vertex of $G_{r,n}$. Thus, we deduce that S' must contain at least $2r-1$ vertices of G'. Since there is still one pair of edges that is not distinguished, namely both edges of G' that do not have an endpoint in the set S, we have to add a vertex to the set S'. Thus, $\dim_e(G_{r,n}) \geq 2r$. On the other hand, a set composed by all but two vertices of G' (the center and one other vertex) is an edge metric generator of $G_{r,n}$. Therefore, $\dim_e(G_{r,n}) \leq 2r$, and using both inequalities it follows that $\dim_e(G_{r,n}) = 2r = t$.

Summing up the above, $\dim(G_{r,n}) = r$ and $\dim_e(G_{r,n}) = t$. Since $n - 2r - 2 \geq 0$ and $t = 2r$, we deduce that $t \leq n-2$ and we are done with this case.

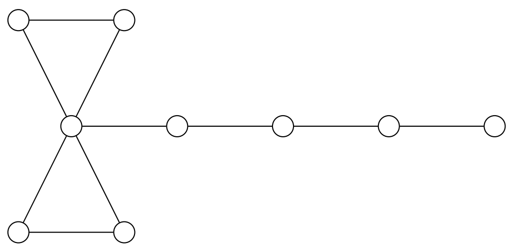


Figure 4.5: The graph $G_{2,9}$.

Now, assume $2 \leq r \leq t \leq 2r-1 \leq n-2$. We consider the graph $G_{x,y,z} \in \mathcal{F}$. We know that $G_{x,y,z}$ has an order of $2x+y+z+2$, and Remark 4.19 satisfies that $\dim(G_{x,y,z}) = x+y-1$ and $\dim_e(G_{x,y,z}) = 2x+y-2$. Since we are looking for a graph G of order n such that $\dim(G) = r$ and $\dim_e(G) = t$, we must find a graph $G_{x,y,z} \in \mathcal{F}$ for some x, y, z that will satisfy the following system of linear equations:

$$\begin{aligned} 2x+y+z+2 &= n \\ x+y-1 &= r \\ 2x+y-2 &= t \end{aligned}$$

We can easily compute that such a system has the solution $x = t - r + 1$, $y = 2r - t$ and $z = n - t - 4$ (note that these values represent integer numbers). Since the graph $G_{x,y,z} \in \mathcal{F}$ satisfies that $x \geq 1$, $y \geq 2$ and $z \geq 0$, we observe that $t - r + 1 \geq 1$, $2r - t \geq 2$ and $n - t - 4 \geq 0$. Thus, it follows that $t \geq r$, $t \leq 2r - 2$ and $t \leq n - 4$.

According to this, only the following cases remain, (1): $t = 2r - 1 \leq n - 2$ or (2): ($2 \leq r \leq t \leq 2r-2$ and $t \in \{n-3, n-2\}$). Assume that $t = 2r-1 \leq n-2$. Consider the graph G_r obtained as follows. We begin with the graph $G'' = K_1 \vee (\bigcup_{i=1}^{r} K_2)$. Then, we add a path of order $(n - 2r - 1) \geq 0$ (the case $n - 2r - 1 = 0$ means that we do not add any path and, clearly $n = 2r + 1$) and use an edge to join a leaf of this path with one non-central vertex of G''. See Figure 4.6 for an example.

Let S be a metric basis of G_r. Any two adjacent vertices of G'' different from the center and different from the endpoints of the K_2 closest to the added path have the same distance to any other vertex of G_r. Thus, S must contain at least $r-1$ vertices of G''. Since there is still one pair of vertices that is not distinguished, namely both endpoints of the K_2 closest to the added path, we have to add additional vertex to the set S and $\dim(G_r) \geq r$. On the other hand, a set composed by one vertex of each graph K_2 used to generate G'' is a metric generator of G_r. Therefore, $\dim(G_r) \leq r$, and using both inequalities it follows that $\dim(G_r) = r$.

Now, let S' be an edge metric basis of G_r. We observe that any two edges from G'' incident with the center of G'', except the edge connecting the center of G'' and vertex of G'' that is attached to the path, have the same distance to every other vertex of G_r. Thus, we deduce that S' must contain at least $2r - 2$ vertices of G''. Since there is still one pair of edges that is not distinguished, namely both edges of G'' that do not have an endpoint in the set S, we have to add one more vertex to the set S'. Thus, $\dim_{\mathrm{e}}(G_{r,n}) \geq 2r - 1$. On the other hand, a set composed by all but two vertices of G'' (the center and the vertex which is attached to the path) is an edge metric generator of G_r. Therefore, $\dim_{\mathrm{e}}(G_r) \leq 2r - 1$, and using both inequalities it follows that $\dim_{\mathrm{e}}(G_{r,n}) = 2r - 1$.

It is straightforward to observe that G_r has an order of $n - 2r - 1 + 2r + 1 = n$. It also satisfies that $\dim(G_r) = r$ and that $\dim_{\mathrm{e}}(G_r) = 2r - 1 = t$. Since $n - 2r - 1 \geq 0$ and $t = 2r - 1$, we deduce that $t \leq n - 2$.

Finally, we assume $2 \leq r \leq t \leq 2r-2$ with $t \in \{n-3, n-2\}$. First suppose that $t =$

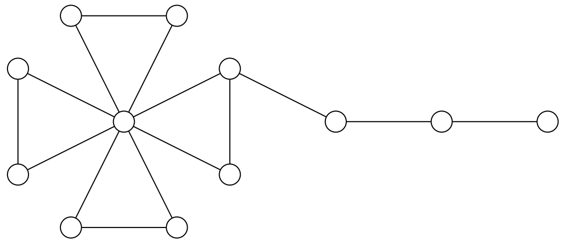
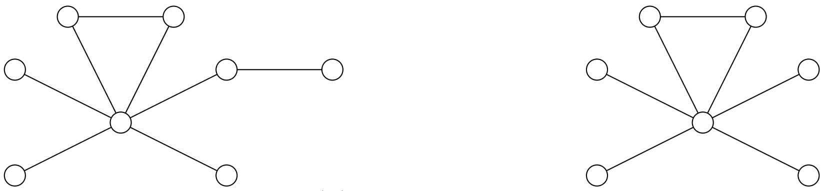


Figure 4.6: The graph G_4.

$n-3$. Consider the graph $G'_{r,t}$ given by the join graph $K_1 \vee [(\bigcup_{i=1}^{2r-t+1} K_1) \cup (\bigcup_{i=1}^{t-r} K_2)]$ and add a pendant vertex to one of its vertices of degree one. See Figure 4.7 (a) for an example.

It is straightforward to observe that $G'_{r,t}$ has an order of $2(t-r)+2r-t+1+2 = t+3 = n$. Following almost the identical arguments as above, we deduce that $\dim(G'_{r,t}) = (2r-t+1)+(t-r)-1 = r$ and that $\dim_{\mathrm{e}}(G'_{r,t}) = 2r-t+1+2(t-r)-1 = t$.



(a) (b)

Figure 4.7: The graph $G'_{4,5}$ (a) and the graph $G''_{4,5}$ (b).

Now, suppose that $t = n-2$. In such a case, we use a similar construction as above. Consider the graph $G''_{r,t}$ given by the join graph $K_1 \vee [(\bigcup_{i=1}^{2r-t+1} K_1) \cup (\bigcup_{i=1}^{t-r} K_2)]$. See Figure 4.7 (b) for an example. It is straightforward to observe that $G''_{r,t}$ has an order of $2(t-r)+2r-t+1+1 = t+2 = n$. Once again, following almost the identical arguments as above, we deduce that $\dim(G''_{r,t}) = (2r-t+1)+(t-r)-1 = r$ and that $\dim_{\mathrm{e}}(G''_{r,t}) = (2r-t+1)+2(t-r)-1 = t$, and we are done for this case, which completes the whole proof. □

As a consequence of the theorem above, one could think that for any graph G it follows that $\dim_{\mathrm{e}}(G) \leq 2\dim(G)$. However, this is not true and can be seen in the next example.

Example 4.21. *Let us take the wheel graph $W_{1,6}$. In Section 4.2.2, we recall the formulae for the metric dimension (from* [7]*) and compute the edge metric dimension of* *wheel graphs. We get that* $\dim_{\mathrm{e}}(W_{1,6}) = 5$ *and* $\dim(W_{1,6}) = 2$*. Thus, it follows that* $\dim_{\mathrm{e}}(W_{1,6}) > 2\dim(W_{1,6})$*.*

Other similar examples can easily be presented for wheels or fan graphs of a higher order. Moreover, we also observe that the difference between edge metric dimension and metric dimension can be as large as possible.

Proposition 4.22. *For any integer* $q \geq 1$*, a connected graph* G *exists, such that* $\dim_{\mathrm{e}}(G) - \dim(G) \geq q$*.*

Proof. The result can be obtained by using the wheel or fan graphs. For instance, from [7] we know that for every $n \geq 6$ it holds that $\dim(W_{1,n}) = \left\lfloor \frac{2n+2}{5} \right\rfloor$ and by Proposition 4.15 that $\dim_{\mathrm{e}}(W_{1,n}) = n-1$. Thus, by taking a wheel graph $W_{1,n}$, such that $n \geq \frac{5q+2}{3}$, we deduce that $n - 1 - \left\lfloor \frac{2n+2}{5} \right\rfloor \geq q$. □

According to the results for the case $\dim(G) < \dim_{\mathrm{e}}(G)$ obtained in this subsection, it remains for us to complete the realization of the triplet r, t, n for the case $r \geq 2$ and $t > 2r$ (if $2r < n-2$). Thus, we point out the following open problem.

Problem 4.23. *Is it possible to find a graph* G *of order* n *such that* $\dim(G) = r$ *and* $\dim_{\mathrm{e}}(G) = t$ *for any integers* r, t, n *with* $r \geq 2$ *and* $2r < t \leq n-2$*?*

Finally, we analyse the realizability of graphs G for which $\dim_{\mathrm{e}}(G) < \dim(G)$. In contrast to the other possibility $\dim(G) < \dim_{\mathrm{e}}(G)$, it seems that given a triplet of integers r, t, n with $2 \leq t < r \leq n-2$, it is quite a challenging problem to provide a connected graph G of order n such that $\dim(G) = r$ and $\dim_{\mathrm{e}}(G) = t$. By Theorem 4.17 we know that if $r = 4$ and $t = 3$, then for any $n = 16k$, for some integer $k \geq 1$, it is possible to provide a graph satisfying the conditions above. On the other hand, we have not found any other example in which this is also satisfied and we pose the following question.

Problem 4.24. *Given any three integers* r, t, n *with* $2 \leq t < r \leq n-2$*: Is it possible to construct a connected graph* G *of order* n *such that* $\dim(G) = r$ *and* $\dim_{\mathrm{e}}(G) = t$*?*

Another approach could be related to finding a possible bound for $\dim_{\mathrm{e}}(G)$ in terms of $\dim(G)$ for any connected graph G, under the supposition that $\dim_{\mathrm{e}}(G) <$

$\dim(G)$. For instance, if G is the torus graph $C_{4r}\square C_{4t}$, then $3 = \dim_e(G) = 4 - 1 = \dim(G) - 1$. In this sense, we pose the following question.

Problem 4.25. *Is there a positive constant c such that $\dim_e(G) \geq \dim(G) - c$ for any connected graph G?*

4.3 Complexity of the Edge Metric Dimension Problem

We studied the relationships between the edge metric dimension and the standard metric dimension in the previous section. The edge metric dimension is an interesting invariant and it can be used in several applications. Therefore, we want to know the complexity of the problem of computing the edge metric dimension of a graph. The decision problem concerning the metric dimension of a graph is already known as one of the classic NP-complete problems presented in book [25] (a formal proof of it appeared in [32]). We show that the corresponding problem for the edge metric dimension is also NP-complete. First, we need to introduce the 3-SAT problem. We will prove the NP-completeness of our problem, using the reduction from the 3-SAT problem, as in the case of the metric dimension proof of [32].

The 3-SAT problem is one of the most classic problems known as NP-complete.

3-SATISFIABILITY (3-SAT problem for short) INSTANCE: Collection $C = \{c_1, \ldots, c_m\}$ of clauses on a finite set U of variables, such that $\lvert c_i \rvert = 3$ for $1 \leq i \leq m$. QUESTION: Is there a truth assignment for U that satisfies all the clauses in C?

For more information on this problem, and on NP-completeness reductions in general, we suggest [25].

From now on in this section, we show that the problem of finding the edge metric dimension of an arbitrary connected graph is NP-hard. We first deal with the decision problem for the edge metric dimension.

EDGE METRIC DIMENSION PROBLEM (EDIM problem for short)
INSTANCE: A connected graph G of order $n \geq 3$ and an integer $1 \leq r \leq n-1$.
QUESTION: Is $\dim_e(G) \leq r$?

To study the complexity of the EDIM problem we make a reduction from the 3-SAT problem.

Theorem 4.26. *The EDIM problem is NP-complete.*

Proof. The problem is easily seen to be in NP. For a set of vertices S guessed by a non-deterministic algorithm for the problem, one needs to check that this is an edge metric generator. This can be done in polynomial time by calculating the distances from vertices to edges and checking that all pairs of distinct edges have different distance vectors with respect to the set S. We now describe a polynomial transformation of the 3-SAT problem to the EDIM problem.

Consider an arbitrary input of the 3-SAT problem, a collection $C = \{c_1, c_2, \ldots, c_m\}$ of clauses over a finite set $U = \{u_1, u_2, \ldots, u_n\}$ of Boolean variables. We shall construct a connected graph $G = (V(G), E(G))$ and set a positive integer $r \leq |V(G)| - 1$, such that the graph G has an edge metric generator of size r or less if and only if C is satisfiable. The construction will be made up of several components augmented by some additional edges for communication between various components.

For each variable $u_i \in U$ we construct a truth-setting component $X_i = (V_i, E_i)$, with $V_i = \{T_i, F_i, a_i^1, a_i^2, b_i^1, b_i^2\}$ and $E_i = \{T_i a_i^1, T_i a_i^2, b_i^1 a_i^1, b_i^2 a_i^2, F_i b_i^1, F_i b_i^2\}$ (see Figure 4.8 for reference). The vertices T_i and F_i are the `TRUE` and `FALSE` ends of the component, respectively. Each component is connected to the rest of the graph only through these two vertices, which gives us the following claim:

Lemma 4.27. *Let u_i be an arbitrary variable in U. Any edge metric generator must contain at least one of the vertices $\{a_i^1, a_i^2, b_i^1, b_i^2\}$.*

Proof. Suppose that an edge metric generator S without any of these vertices in it exists. Since the component X_i is attached to the rest of the graph only through the

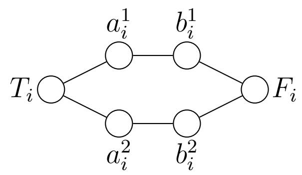


Figure 4.8: The truth-setting component for the variable u_i.

vertices T_i and F_i, due to the symmetry, this implies that the edges $T_i a_i^1$ and $T_i a_i^2$ have the same distances to all vertices in the set S, a contradiction. $\square$

Now, suppose that $c_j = y_j^1 \vee y_j^2 \vee y_j^3$, where y_j^k is a literal in the clause c_j. For such a clause c_j, we construct a satisfaction testing component $Y_j = (V_j', E_j')$, with $V_j' = \{c_j^1, \ldots, c_j^{10}\}$ and $E_j' = \{c_j^1 c_j^2, c_j^1 c_j^3, c_j^4 c_j^2, c_j^4 c_j^3, c_j^2 c_j^5, c_j^5 c_j^6, c_j^5 c_j^7, c_j^3 c_j^8, c_j^8 c_j^9, c_j^8 c_j^{10}\}$ (see Figure 4.9 for reference). The component is attached to the rest of the graph only through the vertices c_j^1 and c_j^2, which gives us the following claim.

Lemma 4.28. *Let c_j be an arbitrary clause in C. Any edge metric generator must contain at least one of the vertices $\{c_j^6, c_j^7\}$ and at least one of the vertices $\{c_j^9, c_j^{10}\}$.*

Proof. Suppose that an edge metric generator S exists, containing no of the vertices $\{c_j^6, c_j^7\}$. Since all the shortest paths from any vertex $x \neq c_j^6, c_j^7$ to the edges $c_j^5 c_j^6$ and $c_j^5 c_j^7$ go through the vertex c_j^5, this implies that the edges $c_j^5 c_j^6, c_j^5 c_j^7$ have the same distance to all vertices in the set S, a contradiction. A similar process works for the vertices $\{c_j^9, c_j^{10}\}$ and the edges $c_j^8 c_j^9$ and $c_j^8 c_j^{10}$. $\square$

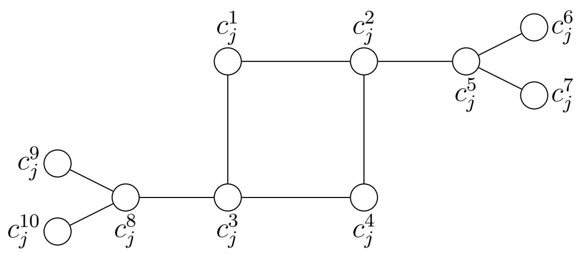


Figure 4.9: The satisfaction testing component for the clause c_j.

We also add some edges between truth-setting and satisfaction testing components, as follows. If a variable u_i occurs as a positive literal in a clause c_j, then we add the edges $T_i c_j^1$ and $F_i c_j^2$. If a variable u_i occurs as a negative literal in a clause c_j, then we add the edges $T_i c_j^2$ and $F_i c_j^1$. For each clause $c_j \in C$, denote those six added edges with E''_j. We call them *the communication* edges. Figure 4.10 shows the edges that were added corresponding to the clause $c_j = (u_1 \vee \overline{u_2} \vee u_3)$, where $\overline{u_2}$ represents the negative literal corresponding to the variable u_2.

For all $k \in \{1, \ldots, n\}$ such that neither of u_k and $\overline{u_k}$ occur in clause c_j, add the edges $T_k c_j^2$ to the graph G. For each clause $c_j \in C$, denote them with E'''_j. Those edges ensure that the graph is connected. We call them *the neutralizing* edges, because no matter what value is assigned to the variable u_k (or equivalently, which vertex v_k from the corresponding truth-setting component is chosen for an edge metric generator), this gives the same distance from the chosen vertex v_k to the edges $c_j^1 c_j^2$ and $c_j^2 c_j^4$ from the satisfaction testing component corresponding to the clause c_j. These two edges play an important role later in the proof.

Finally, for each clause c_j and every $k \in \{1, \ldots, m\}, k \neq j$, add the edges $c_j^2 c_k^2$ to the graph G. For each clause $c_j \in C$, denote them with E''''_j. We call these edges *the correcting* edges.

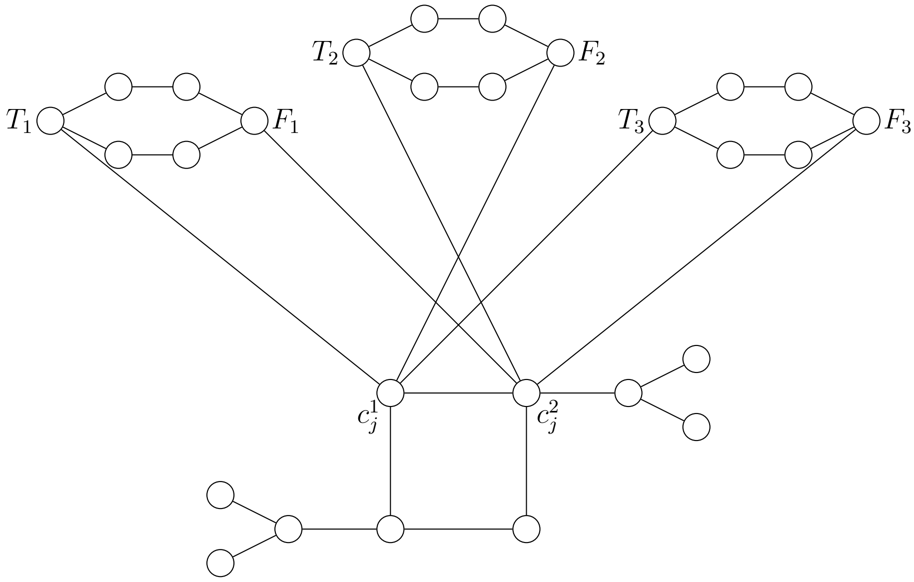


Figure 4.10: The subgraph associated with the clause $c_j = (u_1 \vee \overline{u_2} \vee u_3)$.

The construction of our instance of the EDIM problem is then completed by setting $r = 2m + n$ and $G = (V(G), E(G))$, where

$$V(G) = \left(\bigcup_{i=1}^{n} V_i \right) \cup \left(\bigcup_{j=1}^{m} V'_j \right)$$

and

$$E(G) = \left(\bigcup_{i=1}^{n} E_i\right) \cup \left(\bigcup_{j=1}^{m} E'_j\right) \cup \left(\bigcup_{j=1}^{m} E''_j\right) \cup \left(\bigcup_{j=1}^{m} E'''_j\right) \cup \left(\bigcup_{j=1}^{m} E''''_j\right)$$

It is not hard too see that the construction can be done in polynomial time. What remains is to show that C is satisfiable if and only if G has an edge metric generator of size r. From Lemmas 4.27 and 4.28 we get the following.

Corollary 4.29. *The edge metric dimension of the graph G is at least $r = 2m + n$.*

We now continue with the following lemmas, which constitute the heart of our NP-completeness reduction from 3-SAT.

Lemma 4.30. *If C is satisfiable, then the edge metric dimension of the graph G is r.*

Proof. We know that the edge metric dimension is at least r. We now construct an edge metric generator S of size r based on a satisfying truth assignment for C. Let $t : U \to \{\texttt{TRUE}, \texttt{FALSE}\}$ be a satisfying truth assignment for C. For each clause $c_j \in C$, put in the set S the vertices c_j^6 and c_j^9. For each variable $u_i \in U$, put in the set S either the vertex a_i^1 if $t(u_i) = \texttt{TRUE}$, or the vertex b_i^1 if $t(u_i) = \texttt{FALSE}$. We now show that S is an edge metric generator for the graph G.

Let $e_{j,k}$ be an arbitrary correcting edge between the satisfaction testing components c_j and c_k. We notice that $e_{j,k}$ is uniquely determined by the set of vertices $\{c_j^6, c_k^6\}$, because this is the only edge in the graph G that has a distance of 2 to both c_j^6 and c_k^6.

Let $i \in \{1, \ldots, n\}$ and $j \in \{1, \ldots, m\}$ be arbitrary indices and let $v_i \in V_i \cap S$. Since we have already checked that any correcting edge is uniquely determined by some vertex in S, we do not have to check any pair of edges in which at least one correcting edge occurs. Also, it is easy to check that each communication edge and each neutralizing edge between a truth-setting component X_i and a satisfaction testing component Y_j is distinguished from all the remaining edges by the vertices v_i, c_j^6 and c_j^9.

We next take a look at the edges in a truth-setting component. Let $i \in \{1, \ldots, n\}$ be an arbitrary index and let $e \in E_i$ be an arbitrary edge from X_i. Since we have already checked that all correcting, communication, and neutralizing edges are distinguished by some vertex from S, we only need to check that e has different distance vectors: (1) from all other edges in X_i, (2) from all edges in other truth-setting components, and (3) from all edges in the satisfaction testing components. This is addressed next. (1) In checking that e has different distance vectors to all other edges in X_i, we consider two possibilities.

- u_i or $\overline{u}_i$ is a literal in at least one clause c_j. Thus, the vertices v_i, c_j^6 and c_j^9 distinguish the edge e from all other edges in X_i;
- neither u_i nor $\overline{u}_i$ are literals in any clause c_j. Thus, for an arbitrary $j \in \{1, \ldots, m\}$, the vertices v_i, c_j^6 distinguish the edge e from all other edges in X_i.

For (2), let $k \in \{1, \ldots, n\}, k \neq i$, be an arbitrary index. The vertex v_i distinguishes the edge e from all edges $f \in E_k$ (the edges in the truth-setting component X_k). For (3), let $j \in \{1, \ldots, m\}$ be an arbitrary index. Hence, the vertices c_j^6 and c_j^9 distinguish the edge e from all edges $f \in E'_j$ (the edges in the satisfaction testing component Y_j).

Finally, we take a look at the edges from the satisfaction testing components. Let $j \in \{1, \ldots, m\}$ be an arbitrary index. Each one of the edges $\{c_j^2c_j^5, c_j^5c_j^6, c_j^5c_j^7, c_j^3c_j^8, c_j^8c_j^9, c_j^8c_j^{10}\}$ is uniquely determined by the set of vertices $\{c_j^6, c_j^9\}$. Those two vertices also distinguish the edges $c_j^1c_j^2, c_j^2c_j^4$ from all other edges but they do not distinguish $c_j^1c_j^2$ and $c_j^2c_j^4$ themselves. Similarly, the same holds for the edges $c_j^1c_j^3, c_j^3c_j^4$. To complete the proof, we need to show that for precisely this pair of edges there exists a vertex in the set S that distinguishes them. Since C is satisfiable, suppose that c_j is satisfied by the variable u_i. For the variable u_i there are two possibilities:

- u_i occurs as a positive literal in c_j and $t(u_i) = \texttt{TRUE}$
- u_i occurs as a negative literal in c_j and $t(u_i) = \texttt{FALSE}$.

Thus, if $t(u_i) = \texttt{TRUE}$, then we have added the vertex a_i^1 to the set S. In this case, the distance from a_i^1 to the edge $c_j^1c_j^2$ is 2, while the distance to the edge $c_j^2c_j^4$ is 3.

Similarly, the distance from a_i^1 to the edge $c_j^1 c_j^3$ is 2 and to the edge $c_j^3 c_j^4$ is 3. The case when $t(u_i) = \texttt{FALSE}$ is symmetric.

Therefore, any two edges are distinguished by a vertex of S and it follows that S is an edge metric generator of graph G, which completes the proof of this lemma. $\square$

Lemma 4.31. *If the edge metric dimension of the graph G is r, then C is satisfiable.*

Proof. Let S be an arbitrary edge metric generator with cardinality r of the graph G. From Lemmas 4.27 and 4.28, the set S must contain at least one vertex from each truth-setting component and at least two vertices from each satisfaction testing component. Since the cardinality of S equals $r = 2m + n$, it follows that in the set S there is exactly one vertex from each truth-setting component and there are exactly two vertices from each satisfaction testing component. We shall find a function $t : U \to \{\texttt{TRUE}, \texttt{FALSE}\}$ such that it represents a satisfying truth assignment for the collection of clauses C. For an arbitrary $i \in \{1, \ldots, n\}$, let $v_i \in V_i \cap S$. Hence, we define a function t as follows:

$$t(u_i) = \begin{cases} \texttt{TRUE}, & \text{if } v_i \in \{a_i^1, a_i^2\}, \\ \texttt{FALSE}, & \text{if } v_i \in \{b_i^1, b_i^2\}. \end{cases}$$

We shall show that t produces a satisfying truth assignment for C. To this end, let c_j be an arbitrary clause. We claim that at least one of its literals has the value $\texttt{TRUE}$. We prove that fact by tracing which vertex from S distinguishes the edges $e_j^1 = c_j^1 c_j^2$ and $e_j^2 = c_j^2 c_j^4$, and showing that the corresponding function t satisfies c_j.

Let $k \in \{1, \ldots, m\}$ be an arbitrary index. For the clause c_k we assume, without loss of generality, that the vertices in the set S are c_k^6 and c_k^9. If $j = k$, then both edges e_j^1 and e_j^2 are at a distance of 2 from c_k^6 and at a distance of 3 from c_k^9. If $j \neq k$, then by using the correcting edges, we deduce that the edges e_j^1 and e_j^2 are at a distance of 3 from c_k^6 and at a distance of 5 from c_k^9. Therefore, none of these vertices distinguish e_j^1 from e_j^2.

Now, consider any variable u_i that does not occur in c_j. If $v_i \in \{a_i^1, a_i^2\}$, then both edges e_j^1, e_j^2 are at a distance of 2 from v_i. If $v_i \in \{b_i^1, b_i^2\}$, then both edges are at a distance of 3 from v_i. Thus, the vertex of S distinguishing the edges e_j^1, e_j^2 must belong to one of the truth-setting components that corresponds to a variable

u_k, which occurs in the clause c_j. We recall that we have added communication edges in such a manner that v_k distinguishes the edges e_j^1 and e_j^2 only if one of the following statements holds true:

- u_k occurs as a positive literal in c_j and $v_k \in \{a_k^1, a_k^2\}$ (in this case $t(u_k) =$ `TRUE`);
- u_k occurs as a negative literal in c_j and $v_k \in \{b_k^1, b_k^2\}$ (in this case $t(u_k) =$ `FALSE`).

In both cases, the clause c_j is satisfied by the setting assigned to the variable u_k. As a consequence, the formula C is satisfiable, which completes the proof of this lemma. □

As a consequence of the Lemmas 4.30 and 4.31, the polynomial transformation from 3-SAT to the EDIM problem is done, and the proof of the theorem is now completed. □

From Theorem 4.26, we obtain the following result.

Corollary 4.32. *The problem of finding the edge metric dimension of a connected graph is NP-hard.*

4.3.1 Approximation of the EDIM Problem

In concordance with Corollary 4.32, finding the edge metric dimension of a graph is NP-hard in general. Thus, it is reasonable to look for an approximation algorithm for it. We use an approach similar to that in [32], obtaining an approximation in polynomial time within a factor of $O(\log m)$, where m is the number of edges of the graph. We show that the problem of finding the edge metric dimension can be transformed in polynomial time to the set cover problem. Once we have the set cover problem, we use the $O(\log m)$ factor approximation algorithm for the set cover problem [29] to obtain an approximation algorithm for the EDIM problem.

Theorem 4.33. *If G is an arbitrary connected graph with m edges, then $\dim_{\mathrm{e}}(G)$ can be approximated within a factor of $O(\log m)$ in polynomial time.*

Proof. Starting from the graph G, we first construct an instance of the set cover problem, similar to the one in [29]. Let F be a finite family $\{M_1, M_2, \ldots, M_p\}$ of finite sets, and let $U = \bigcup_{M \in F} M$ be the universe set. We are looking for a subfamily $F' \subseteq F$ with the minimum cardinality for which it holds that $\bigcup_{M \in F'} M = U$.

For each vertex in the graph G we can compute in polynomial time all the pairs of edges that have different distances to that vertex. For a vertex v, use M_v to denote the set of all such pairs of edges. To solve the EDIM problem one has to find a set of vertices S with minimum cardinality, such that every pair of edges is distinguished by some vertex $v \in S$. We can easily transform the EDIM problem to the set cover problem by setting $F = \{M_{v_1}, \ldots, M_{v_n}\}$, where $v_1, \ldots, v_n$ are all the vertices from the graph G. Observe that the universe set U is the set of all possible pairs of distinct edges in the graph G with cardinality $\binom{m}{2}$. It remains to be shown that in the graph G there exists an edge metric basis of size k if and only if there is a set cover of size k for the corresponding instance of the set cover problem.

First, suppose that an edge metric basis S of size k exists. Take the sets M_{v_i} for all the $v_i \in S$ into the subfamily F'. There are clearly k sets in the F'. Any element (a pair of edges $e_i e_j$) from the universe set U is then covered by the set M_{v_a}, where $v_a \in S$ is a vertex that distinguishes edges e_i and e_j in the graph G.

For the converse, suppose that F' is a set cover of cardinality k for the universal set U. For the edge metric basis S take all the vertices that correspond to the sets of the subfamily F'. From the construction of the instance of the set cover problem it follows that such a set S distinguishes all the pairs of distinct edges. The cardinality of the set S is the smallest possible, otherwise a set cover F'' with a smaller cardinality than F' would exist, which would be a contradiction.

For the set cover problem there is a polynomial approximation algorithm that finds a set cover within a factor of $O(\log m)$. Therefore, we get the same approximation for the EDIM problem. □

The approximation algorithm that finds a set cover within a factor of $O(\log m)$ is a greedy algorithm. It starts with an empty subfamily F'. At each step, it takes into the subfamily F' the set with maximum cardinality at the current step. It stops when the whole universe set is covered.

5

Mixed Metric Dimension

Metric dimension deals with a subset of vertices that distinguishes pairs of distinct vertices, while edge metric dimension deals with a subset of vertices that distinguishes pairs of distinct edges. What if we want to locate an intruder in a network and we do not know if the intruder is at some vertex or on some edge? In this situation, we want to distinguish not only pairs of distinct vertices and pairs of distinct edges separately, but also pairs consisting of a vertex and an edge. Thus, we need a kind of mixed version of the metric dimension and the edge metric dimension. That is, given a connected graph G, we wish to uniquely identify the elements (edges and vertices) of G by means of vector distances to a fixed set of vertices of G.

We say that a vertex v of a connected graph G *distinguishes* two elements $x, y \in V(G) \cup E(G)$ of a graph G if $d_G(x, v) \neq d_G(y, v)$. A set S of vertices of G is a *mixed metric generator* if any two elements $x, y \in V(G) \cup E(G)$ of G, where $x \neq y$, are distinguished by some vertex of S. The smallest cardinality of a mixed metric generator of G is called the *mixed metric dimension* and is denoted by $\dim_{\mathrm{m}}(G)$. A *mixed metric basis* of G is a mixed metric generator of G of cardinality $\dim_{\mathrm{m}}(G)$.

The mixed metric dimension was introduced recently [34], and we have not found any other literature on this topic.

We consider the structure of mixed metric generators and characterize graphs for which the mixed metric dimension equals the trivial lower and upper bounds. We also give results on the mixed metric dimension of some families of graphs and present an upper bound with respect to the girth of a graph. Finally, we prove that the problem of determining the mixed metric dimension of a graph is NP-hard in the general case.

Similar to the case of edge metric dimension, the problem of determining the mixed metric dimension of a given graph can also be restated as an optimization problem. Let us now present this mathematical programming model, which can be used to solve the problem of computing the mixed metric dimension or finding a mixed metric basis of a graph G.

Let G be a graph of order n and size m with the vertex set $V = \{v_1, v_2, \ldots, v_n\}$ and the edge set $E = \{e_1, e_2, \ldots, e_m\}$. We consider the $n \times (n+m)$ dimensional matrix $D = [d_{ij}]$ such that $d_{ij} = d_G(x_i, x_j)$, where $x_i \in V$ and $x_j \in V \cup E$. Now, given the variables $y_i \in \{0, 1\}$ with $i \in \{1, 2, \ldots, n\}$ we define the following function:

$$\mathcal{F}(y_1, y_2, \ldots, y_n) = y_1 + y_2 + \cdots + y_n.$$

Minimizing the function $\mathcal{F}$ subject to the following constraints

$$\sum_{i=1}^{n} |d_{ij} - d_{il}| y_i \geq 1, \quad \text{for every } 1 \leq j < l \leq n + m,$$

is equivalent to finding a mixed metric basis of G. Namely, the solution for $y_1, y_2, \ldots, y_n$ represents a set of values for which the function $\mathcal{F}$ achieves the minimum value possible. This is equivalent to saying that the set $W = \{v_i \in V \mid y_i = 1\}$ is a mixed metric basis of G. On the other hand, let W' be a mixed metric basis of G and let $(y'_1, y'_2, \ldots, y'_n)$ be a vector such that for any $i \in \{1, 2, \ldots, n\}$, $y'_i = 0$ if $v_i \notin W'$, or $y'_i = 1$ if $v_i \in W'$. It is straightforward to observe that $\mathcal{F}(y'_1, y'_2, \ldots, y'_n)$ gives a minimum subject to the constraints given before.

5.1 The Structure of Mixed Metric Generators

We continue with several combinatorial properties of mixed metric generators. First, it follows from definition that any mixed metric generator is also a metric generator and an edge metric generator. In this sense, the following relationship immediately follows. For any connected graph G,

$$\dim_{\mathrm{m}}(G) \geq \max\{\dim(G), \dim_{\mathrm{e}}(G)\}. \tag{5.1}$$

On the other hand, it is not difficult to see that the whole vertex set of any graph G forms a mixed metric generator. Also, any vertex of G and any edge incident with it are at the same distance to the vertex itself. In this sense, a vertex alone cannot form a mixed metric generator of G. As a consequence of these situations, the following remark is readily seen to be true.

Remark 5.1. *For any graph G of order n, $2 \leq \dim_{\mathrm{m}}(G) \leq n$.*

Let us recall that two vertices u, v of G are called *false twins* if they have the same open neighbourhoods, *i.e.* $N(u) = N(v)$. Similarly, the vertices u, v are called *true twins* if $N[u] = N[v]$. A vertex v is a *true twin* or a *false twin* of G if there exists $u \neq v$ such that u, v are true twins or false twins, respectively.

Proposition 5.2. *If u, v are true twins of a graph G, then u and v belong to every mixed metric generator of G.*

Proof. Since u, v are adjacent, it follows that the edge uv and the vertex v are at the same distance to every vertex of the graph except u. Similarly, the edge uv and the vertex u are at the same distance to every vertex of the graph except v. As a consequence, u, v must belong to every mixed metric generator of G. □

Proposition 5.3. *If u, v are false twins of a graph G, and S is a mixed metric generator of G, then $\{u, v\} \cap S \neq \emptyset$.*

Proof. If u, v are false twins, it follows that they are at the same distance to every vertex of G except themselves. Thus, if S is a mixed metric generator of G, then at least one of them must belong to S. □

Proposition 5.4. *If u is a simplicial vertex of a graph G, then u belongs to every mixed metric generator of G.*

Proof. Since $N(u)$ induces a complete graph for any vertex $v \in N(u)$, it follows that the edge uv and the vertex v are at the same distance to every vertex of the graph, except u. Therefore, the vertex u must belong to every mixed metric generator of G. □

As a direct consequence of Proposition 5.4, we obtain the following result.

Corollary 5.5. *If u is a vertex of degree 1 of a graph G, then u belongs to every mixed metric generator of G.*

Following this, we deal with characterizing the families of graphs achieving the equality in the bounds from Remark 5.1.

Theorem 5.6. *Let G be a graph of order n. It holds that $\dim_m(G) = 2$ if and only if G is a path.*

Proof. Following Corollary 5.5, both end-vertices of the path must be in every mixed metric generator, therefore $\dim_m(P_n) \geq 2$. It is straightforward to observe that for any path P_n the two leaves of the path distinguish all pairs of distinct elements (vertices and/or edges) of the path. It follows that $\dim_m(P_n) = 2$.

For the converse, assume G satisfies that $\dim_m(G) = 2$, and let $S = \{u, v\}$ be any mixed metric basis. If there is a neighbour v' of v such that $d(v', u) \geq d(v, u)$, then $d(v'v, u) = d(v, u)$, which means that the edge $v'v$ and the vertex v are not distinguished by any vertex of S, a contradiction. Thus, for any vertex v' adjacent to v it follows that $d(v', u) = d(v, u) - 1$.

Now, if two vertices x and y belonging to two different shortest $u - v$ paths exist, such that $d(x, u) = d(y, u)$, then also $d(x, v) = d(y, v)$, which means x, y are not distinguished by S, a contradiction again.

Thus, exactly one shortest $u - v$ path in G exists, say $P = uw_1w_2 \dots w_rv$. Suppose $i \in \{1, \dots, r\}$ exists, such that the vertex w_i in P is of a degree of at least three, and let w' be a neighbour of w_i which is not in P. Since S is a mixed metric basis, the edge w_iw' and the vertex w_i are distinguished by some $x \in S$. This means that $d(w_i, x) \neq d(w_iw', x) = \min\{d(w_i, x), d(w', x)\}$. It follows that $d(w', x) < d(w_i, x)$. Let $x' \in S \setminus \{x\}$. Since $d(w', x) \leq d(w_i, x) - 1$, there is a path $Q = x \dots w'w_i \dots x'$ of length $d(x,w')+d(w',w_i)+d(w_i,x') \leq d(w_i,x)-1+1+d(w_i,x') = d(w_i,x)+d(w_i,x')$ from x to x' (note that $\{x,x'\} = \{u,v\}$), a contradiction since this is either a $u-v$ path shorter than P (which is the shortest $u-v$ path) or a path of the same length as P (contradicting the uniqueness of P). Thus, every vertex w_i, with $i \in \{1,\ldots,r\}$, in P has a degree of two.

It remains to be proven that u and v are both of degree 1. Suppose u is of a degree of at least 2. Let u' be the neighbour of u, which is not in P. Since S is a mixed metric basis, the vertex v must distinguish the edge uu' and the vertex u. It follows that $d(u',v) < d(u,v)$. Following the same line of thought as for the case above we obtain contradictions for all possibilities. Therefore, u is of degree 1. Analogously, v is of degree 1. Since G is connected, it follows that G must be a path. □

Lemma 5.7. *Let v be an arbitrary vertex of a graph G and let $S = V(G) \setminus \{v\}$. If for every $w \in N(v)$ there exists $x \in S$ such that $d(vw,x) \neq d(w,x)$, then S is a mixed metric generator of the graph G.*

Proof. Assume that for every $w \in N(v)$ there exists $x \in S$ such that $d(vw,x) \neq d(w,x)$. If we want to prove that S is a mixed metric generator, we have to show that any two distinct elements (vertices or edges) of the graph G are distinguished by some vertex from the set S. Any subset of $V(G)$ with cardinality $n-1$ is a metric generator and also an edge metric generator. Thus, we only have to check pairs of elements in which one element is a vertex and the other is an edge. Let $e \in E(G)$ be an arbitrary edge. The vertex v and the edge e are distinguished by at least one endpoint of the edge e. All vertices different from v are in the set S. This means that for an arbitrary vertex $u \in V(G) \setminus \{v\}$ we only have to check the edges that are incident with the vertex u. If both endpoints of the edge $e = uw$ are in the set S, then u and e are distinguished by the vertex w. What remains is for us to check only the pairs of vertices w and edges wv, for all $w \in N(v)$. Since we know that for all such pairs there exists $x \in S$ such that $d(vw,x) \neq d(w,x)$, it follows that S is a mixed metric generator. □

Let v be a vertex of a graph G. A vertex $u \in N(v)$ is said to be a *maximal neighbour* of the vertex v if all neighbours of v (and v itself) are also in the closed neighbourhood of u. Now, we are ready to characterize the family of graphs G of order n, satisfying that $\dim_{\mathrm{m}}(G) = n$.

Theorem 5.8. *Let G be a graph of order n. Then $\dim_{\mathrm{m}}(G) = n$ if and only if every vertex of the graph G has a maximal neighbour.*

Proof. First let $\dim_{\mathrm{m}}(G) = n$. We want to prove that for every $v \in V(G)$ there exists $u \in N(v)$ such that $N[v] \subseteq N[u]$. Towards contradiction, suppose that there exists $v \in V(G)$ such that for every $u \in N(v)$ it holds that $N[v] \nsubseteq N[u]$. Let $S = V(G) \backslash \{v\}$. We claim that S is a mixed metric generator.

If S is not a mixed metric generator, then due to Lemma 5.7 there exists $w \in N(v)$ such that for every $x \in S$ it holds that $d(vw, x) = d(w, x)$. Since $w \in N(v)$, it follows that $N[v] \nsubseteq N[w]$, so there exists $v' \in N(v)$ such that $wv' \notin E(G)$. It follows that $1 = d(vw, v') \neq d(w, v') = 2$, a contradiction. So S is a mixed metric generator and $\dim_{\mathrm{m}}(G) < n$, a contradiction.

For the converse, assume that for every $v \in V(G)$ there exists $u \in N(v)$ such that $N[v] \subseteq N[u]$. Suppose that $\dim_{\mathrm{m}}(G) < n$. Therefore, a mixed metric generator S with cardinality $n - 1$ exists, and $v \in V(G)$, such that $v \notin S$. Let $u \in N(v)$ be a neighbour of v for which it holds that $N[v] \subseteq N[u]$. Since S is a mixed metric generator, there must exist $x \in S$, such that $d(u, x) \neq d(uv, x)$. Thus, it follows that $d(v, x) < d(u, x)$. On an arbitrary shortest path between x and v there exists $v' \in N(v)$ such that $d(v, x) = d(v', x) + 1$. Since $N[v] \subseteq N[u]$ it follows that $d(v, x) \geq d(u, x)$, a contradiction. Therefore $\dim_{\mathrm{m}}(G) = n$. $\square$

5.2 Mixed Metric Dimension of Some Families of Graphs

In this section, we determine the mixed metric dimension of cycles, complete bipartite graphs, trees, and grid graphs.

Proposition 5.9. *For any positive integer $n \geq 4$, $\dim_{\mathrm{m}}(C_n) = 3$.*

Proof. From Remark 5.1 and Theorem 5.6 we know that $\dim_{\mathrm{m}}(C_n) \geq 3$. On the other hand, let $V(C_n) = \{v_0, v_1, \ldots, v_{n-1}\}$, where $v_i v_{i+1} \in E(C_n)$ for every $i \in$

$\{0, \ldots, n-1\}$ and operation $i+1$ is done modulo n. Let $S = \{v_0, v_1, v_{\lceil \frac{n}{2} \rceil}\}$. It is clear that the vertices v_0, v_1 distinguish every pair of two distinct vertices or two distinct edges. Now, let e be an edge and let v_i be a vertex. If $d(e, v_0) = d(v_i, v_0)$ and $d(e, v_1) = d(v_i, v_1)$, then either $e = v_i v_{i+1}$ or $e = v_{i-1} v_i$ must happen. Thus, it follows that either $d(e, v_{\lceil \frac{n}{2} \rceil}) = d(v_{i+1}, v_{\lceil \frac{n}{2} \rceil}) < d(v_i, v_{\lceil \frac{n}{2} \rceil})$ or $d(e, v_{\lceil \frac{n}{2} \rceil}) = d(v_{i-1}, v_{\lceil \frac{n}{2} \rceil}) < d(v_i, v_{\lceil \frac{n}{2} \rceil})$. Therefore, the edge e and the vertex v_i are distinguished by $v_{\lceil \frac{n}{2} \rceil}$ and, as a consequence, S is a mixed metric generator of cardinality three, which completes the proof. □

Proposition 5.10. *For any positive integers $r, t \geq 2$,*

$$\dim_{\mathrm{m}}(K_{r,t}) = \begin{cases} r+t-1, & \text{if } r = 2 \text{ or } t = 2, \\ r+t-2, & \text{otherwise.} \end{cases}$$

Proof. From [11] and [36] we know that $\dim(K_{r,t}) = \dim_{\mathrm{e}}(K_{r,t}) = r+t-2$. Thus, by using (5.1) we have $\dim_{\mathrm{m}}(K_{r,t}) \geq r+t-2$. Let U and V be the bipartition sets of $K_{r,t}$ with $|U| = r$ and $|V| = t$. We first consider the case of $r = 2$. Suppose $\dim_{\mathrm{m}}(K_{r,t}) = r+t-2$ and let S be a mixed metric basis for $K_{2,t}$. Since any metric basis or edge metric basis must contain at least $r-1$ vertices of U and $t-1$ vertices of V, we deduce that $|U \cap S| = 1$ and $|V \cap S| = t-1$. Let $u \in U \cap S$ and $v \in V \setminus S$. We observe that the vertex u has a distance of 0 to itself (vertex u) and a distance of 1 to every other vertex in S. Moreover, the edge uv has a distance of 0 to the vertex u and a distance of 1 to every other vertex in S. Thus, the vertex u and the edge uv are not distinguished by S, a contradiction. A similar contradiction is obtained if $t = 2$. Therefore, $\dim_{\mathrm{m}}(K_{r,t}) \geq r+t-1$ and the proof is completed by using Theorem 5.8, since no vertex of $K_{r,t}$ admits a maximal neighbour.

From now on, assume $r, t \geq 3$. Let S be a set of vertices of G of cardinality $r+t-2$, such that it does not contain exactly one vertex from each bipartition set of $K_{r,t}$. Since S is a metric basis and also an edge metric basis, we only need to check that S distinguishes those pairs given by an edge and by a vertex. But this is straightforward to observe since any edge of $K_{r,t}$ has a distance of 0 or 1 to every vertex of S, and for any vertex there is at least one vertex in S at a distance of 2, since $r \geq 3$ and $t \geq 3$. Therefore, S is a mixed metric generator of cardinality $r+t-2$ and the result follows. □

Theorem 5.11. *For any tree T with $l(T)$ leaves,* $\dim_{\mathrm{m}}(T) = l(T)$.

Proof. Let S be the set of all leaves of T and let x, y be any two distinct elements of T. From [32] and [36] it is known that for any tree T a metric basis and an edge metric basis are both subsets of leaves in T. Thus, if x, y are either two vertices or two edges, they are distinguished by S, which is formed by all the leaves of T. Now, assume $x = x_1x_2$ is an edge and y is a vertex. Without loss of generality, we assume there is an $x_1 - y$ path containing x_2 (notice that it could happen $y = x_2$). Now, let x' and y' be two leaves of T such that x_1, x_2, y lie in the $x' - y'$ path (notice that it could be $x' = x_1$ and $y' = y$ or vice-versa). Thus, it is easy to see that at least one of the leaves x' or y' distinguishes x and y. The case of only one of these two leaves distinguishing x and y is given when $x_2 = y$. Therefore, S is a mixed metric generator and we observe that $\dim_{\mathrm{m}}(T) \leq l(T)$. On the other hand, since every leaf of T is of degree 1 from Corollary 5.5, we obtain that $\dim_{\mathrm{m}}(T) \geq l(T)$, which completes the proof. □

Next, we give the value of the mixed metric dimension of the grid graph, which is the Cartesian product of two paths P_r and P_t with r and t vertices, respectively.

Proposition 5.12. *If G is the grid graph $G = P_r \Box P_t$, with $r \geq t \geq 2$, then* $\dim_{\mathrm{m}}(G) = 3$.

Proof. In order to simplify the procedure, we shall embed G into $\mathbb{Z}^2$. That is, each vertex of the grid is represented as an ordered pair of coordinates (x, y). In this sense, G is embedded into $\mathbb{Z}^2$ in such way that $(0,0), (r-1,0), (0,t-1), (r-1,t-1)$ are the corner vertices of G (the vertices of degree two). We shall prove that the set $S = \{(0,0), (0,t-1), (r-1,0)\}$ is a mixed metric generator of G. Consider any two different elements x, y of G.

Case 1: x, y are vertices. From [32] we know that $S' = \{(0,0), (0,t-1)\}$ is a metric generator of G. Thus, x and y are distinguished by $(0,0)$ or by $(0,t-1)$. Notice that $S = \{(0,0), (r-1,0)\}$ is also a metric generator of G.

Case 2: x, y are edges. From [36] we know that $S' = \{(0,0), (0,t-1)\}$ or $S = \{(0,0), (r-1,0)\}$ are edge metric generators of G and we are done for this case.

Case 3: x is a vertex and y is an edge, say $x = (i,j)$ and $y = (k,a)(k,b)$ (notice that endpoints of any edge either have equal first components or equal second components). Without loss of generality we assume $a < b$ (which means $b = a+1$). Suppose the vertex x and the edge y are not distinguished by S. This means the following:

$$i + j = d(x,(0,0)) = d(y,(0,0)) = k + a,$$

$$i + t - 1 - j = d(x,(0,t-1)) = d(y,(0,t-1)) = k + t - 1 - b = k + t - 2 - a,$$

$$j + r - 1 - i = d(x,(r-1,0)) = d(y,(r-1,0)) = a + r - 1 - k.$$

Thus, we obtain the following system of equations:

$$\begin{aligned} i + j - k - a &= 0 \\ i - j - k + a &= -1 \\ -i + j + k - a &= 0, \end{aligned}$$

which is straightforward to observe to be a system of linear equations without solutions, a contradiction. An analogous procedure gives a similar contradiction in the cases $x = (i,j)$ and $y = (a,k)(b,k)$. Thus, at least one of the vertices in S distinguishes the pair x, y. As a consequence, S is a mixed metric generator of cardinality three. Therefore, we complete the proof using Theorem 5.6. □

5.3 An Upper Bound for the Mixed Metric Dimension of Graphs

We can give an upper bound for $\dim_{\mathrm{m}}(G)$ in terms of the girth of the graph.

Theorem 5.13. *Let G be a graph of order n. If G has a cycle, then* $\dim_{\mathrm{m}}(G) \leq n - g(G) + 3$.

Proof. Let $C = v_0 v_1 \ldots v_{r-1}$ be a cycle of order $r = g(G)$ of the graph G. We claim that $S = (V(G) \setminus V(C)) \cup \{v_0, v_1, v_{\lceil \frac{r}{2} \rceil}\}$ is a mixed metric generator.

Let $x, y \in V(G)$ be two arbitrary distinct vertices. If at least one of them, say x, is in S, then they are clearly distinguished by x, since $0 = d(x,x) \neq d(x,y) > 0$. If none of them is in S, then they are vertices of the cycle C and, following Proposition

5.9, they are distinguished by at least one of $\{v_0, v_1, v_{\lceil \frac{r}{2} \rceil}\}$. Therefore, S is a metric generator.

Now, let $e, f \in E(G)$ be two distinct edges of G. If at least one of them, say e, has both endpoints in S, then they are clearly distinguished by at least one endpoint of e. Suppose now that $e = uv$, with $u \in S$ and $v \in V(G) \setminus S$. If e and f are disjoint or their common endpoint is v, then they are distinguished by u. If $e = uv$ and $f = uv'$ and $v, v' \in V(C)$, then the vertex that distinguishes v and v' also distinguishes e and f. The remaining case, where e and f have no endpoints in S, is covered by Proposition 5.9. It follows that S is an edge metric generator.

To conclude the proof, we need to prove that any vertex and any edge are distinguished by at least one vertex of S. Towards contradiction, suppose that there exist $e \in E(G)$ and $v \in V(G)$, which are not distinguished by any vertex of S; in other words, for every $x \in S$ it holds that $d(e, x) = d(v, x)$. Suppose both endpoints of $e = xy$ are in S (note that it could happen that $v \in \{x, y\}$). Then e and v are distinguished by the endpoint of e that is not v, a contradiction. Suppose that both endpoints of $e = xy$ are in $V(G) \setminus S$ (again, it could be that $v \in \{x, y\}$). If $v \in S$, then e and v are distinguished by v, a contradiction. The case in which $v \notin S$ is covered by the fact that C is a smallest cycle of G and Proposition 5.9, again a contradiction. The remaining case is where $e = xy$, with $x \in S$ and $y \in V(G) \setminus S$. If v is not an endpoint of e or $v = y$, then e and v are distinguished by x, a contradiction. Finally, say $v = x$. If $x \in V(C)$ again, since C is a smallest cycle in G, at least one vertex of $\{v_0, v_1, v_{\lceil \frac{r}{2} \rceil}\}$ distinguishes the edge e and the vertex v following Proposition 5.9, a contradiction. Therefore, $x \notin V(C)$. Let $v' \in \{v_0, v_1, v_{\lceil \frac{r}{2} \rceil}\}$ be a vertex closest to y. Then $d(e, v') \leq d(y, v') \leq \frac{r}{4}$. On the other hand, since $v' \in S$, by assumption $d(v, v') = d(e, v') \leq d(y, v') \leq \frac{r}{4}$. Let $P_{v',y}$ be the shortest path in C from v' to y. Let $P_{v',v}$ be the shortest path in G from v' to v. However, then the subgraph of G induced by vertices of $P_{v',y}$ and $P_{v',v}$ admits a cycle of size at most $d(v, v') + d(y, v') + d(y, v) \leq \frac{r}{4} + \frac{r}{4} + 1 = \frac{r}{2} + 1 < r$ (the case where the two paths $P_{v',y}$ and $P_{v',v}$ have no internal vertices in common; otherwise the cycle in question is even smaller), a contradiction with the fact that r is the girth of the graph G. Since we obtained a contradiction in all cases, it follows that any vertex and any edge are distinguished by at least one vertex of S.

Upon combining all of the above, it follows that S is a mixed metric generator and the proof is completed. □

As the following examples show, the bound from Theorem 5.13 is sharp. For any cycle C_n, $\dim_{\mathrm{m}}(C_n) = n - g(C_n) + 3 = 3$. For any complete graph $\dim_{\mathrm{m}}(K_n) = n - g(K_n) + 3 = n$. For any complete bipartite graph $K_{2,t}$ we have $\dim_{\mathrm{m}}(K_{2,t}) = t + 2 - g(K_{2,t}) + 3 = t + 1$. For any graph G, such that every vertex has a maximal neighbour, the girth is $g(G) = 3$, therefore by Theorem 5.8, $\dim_{\mathrm{m}}(G) = n - g(G) + 3$.

5.4 The Complexity of the Mixed Metric Dimension Problem

Owing to the close relationship between the mixed metric dimension, edge metric dimension and the standard metric dimension, it is natural to think how computationally difficult the problem of computing the mixed metric dimension of a graph is. We already mention in Chapter 4 that the decision problem concerning the metric dimension is NP-complete. Proof that the decision problem concerning the edge metric dimension problem is NP-complete is presented in Theorem 4.26. In this section, we show that the problem of finding the mixed metric dimension of an arbitrary connected graph is NP-hard. We will use a reduction from the 3-SAT problem, as in the case of the metric dimension proof in [32] and edge metric dimension proof in [36], with slight improvements to the gadgets in the construction. We first define the decision problem for the mixed metric dimension.

MIXED METRIC DIMENSION PROBLEM (MDIM problem for short)
INSTANCE: A connected graph G of order $n \geq 3$ and an integer $2 \leq r \leq n$.
QUESTION: Is $\dim_{\mathrm{m}}(G) \leq r$?

To prove that the problem stated above is NP-complete, we make a reduction from the 3-SAT problem. We briefly described the 3-SAT problem in Section 4.3.

Theorem 5.14. *The MDIM problem is NP-complete.*

Proof. First, let us show that MDIM is in NP. For a set of vertices S guessed by a non-deterministic algorithm for the problem, one needs to check that this is a mixed metric generator. This can be checked in polynomial time. One has to compute the distances from the vertices of S to all elements (edges and vertices) and check that all pairs of distinct elements have different distance vectors with respect to the set S.

We now describe a polynomial transformation of the 3-SAT problem to the MDIM problem. Consider an arbitrary input of the 3-SAT problem, a collection $C = \{c_1, c_2, \ldots, c_m\}$ of clauses over a finite set $U = \{u_1, u_2, \ldots, u_n\}$ of Boolean variables. We shall construct a connected graph $G = (V(G), E(G))$ and set a positive integer $r \leq |V(G)|$, such that the graph G has a mixed metric generator of a size that is at most r if and only if C is satisfiable. The construction will be made up of several components augmented by some additional edges for communication between various components.

For each variable $u_i \in U$ we construct a truth-setting component $X_i = (V_i, E_i)$, with $V_i = \{T_i, F_i, a_i, b_i, c_i, d_i\}$ and $E_i = \{T_ic_i, a_ic_i, a_ib_i, b_id_i, c_id_i, d_iF_i\}$ (see Figure 5.1 for reference). The vertices T_i and F_i are the `TRUE` and `FALSE` ends of the component, respectively. Each component is connected to the rest of the graph only through these two vertices, which gives us the following proposition.

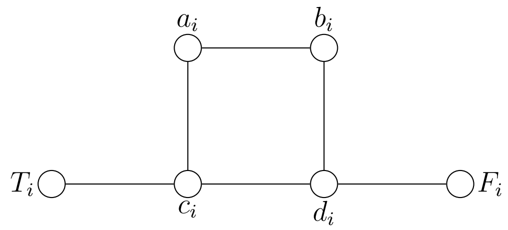


Figure 5.1: The truth-setting component for the variable u_i.

Lemma 5.15. *Let u_i be an arbitrary variable in U. Any mixed metric generator must contain at least one vertex from the set $\{a_i, b_i\}$.*

Proof. Suppose that an edge metric generator S exists without any of these vertices in it. Since the component X_i is attached to the rest of the graph only through the vertices T_i and F_i, due to the symmetry, this implies that the vertex c_i and edge a_ic_i have the same distances to all vertices in the set S, a contradiction. $\square$

For each clause $c_j \in C$ we construct a satisfaction-testing component $Y_j = (V'_j, E'_j)$, with $V'_j = \{c_j^1, \ldots, c_j^6\}$ and $E'_j = \{c_j^1c_j^2, c_j^2c_j^5, c_j^1c_j^3, c_j^2c_j^4, c_j^6c_j^3, c_j^3c_j^4\}$ (see Figure 5.2 for reference). The component is attached to the rest of the graph only through vertices c_j^1 and c_j^2 which gives us the following proposition.

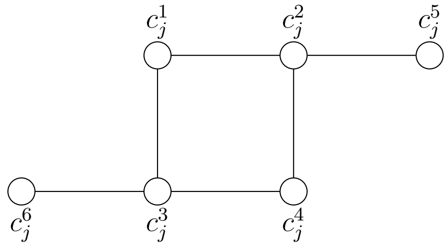


Figure 5.2: The satisfaction-testing component for clause c_j.

Lemma 5.16. *Let c_j be an arbitrary clause in C. Any mixed metric generator must contain the vertices c_j^5 and c_j^6.*

Proof. Suppose that an edge metric generator S exists without vertex c_j^5 in it. Since all the shortest paths from any vertex $x \neq c_j^5$ to the vertex c_j^2 and to the edge $c_j^2c_j^5$ go through the vertex c_j^2, this implies that the vertex c_j^2 and the edge $c_j^2c_j^5$ are at same distance to all vertices in the set S, a contradiction. A similar argument applies to the vertex c_j^6. □

We also add some edges between the truth-setting and the satisfaction-testing components as follows. If the variable u_i occurs as a positive literal in the clause c_j, then we add the edges $T_ic_j^1$ and $F_ic_j^2$. If the variable u_i occurs as a negative literal in the clause c_j, then we add the edges $T_ic_j^2$ and $F_ic_j^1$. For each clause $c_j \in C$ denote the six added edges with E''_j. We call them *the communication* edges. Figure 5.3 shows the edges that were added corresponding to the clause $c_j = (u_1 \vee \overline{u_2} \vee u_3)$, where $\overline{u_2}$ represents the negative literal of the variable u_2.

For all $k \in \{1, \ldots, n\}$, such that neither of u_k and $\overline{u_k}$ occur in the clause c_j, add the edges $T_kc_j^2$ to the graph G. For each clause $c_j \in C$ denote them with E'''_j. We call them *the neutralizing* edges, because no matter what value is assigned to the variable u_k the value of the clause c_j does not change. Equivalently, no matter which vertex v_k from the corresponding truth-setting component X_k is chosen for a mixed metric generator, it gives the same distance from such v_k to the edges $c_j^1c_j^2$

and $c_j^2 c_j^4$ from the satisfaction-testing component corresponding to the clause c_j. These two edges play an important role later in the proof.

Finally, for each clause c_j and every $k \in \{1, \ldots, m\}, k \neq j$, add the edge $c_j^2 c_k^2$ to the graph G (if it does not exist). For each clause $c_j \in C$ denote them with E_j''''. These edges ensure that the graph is connected. We call these edges *correcting* edges.

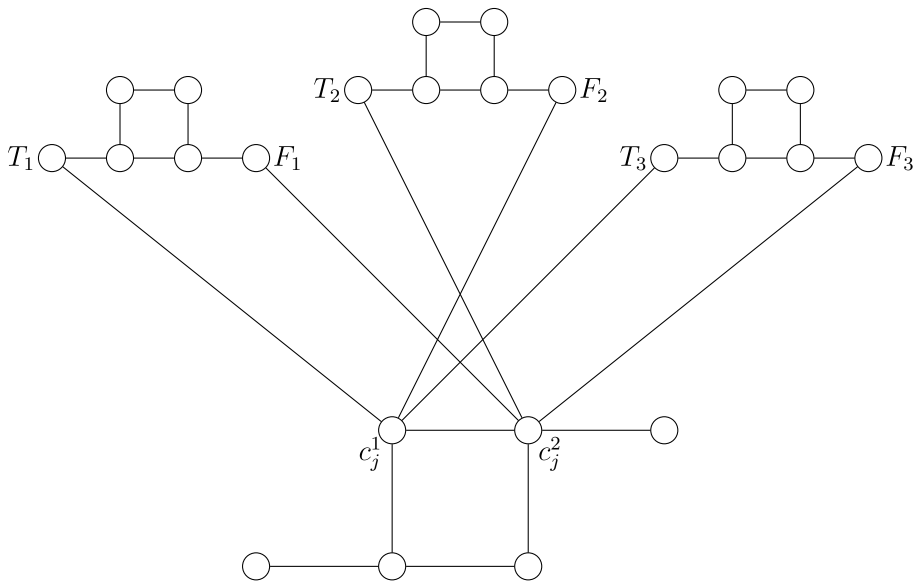


Figure 5.3: The subgraph associated with the clause $c_j = (u_1 \vee \overline{u_2} \vee u_3)$.

The construction of our instance of the MDIM problem is then completed by setting $r = 2m + n$ and $G = (V(G), E(G))$, where

$$V(G) = \left(\bigcup_{i=1}^{n} V_i \right) \cup \left(\bigcup_{j=1}^{m} V_j' \right)$$

and

$$E(G) = \left(\bigcup_{i=1}^{n} E_i \right) \cup \left(\bigcup_{j=1}^{m} \left(E_j' \cup E_j'' \cup E_j''' \cup E_j'''' \right) \right).$$

It is not hard too see that the construction can be done in polynomial time. It remains to be shown that C is satisfiable if and only if G has a mixed metric generator of size r. From Lemmas 5.15 and 5.16 we get the following.

Corollary 5.17. *The mixed metric dimension of the graph G is at least $r = 2m + n$.*

We now continue with the following lemmas which complete the proof of NP-completeness of the MDIM problem.

Lemma 5.18. *If C is satisfiable, then the mixed metric dimension of the graph G is r.*

Proof. We know that the mixed metric dimension is at least r. We now construct a mixed metric generator S of size r based on a satisfying truth assignment for C. Let $t : U \to \{\texttt{TRUE}, \texttt{FALSE}\}$ be a satisfying truth assignment for C. For each clause $c_j \in C$, put the vertices c_j^5 and c_j^6 in the set S. For each variable $u_i \in U$, put either the vertex a_i if $t(u_i) = \texttt{TRUE}$, or the vertex b_i if $t(u_i) = \texttt{FALSE}$ in the set S. We now show that S is a mixed metric generator of the graph G.

Let $e_{j,k}$ be an arbitrary correcting edge between the satisfaction testing components c_j and c_k. We notice that $e_{j,k}$ is distinguished from all other elements of the graph G by the set of vertices $\{c_j^5, c_k^5\}$, because this is the only element in the graph G that has a distance of 1 to both of the vertices c_j^5 and c_k^5.

Let $i \in \{1, \ldots, n\}$ and $j \in \{1, \ldots, m\}$ be arbitrary indices and let $v_i \in V_i \cap S$. Since we have already checked that any correcting edge is uniquely determined by some vertices in S, we do not have to check any pair of elements in which at least one correcting edge occurs. Also, one can check that each communication edge and each neutralizing edge between the truth-setting component X_i and the satisfaction-testing component Y_j is distinguished from all the remaining elements by the vertices v_i, c_j^5 and c_j^6.

Next we take a look at the elements in a truth-setting component. Let $i \in \{1, \ldots, n\}$ be an arbitrary index and let $x \in V_i \cup E_i$ be an arbitrary element from X_i. Since we have already checked that all correcting, communication, and neutralizing edges are distinguished from all other elements by some vertices from S, we only need to check that x has different distance vectors: (1) from all other elements in X_i, (2) from all elements in other truth-setting components, and (3) from all elements in the satisfaction-testing components. This is addressed next. (1) For checking that x has different distance vectors to all other elements in X_i, suppose that u_i or $\overline{u}_i$ is a literal in the clause c_j. It is not difficult to check that the vertices v_i, c_j^5 and c_j^6 distinguish the element x from all other elements in X_i. For (2), let $k \in \{1, \ldots, n\}, k \neq i$, be an arbitrary index. The vertex v_i distinguishes the element x from all elements $x' \in V_k \cup E_k$ (the elements in the truth-setting component

X_k). For (3), let $j \in \{1,\ldots,m\}$ be an arbitrary index. Hence, the vertices c_j^5 and c_j^6 distinguish the element x from all elements $y \in V_j' \cup E_j'$ (the elements in the satisfaction testing component Y_j).

Finally, we take a look at the elements from the satisfaction-testing components. Let $j \in \{1,\ldots,m\}$ be an arbitrary index. Every element of $\{c_j^2, c_j^3, c_j^5, c_j^6, c_j^2c_j^5, c_j^3c_j^6\}$ and any other element not covered in previous cases is distinguished by the set of vertices $\{c_j^5, c_j^6\}$. Let $D_1 = \{c_j^1c_j^2, c_j^2c_j^4\}$, $D_2 = \{c_j^1c_j^3, c_j^3c_j^4\}$ and $D_3 = \{c_j^1, c_j^4\}$. The set of vertices $\{c_j^5, c_j^6\}$ also distinguishes any pair of elements where one element is from D_i, for $i \in \{1,2,3\}$, and the other element is any element that has not been covered in previous cases and is not in D_i.

To complete the proof, we have to show that for any pair (x,y), where $x \neq y$ and $x,y \in D_i$, for some $i \in \{1,2,3\}$, there exists a vertex in the set S that distinguishes x and y. Since C is satisfiable, suppose that c_j is satisfied by the variable u_i. For the variable u_i there are two possibilities:

- u_i occurs as a positive literal in c_j and $t(u_i) = \texttt{TRUE}$,
- u_i occurs as a negative literal in c_j and $t(u_i) = \texttt{FALSE}$.

Thus, if $t(u_i) = \texttt{TRUE}$, then we have added the vertex a_i to the set S. In this case, the distance from a_i to the edge $c_j^1c_j^2$ is 3, while the distance to the edge $c_j^2c_j^4$ is 4. Similarly, the distance from a_i to the edge $c_j^1c_j^3$ is 3 and to the edge $c_j^3c_j^4$ it is 4. The distance from a_i to the vertex c_j^1 is 3 and to the vertex v_j^4 it is 5. The case when $t(u_i) = \texttt{FALSE}$ is symmetric.

Therefore, any two distinct elements of the graph G are distinguished by at least one vertex from the set S, and as a consequence, S is a mixed metric generator for a graph G, which completes the proof of this lemma. □

Lemma 5.19. *If the mixed metric dimension of graph G is r, then C is satisfiable.*

Proof. Let S be an arbitrary mixed metric generator of graph G with cardinality r. As seen in Lemmas 5.15 and 5.16, the set S must contain at least one vertex from the set $\{a_i, b_i\}$ for each truth-setting component X_i, and at least vertices c_j^5, c_j^6 from each satisfaction-testing component Y_j. Since the cardinality of S equals $r = 2m+n$, it follows that there is exactly one vertex from each truth-setting component and exactly two vertices from each satisfaction-testing component in the set S. We shall find a function $t : U \to \{\texttt{TRUE}, \texttt{FALSE}\}$ such that it represents a satisfying truth assignment for the collection of clauses C. For an arbitrary $i \in \{1, \dots, n\}$, let $v_i \in V_i \cap S$. Hence, we define the function t as follows:

$$t(u_i) = \begin{cases} \texttt{TRUE}, & \text{if } v_i = a_i, \\ \texttt{FALSE}, & \text{if } v_i = b_i. \end{cases}$$

We shall show that t produces a satisfying truth assignment for C. To this end, let c_j be an arbitrary clause. We claim that at least one of its literals has the value TRUE. We prove that fact by tracing which vertex from S distinguishes the edges $e_j^1 = c_j^1 c_j^2$ and $e_j^2 = c_j^2 c_j^4$, and by showing that the corresponding function t satisfies c_j.

Let $k \in \{1, \dots, m\}$ be an arbitrary index. For the clause c_k the vertices in the set S are c_k^5 and c_k^6. If $j = k$, then both edges e_j^1 and e_j^2 are at a distance of 1 from c_k^5 and at a distance of 2 from c_k^6. If $j \neq k$, then by using the correcting edges, we deduce that the edges e_j^1 and e_j^2 are at a distance of 2 from c_k^5 and at a distance of 4 from c_k^6. Therefore, none of these vertices distinguishes e_j^1 from e_j^2.

Now, consider any variable u_i, which does not occur in c_j. If $v_i = a_i$, then both edges e_j^1, e_j^2 are at a distance of 3 from v_i. If $v_i = b_i$, then both edges are at a distance of 4 from v_i. Thus, the vertex of S distinguishing the edges e_j^1, e_j^2 must belong to one of the truth-setting components that corresponds to a variable u_k, which occurs in the clause c_j. We recall that we have added communication edges in such a manner that v_k distinguishes the edges e_j^1 and e_j^2 only if one of the following statements holds true:

- u_k occurs as a positive literal in c_j and $v_k = a_k$ (in this case $t(u_k) = \texttt{TRUE}$),
- u_k occurs as a negative literal in c_j and $v_k = b_k$ (in this case $t(u_k) = \texttt{FALSE}$).

In both cases, the clause c_j is satisfied by the setting assigned to the variable u_k. As a consequence, the formula C is satisfiable, which completes the proof of this lemma. □

As a consequence of Lemmas 5.18 and 5.19 above, the polynomial transformation from 3-SAT to the MDIM problem is done, and the proof of the theorem is now completed. □

From Theorem 5.14 we immediately obtain the following result.

Corollary 5.20. *The problem of finding the mixed metric dimension of a connected graph is NP-hard.*

6 CONCLUSION

We have studied some distance-based measures and invariants of graphs. For the Hausdorff distance of graphs we have presented some results on general graphs and results on some families of graphs. We have determined exact formulae for the Hausdorff distance between two paths, between two cycles, and between a path and a cycle. A polynomial-time algorithm for determining the Hausdorff distance between two trees has been developed. It utilizes the divide and conquer problem-solving technique. As as subroutine it also uses an algorithm for maximal bipartite matching problem.

We have introduced and initiated the study of a new variant of the metric dimension of connected graphs concerned with uniquely identifying the edges of a graph, namely, the edge metric dimension. We have represented the problem of computing the edge metric dimension from a different point of view with a linear programming model. We have given some realization results on this new parameter in connection with the standard metric dimension, as well as some comparisons between both mentioned parameters. We have found that graphs exist in which the metric dimension and the edge metric dimension are in three different correlations, namely $\dim(G) = \dim_{\mathrm{e}}(G)$, $\dim(G) < \dim_{\mathrm{e}}(G)$ or $\dim(G) > \dim_{\mathrm{e}}(G)$. This, together with the fact that $\dim_{\mathrm{e}}(G) - \dim(G)$, can be arbitrarily large indicates that the structure of the edge metric generators is quite different from the structure of the metric generators. Using a polynomial reduction from the 3-SAT problem we have proved that computing the edge metric dimension of a connected graph is NP-hard. We have also computed the value of the edge metric dimension of several graph families, namely paths, cycles, complete graphs, complete bipartite graphs, trees, grids, wheels, fans, and some special cases of torus graphs. We have bounded the value of the edge metric dimension in some other cases.

We have introduced another new parameter concerning distances in graphs, namely the mixed metric dimension of a graph. It is a kind of mixed version of the metric dimension and the edge metric dimension, where not only pairs of distinct vertices and pairs of distinct edges are distinguished, but also pairs consisting of a vertex and an edge. We have begun the study of its combinatorial and computational properties. We have presented a linear programming model, which can be used to solve the problem of computing the mixed metric dimension. We have presented several tight bounds for the mixed metric dimension. We have characterized the graphs that achieve the lower bound and the upper bound for the mixed metric dimension. In addition, we have computed the exact value of the mixed metric dimension for paths, cycles, complete bipartite graphs, trees, and grids. We have given an upper bound for the mixed metric dimension in terms of the girth of the graph together with some families that achieve this bound. Finally, we have proved that this problem is computationally NP-hard.

In metric graph theory, results like this are important for the theoretical development of the field and practical applications. Introducing new invariants, such as edge metric dimension and mixed metric dimension, is important from the theoretical point of view, since they open new options for research. On the other hand, they also have direct practical applications. With the edge metric generators and mixed metric generators we can locate an intruder or a robot not only at the vertices of the network but also on the edges. The Hausdorff distance between graphs is a measure of similarity of two graphs and, therefore, offers wide options for applications. Searching for the similar molecules in a database in chemistry is an example of the application. By developing a new algorithm on graphs, our work also contributes to the theoretical aspects.

There are many open problems connected to the results presented in this dissertation. We begin with the open problems concerning the Hausdorff distance of graphs. We have presented a polynomial-time algorithm for the Hausdorff distance between two trees. The natural question that arises deals with time complexity.

Problem 6.1. *Is there an algorithm for the Hausdorff distance between two trees T_1 and T_2 with time complexity less than*

$$\mathcal{O}\left(|V(T_1)|^2 \cdot |V(T_2)|^2 \cdot \left(|V(T_1)|^{\frac{3}{2}} + |V(T_2)|^{\frac{3}{2}}\right)\right)?$$

Another question for the Hausdorff distance of graphs is about the general complexity of the problem.

Problem 6.2. *Is the problem of determining the Hausdorff distance of graphs NP-hard?*

We can look to the complexity of the problem from a different perspective and try to find efficient algorithms for some specific graphs families.

Problem 6.3. *For which graphs does a polynomial-time algorithm exist for determining the Hausdorff distance of graphs?*

The Hausdorff distance of graphs is useful for studying chemical grahps. In chemical graphs, the vertices represent atoms and the edges represent bonds. So when determining the similarity of two (chemical) graphs, it would make sense to restrict which vertices can map to each other when making an amalgam.

Problem 6.4. *Define a measure of similarity of two graphs based on the Hausdorff distance for labelled graphs with an additional restriction on which labels are allowed to map to each other.*

There are also questions that are of interest for the continuation of the research on the edge metric dimension. We have not completely answered the realization question regarding the order, metric dimension, and edge metric dimension of the graph.

Problem 6.5. *Is it possible to completely settle the realization result concerning the triplet r, t, n from Question 4.10, which was partly answered in subsection 4.2.4?*

We have compared the edge metric dimension with the metric dimension. We have found graphs that have $\dim(G) = \dim_{\mathrm{e}}(G)$, $\dim(G) < \dim_{\mathrm{e}}(G)$ and $\dim(G) >$

$\dim_e(G)$. Most of the families for which we have determined the edge metric dimension satisfied the equality $\dim(G) = \dim_e(G)$. Therefore, we post the following open problem.

Problem 6.6. *Characterize graphs G for which* $\dim_e(G) = \dim(G)$.

On the other hand, we have found only one family of graphs for which it holds that $\dim_e(G) < \dim(G)$. It is natural to wonder whether there are any others.

Problem 6.7. *Are there any other families of graphs G (different from the torus graph $C_{4r} \square C_{4t}$) such that* $\dim_e(G) < \dim(G)$*?*

We have proved that the problem of determining the edge metric dimension is NP-hard in general. The problem of computing the standard metric dimension of a graph is proved to be NP-hard even when restricted to planar graphs, and it is polynomial for the case of outerplanar graphs (see [18]). Maybe the same holds true for the edge metric dimension.

Problem 6.8. *What is the complexity of the problem of determining the edge metric dimension of a graph in the case of planar graphs and in the case of outerplanar graphs?*

For the torus graph we have only determined results for $C_{4r} \square C_{4t}$, so there are some cases that have to be studied to complete the formula for the torus graph. We have computed the edge metric dimension of some small torus graphs through exhaustive search on the computer. The edge metric dimension of all checked torus graphs is 3 or 4. Therefore, we assume that the edge metric dimension of the torus graphs is 3 or 4, where the value depends on the parity of the order of the factors.

Problem 6.9. *Complete the formula for the edge metric dimension of the torus graph $C_r \square C_t$, where $t, r \geq 3$.*

As a consequence of studying the mixed metric dimension, a number of the following open problems have arisen. Considering the close relation between the metric dimension, the edge metric dimension, and the mixed metric dimension, the following two problems arise.

Problem 6.10. *Characterize graphs G for which* $\dim_m(G) = \dim(G)$.

Problem 6.11. *Characterize graphs G for which* $\dim_{\mathrm{m}}(G) = \dim_{\mathrm{e}}(G)$.

The bound from Theorem 5.13 (the bound in terms of the girth of a graph) is achieved for several families of graphs, therefore the following problem would also be interesting to explore.

Problem 6.12. *Characterize graphs G for which the bound from Theorem 5.13 is achieved.*

Computing the (standard, edge, and mixed) metric dimension of graphs is NP-hard. Moreover, the metric dimension can be approximated within a factor of $O(\log n)$ in polynomial time, where n is the number of vertices of the graph. Similarly, the edge metric dimension can be approximated within a factor of $O(\log m)$ in polynomial time, where m is the number of edges of the graph.

Problem 6.13. *Can the mixed metric dimension be approximated within a factor of* $O(\log(n + m))$ *in polynomial time?*

The standard metric dimension has been studied for several families of graph products. The edge metric dimension has also been considered for some graph products. Therefore, the mixed metric dimension of graph products can also be investigated.

Problem 6.14. *Provide relationships between the mixed metric dimension of product graphs and that of its factors.*

A mixed metric generator is a set of vertices of a graph that uniquely distinguishes all the elements (vertices and edges) of the graph. Considering a different kind of generator might also be interesting.

Problem 6.15. *Study a different kind of a mixed metric generator in which the distinguishing elements would not only be vertices, but vertices and edges of the graph.*

In [39], the authors propose a genetic algorithm for the problem of determining the metric dimension. The genetic algorithm does not necessarily give optimal solutions but it gives satisfactory results in a reasonable amount of time. They use a linear programming model for the metric dimension to get the results with

CPLEX solver and compare them to the results they get with the genetic algorithm. In this sense, one can use a similar approach for the problem of determining the edge metric dimension or the mixed metric dimension.

Problem 6.16. *Apply some heuristics to the problem of determining the edge metric dimension or the mixed metric dimension and analyse the results.*

BIBLIOGRAPHY

[1] A. V. Aho, J. E. Hopcroft, J. D. Ullman, *The design and analysis of computer algorithms*, Addison-Wesley Longman Publishing Co., Inc. Boston, MA, USA, 1974.

[2] R. F. Bailey, P. J. Cameron, Base size, metric dimension and other invariants of groups and graphs, *Bull. Lond. Math. Soc.* 43 (2) (2011) 209–242.

[3] H.-J. Bandelt, H. M. Mulder, E. Wilkeit, Quasi-median graphs and algebras, *J. Graph Theory* 18 (7) (1994) 681–703.

[4] I. Banič, A. Taranenko, Measuring closeness of graphs – the Hausdorff distance, *Bull. Malays. Math. Sci. Soc.* 40 (1) (2017) 75–95.

[5] G. Benadé, W. Goddard, T. A. McKee, P. A. Winter, On distances between isomorphism classes of graphs, *Math. Bohem.* 116 (2) (1991) 160–169.

[6] R. C. Brigham, G. Chartrand, R. D. Dutton, P. Zhang, Resolving domination in graphs, *Math. Bohem.* 128 (1) (2003) 25–36.

[7] P. S. Buczkowski, G. Chartrand, C. Poisson, P. Zhang, On k-dimensional graphs and their bases, *Period. Math. Hungar.* 46 (1) (2003) 9–15.

[8] H. Bunke, On a relation between graph edit distance and maximum common subgraph, *Pattern Recogn. Lett.* 18 (8) (1997) 689–694.

[9] H. Bunke, K. Shearer, A graph distance metric based on the maximal common subgraph, *Pattern Recogn. Lett.* 19 (3-4) (1998) 255–259.

[10] J. Caceres, C. Hernando, M. Mora, I. M. Pelayo, M. L. Puertas, On the metric dimension of infinite graphs, *Discrete Appl. Math.* 160 (18) (2012) 2618–2626.

[11] J. Cáceres, C. Hernando, M. Mora, I. M. Pelayo, M. L. Puertas, C. Seara, D. R. Wood, On the metric dimension of some families of graphs, *Electron. Notes Discrete Math.* 22 (2) (2005) 129–133.

[12] J. Cáceres, C. Hernando, M. Mora, I. M. Pelayo, M. L. Puertas, C. Seara, D. R. Wood, On the metric dimension of Cartesian product of graphs, *SIAM J. Discrete Math.* 21 (2) (2007) 423–441.

[13] G. Chartrand, L. Eroh, M. A. Johnson, O. R. Oellermann, Resolvability in graphs and the metric dimension of a graph, *Discrete Appl. Math.* 105 (1–3) (2000) 99–113.

[14] G. Chartrand, C. Poisson, P. Zhang, Resolvability and the upper dimension of graphs, *Comput. Math. Appl.* 39 (12) (2000) 19–28.

[15] G. Chartrand, F. Saba, H. B. Zou, Edge rotations and distance between graphs, *Čas. pěst. mat.* 110 (1) (1985) 87–91.

[16] G. Chartrand, V. Saenpholphat, P. Zhang, The independent resolving number of a graph, *Math. Bohem.* 128 (4) (2003) 379–393.

[17] G. Chartrand, E. Salehi, P. Zhang, The partition dimension of a graph, *Aequationes Math.* 59 (1-2) (2000) 45–54.

[18] J. Díaz, O. Pottonen, M. Serna, E. J. van Leeuwen, Complexity of metric dimension on planar graphs, *J. Comput. System Sci.* 83 (1) (2017) 132–158.

[19] G. M. Downs, P. Willett, Similarity searching in databases of chemical structures, in: K. B. Lipkowitz, D. B. Boyd (Eds.), *Reviews in Computational Chemistry* 7, John Wiley & Sons, Inc., Hoboken, NJ, USA, 2007, 1–66.

[20] E. Duesbury, J. D. Holliday, P. Willett, Maximum Common Subgraph Isomorphism Algorithms, *MATCH Commun. Math. Comput. Chem.* 77 (2) (2017) 213–232.

[21] L. Epstein, A. Levin, G. J. Woeginger, The (weighted) metric dimension of graphs: Hard and easy cases, *Algorithmica* 72 (4) (2015) 1130–1171.

[22] P. Erdös, A. Rényi, On two problems of information theory, *Magyar Tud. Akad. Mat. Kutató Int. Közl.* 8 (1963) 229–243.

[23] A. Estrada-Moreno, J. A. Rodríguez-Velázquez, I. G. Yero, The k-metric dimension of a graph, *Appl. Math. Inf. Sci.* 9 (6) (2015) 2829–2840.

[24] X. Gao, B. Xiao, D. Tao, X. Li, A survey of graph edit distance, *Pattern Anal. Appl.* 13 (1) (2010) 113–129.

[25] M. R. Garey, D. S. Johnson, *Computers and Intractability: A Guide to the Theory of NP-Completeness*, W. H. Freeman & Co., New York, USA, 1979.

[26] F. Harary, R. A. Melter, On the metric dimension of a graph, *Ars Combin.* 2 (1976) 191–195.

[27] C. Hernando, M. Mora, C. Seara, I. M. Pelayo, D. R. Wood, Extremal graph theory for metric dimension and diameter, *Electron. J. Combin.* 17 (1) (2010) # R30.

[28] J. E. Hopcroft, R. M. Karp, An $n^{5/2}$ Algorithm for Maximum Matchings in Bipartite Graphs, *SIAM J. Comput.* 2 (4) (1973) 225–231.

[29] D. S. Johnson, Approximation Algorithms for Combinatorial Problems, *J. Comput. System Sci.* 9 (3) (1974) 256–278.

[30] M. Johnson, An ordering of some metrics defined on the space of graphs, *Czechoslovak Math. J.* 37 (1) (1987) 75–85.

[31] M. Johnson, Structure-activity maps for visualizing the graph variables arising in drug design, *J. Biopharm. Statist.* 3 (2) (1993) 203–236.

[32] S. Khuller, B. Raghavachari, A. Rosenfeld, Landmarks in graphs, *Discrete Appl. Math.* 70 (3) (1996) 217–229.

[33] A. Kelenc, Determining the Hausdorff Distance Between Trees in Polynomial Time, (2019) arXiv:1907.01299v1.

[34] A. Kelenc, D. Kuziak, A. Taranenko, I. G. Yero, Mixed metric dimension of graphs, *Appl. Math. Comput.* 314 (1) (2017) 429–438.

[35] A. Kelenc, A. Taranenko, On the Hausdorff Distance between Some Families of Chemical Graphs, *MATCH Commun. Math. Comput. Chem.* 74 (2) (2015) 223–246.

[36] A. Kelenc, N. Tratnik, I. G. Yero, Uniquely identifying the edges of a graph: the edge metric dimension, *Discrete Appl. Math.* 251 (2018) 204–220.

[37] S. Klavžar, Wiener index under gated amalgamations, *MATCH Commun. Math. Comput. Chem.* 53 (1) (2005) 181–194.

[38] J. Kratica, V. Filipovic, A. Kartelj, Edge metric dimension of some generalized Petersen graphs, (2018) arXiv:1807.00580v2.

[39] J. Kratica, V. Kovačević-Vujčić, M. Čangalović, Computing the metric dimension of graphs by genetic algorithms, *Comput. Optim. Appl.* 44:343 (2009).

[40] J.-B. Liu, Z. Zahid, R. Nasir, W. Nazeer, Edge version of metric dimension and doubly resolving sets of the necklace graph, *Mathematics* 6 (11) (2018) 243.

[41] R. A. Melter, I. Tomescu, Metric bases in digital geometry, *Computer Vision, Graphics, and Image Processing* 25 (1) (1984) 113–121.

[42] O. R. Oellermann, J. Peters-Fransen, The strong metric dimension of graphs and digraphs, *Discrete Appl. Math.* 155 (3) (2007) 356–364.

[43] F. Okamoto, B. Phinezyn, P. Zhang, The local metric dimension of a graph, *Math. Bohem.* 135 (3) (2010) 239–255.

[44] I. Peterin, I. G. Yero, Edge Metric Dimension of Some Graph Operations, *Bull. Malays. Math. Sci. Soc.* in press (2019).

[45] Y. Ramírez-Cruz, O. R. Oellermann, J. A. Rodríguez-Velázquez, The Simultaneous Metric Dimension of Graph Families, *Discrete Appl. Math.* 198 (2016) 241–250.

[46] J. W. Raymond, P. Willett, Maximum common subgraph isomorphism algorithms for the matching of chemical structures, *J. Comput. Aid. Mol. Des.* 16 (7) (2002) 521–533.

[47] P. J. Slater, Leaves of trees, Proceeding of the 6th Southeastern Conference on Combinatorics, Graph Theory, and Computing, *Congressus Numerantium* 14 (1975) 549–559.

[48] R. Trujillo-Rasua, I. G. Yero, k-Metric antidimension: A privacy measure for social graphs, *Inform. Sci.* 328 (2016) 403–417.

[49] G. Valiente, *Algorithms on Trees and Graphs*, Springer-Verlag, Berlin Heidelberg, 2010.

[50] P. Willett, Matching of Chemical and Biological Structures Using Subgraph and Maximal Common Subgraph Isomorphism Algorithms, in: D. G. Truhlar, W.J. Howe, A. J. Hopfinger, J. Blaney, R. A. Dammkoehler (Eds.), *Rational Drug Design*, Springer, New York, 1999, 11–38.

[51] I. G. Yero, On the strong partition dimension of graphs, *Electron. J. Combin.* 21 (3) (2014) # P3.14.

[52] I. G. Yero, A. Estrada-Moreno, J. A. Rodríguez-Velázquez, Computing the k-metric dimension of graphs, *Appl. Math. Comput.* 300 (2017), 60–69

[53] E. Zhu, A. Taranenko, Z. Shao, J. Xu, On graphs with the maximum edge metric dimension, *Discrete Appl. Math.* 257 (2019) 317–324.

[54] N. Zubrilina, On the edge dimension of a graph, *Discrete math.* 341 (7) (2018), 2083–2088

[55] N. Zubrilina, Asymptotic behavior of the edge metric dimension of the random graph, *Discuss. Math. Graph Theory* in press (2019).

LIST OF FIGURES

Razširjeni povzetek

Razdalja med dvema vozliščema grafa je osnovni koncept, ki ga uporabljamo v različnih invariantah in merah grafov. Definirana je kot dolžina najkrajše poti med dvema vozliščema. V doktorski disertaciji se posvetimo Hausdorffovi razdalji med grafoma, ki je bila vpeljana pred kratkim in dvema novima grafovskima invariantama - povezavni metrični dimenziji grafa in mešani metrični dimenziji grafa. Osnovne definicije vseh treh obravnavanih tem so tesno povezane z razdaljo med dvema vozliščema grafa. Motivacija za študij tem s področja metrične teorije grafov je v njihovih aplikacijah na drugih znanstvenih in strokovnih področjih. Izvirni rezultati, ki so predstavljeni v disertaciji, so bili objavljeni v člankih [33, 34, 35, 36].

Naj bo $G = (V(G), E(G))$ graf z množico vozlišč $V(G)$ in množico povezav $E(G)$. Neurejen par vozlišč $\{u, v\}$ predstavlja povezavo grafa, ki jo krajše zapišemo kar uv. Vozlišči u in v imenujemo *krajišči* povezave uv. Vozlišče u je *sosednje* z vozliščem v, če je $uv \in E(G)$. Vozlišče u je *incidenčno* s povezavo e, če je krajišče povezave e.

Naj bosta $G = (V(G), E(G))$ in $H = (V(H), E(H))$ poljubna grafa. Graf H je podgraf grafa G, če je $V(H) \subseteq V(G)$ in $E(H) \subseteq E(G)$. Podgrafu H rečemo pravi podgraf grafa G, če je $V(H) \subset V(G)$.

Vsi obravnavani grafi v disertaciji so enostavni grafi. To pomeni, da nimajo večkratnih povezav in zank ($uu \notin E(G)$ za vsako vozlišče $u \in V(G)$).

Naj bo G graf in naj bo $S \subseteq V(G)$. Z oznako $\langle S \rangle$ označimo podgraf grafa G induciran na množici vozlišč S; to je za vsaki dve vozlišči $u, v \in S$, $uv \in E(\langle S \rangle)$ natanko tedaj, ko je $uv \in E(G)$.

Grafa G_1 in G_2 sta *izomorfna*, kar označimo z $G_1 \cong G_2$, če obstaja bijektivna preslikava med njunima množicama vozlišč, ki ohranja sosednost in nesosednost vozlišč.

Pot P med vozliščem u in vozliščem v v grafu G je zaporedje $u = v_0 v_1 v_2 \ldots v_{k-1} v_k = v$ paroma različnih vozlišč grafa G, kjer je $v_i v_{i+1}$ povezava grafa G za vsak $i \in \{0, \ldots, k-1\}$. Vozlišči u in v imenujemo *krajišči* poti. *Dolžina* poti P, označimo jo z $\ell(P)$, je število povezav v P. Če poti P dodamo povezavo uv, potem dobimo *cikel*.

Ožina $g(G)$ grafa G je velikost najmanjšega cikla v grafu G.

Če sta vsaki dve različni vozlišči grafa G sosednji, potem grafu G rečemo *polni graf*. Oznaka za polni graf z n vozlišči je K_n.

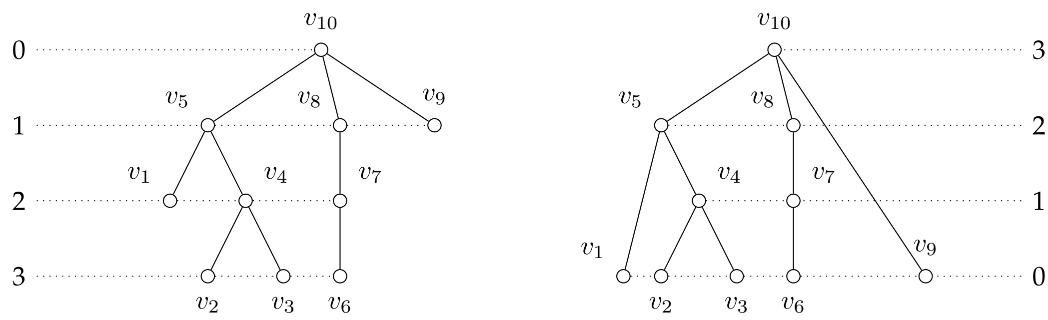
Razdalja med vozliščema u in v grafa G je dolžina najkrajše poti med vozliščema u in v. Označimo jo z $d_G(u, v)$. Razdalja med vozliščem u in podmnožico vozlišč $S \subseteq V(G)$ je definirana kot $d_G(u, S) = \min_{v \in S}\{d_G(u, v)\}$.

Graf G je *povezan*, če za vsak par vozlišč $u, v \in V(G)$ obstaja pot med u in v.

Povezanemu podgrafu H grafa G pravimo *konveksen* v grafu G, če za vsak par vozlišč $u, v \in V(H)$ poljubna najkrajša pot P med u in v v grafu G v celoti leži v H ($P \subseteq H$).

Graf $T = (V(T), E(T))$ je *drevo*, če je povezan in ne premore nobenega cikla. *Drevo s korenom* $T = (V(T), E(T))$ je drevo, ki ima posebno vozlišče $r \in V(T)$, ki mu rečemo *koren*. V drevesu s korenom je vsaka pot od korena do poljubnega vozlišča $v \in V(T)$ enolična. Drevo s korenom lahko narišemo tako, da je koren na vrhu, in potem ostala vozlišča razbijemo na nivoje glede na razdaljo do korena drevesa. *Globina* vozlišča $v \in V(T)$, označimo jo z $depth[v]$, je dolžina poti od korena do vozlišča v. Globina drevesa T pa je največja globina izmed vseh vozlišč drevesa. Vozlišču $v \in V(T)$ rečemo prednik vozlišča $u \in V(T)$, če vozlišče v leži na enolični poti med vozliščem u in korenom, pri čemer $u \neq v$. Vozlišču $v \in V(T)$ rečemo *potomec* vozlišča $u \in V(T)$, če vozlišče u leži na enolični poti med vozliščem v in korenom drevesa in $u \neq v$. Množico vseh prednikov vozlišča v označimo z $ancestors[v]$. Množico vseh potomcev vozlišča v označimo z $descendants[v]$. Vozlišče $v \in V(T)$ se imenuje *starš* vozlišča $u \in V(T)$, označimo ga s $parent[u]$, če je $vu \in E(T)$ in je vozlišče v prednik vozlišča u. V tem primeru vozlišču u rečemo *otrok* vozlišča v. Množici vozlišč $children[v] = \{u \in V(T) \mid u$ je otrok od $v\}$ rečemo *otroci* vozlišča v. Vozlišču brez otrok rečemo *list*. Dve vozlišči $v, u \in V(T)$ sta *sorojenca*, če je $parent[v] = parent[u]$. *Višina* vozlišča $v \in V(T)$, označimo jo s $height[v]$, je dolžina najdaljše poti izmed vseh poti od vozlišča v do poljubnega vozlišča v množici $\{v\} \cup descendants[v]$.

Primer 1. *Na sliki 6.1 je prikazano drevo s korenom T s korenskim vozliščem v_{10}. Drevo T je narisano dvakrat. Na levi strani je drevo T narisano glede na globino vozlišč, medtem ko je na desni strani drevo T narisano glede na višino vozlišč.*



Slika 6.1: Drevo s korenom T, narisano glede na globino (leva stran) in glede na višino (desna stran) vozlišč.

Naj bo G graf in v vozlišče grafa G. *Ekscentričnost* vozlišča v, označimo jo z $\mathrm{e}(v)$, je definirana kot $\mathrm{e}(v) = \max\{d_G(v, u) \mid u \in V(G)\}$. *Radij* grafa G, označimo ga z $\mathrm{rad}(G)$, je najmanjša ekscentričnost izmed vseh ekscentričnosti vozlišč grafa G; torej $\mathrm{rad}(G) = \min\{\mathrm{e}(v) \mid v \in V(G)\}$. *Diameter* grafa G, označimo ga z $\mathrm{diam}(G)$, je največja ekscentričnost izmed vseh ekscentričnosti vozlišč grafa G; torej $\mathrm{diam}(G) = \max\{\mathrm{e}(v) \mid v \in V(G)\}$. *Center* grafa G je množica vozlišč z najmanjšo ekscentričnostjo; torej $\mathrm{center}(G) = \{v \in V(G) \mid \mathrm{e}(v) = \mathrm{rad}(G)\}$. Vozlišču $v \in \mathrm{center}(G)$ rečemo *centralno vozlišče* grafa G. Za poljuben graf G velja, da je $\mathrm{rad}(G) \leq \mathrm{diam}(G) \leq 2 \cdot \mathrm{rad}(G)$.

Graf $G = (V(G), E(G))$ je *dvodelen*, če lahko množico vozlišč $V(G)$ razbijemo na dve množici A in B tako, da ima vsaka povezava iz $E(G)$ eno krajišče v A in drugo krajišče v B. Če med particijskima množicama A in B obstajajo vse možne povezave, potem grafu G rečemo *polni dvodelni* graf. Polni dvodelni graf označimo s $K_{r,t}$, kjer je $r = |A|$ in $t = |B|$.

Prirejanje $M \subseteq E(G)$ je množica povezav, za katero velja, da je vsako vozlišče iz $V(G)$ incidenčno z največ eno povezavo iz M. *Največje prirejanje* je prirejanje, ki vsebuje največje možno število povezav. Prirejanju rečemo *popolno*, če je vsako vozlišče grafa G krajišče neke povezave iz M. Največjemu prirejanju v dvodelnem grafu $G = (V(G), E(G))$ rečemo *največje dvodelno prirejanje*. Problem iskanja največjega dvodelnega prirejanja je rešljiv v polinomskem času. Hopcroft-Karpov algoritem [28] poišče največje dvodelno prirejanje v času $\mathcal{O}(\sqrt{|V(G)|}|E(G)|)$.

Vsa vozlišča, ki so sosednja z vozliščem v, tvorijo *odprto soseščino* $N(v)$ vozlišča v. Če odprti soseščini dodamo še samo vozlišče v, dobimo *zaprto soseščino* vozlišča v, torej $N[v] = N(v) \cup \{v\}$. Vozlišču v rečemo *simplicialno vozlišče*, če soseščina $N(v)$ inducira polni graf. Vozliščema u, v iz grafa G rečemo *neprava dvojčka*, če imata enaki odprti soseščini, to je $N(u) = N(v)$. Vozliščema u, v rečemo *prava dvojčka*, če imata enaki zaprti soseščini, torej $N[u] = N[v]$. Vozlišče v je *pravi dvojček*, če obstaja tako vozlišče $u \neq v$, da sta u in v prava dvojčka. Podobno je vozlišče v *nepravi dvojček*, če obstaja tako vozlišče $u \neq v$, da sta u in v neprava dvojčka.

Kartezični produkt grafov G in H je graf z oznako $G \Box H$, katerega množica vozlišč je $V(G \Box H) = \{(a, b) \mid a \in V(G) \wedge b \in V(H)\}$ in v katerem sta dve vozlišči (a, b) in (c, d) sosednji natanko tedaj, ko velja

- $a = c$ in $bd \in E(H)$ ali
- $b = d$ in $ac \in E(G)$.

Naj bo $h \in V(H)$ poljubno vozlišče grafa H. V kartezičnem produktu grafov G in H množica $V(G) \times \{h\}$ predstavlja G-sloj. Podobno množica $\{g\} \times V(H)$, kjer je $g \in V(G)$, predstavlja H-sloj. Neki konkreten G-sloj označimo z G^h. Podobno neki konkreten H-sloj označimo z ${}^gH$. Za podgraf, ki je induciran z nekim G-slojem, velja, da je izomorfen grafu G. Prav tako za podgraf, ki je induciran z nekim H-slojem, velja, da je izomorfen grafu H.

Stik grafov G in H je graf, ki ga dobimo iz grafov G in H tako, da dodamo vse možne povezave, v katerih je eno krajišče poljubno vozlišče grafa G, drugo krajišče pa poljubno vozlišče grafa H.

Hausdorffova razdalja med grafoma

Za merjenje podobnosti dveh objektov moramo primerjana objekta najprej modelirati z ustreznim orodjem. V ta namen se pogosto uporabljajo grafi. Podobnost grafov izmerimo na podlagi mere, ki določa, kako daleč narazen sta dva grafa. Obstaja več različnih mer za merjenje podobnosti grafov.

Hausdorffovo razdaljo med grafoma sta leta 2014 vpeljala Banič in Taranenko [4]. Je mera, ki temelji na posebnem skupnem podgrafu primerjanih grafov, ki ga določimo na podlagi strukturnih lastnosti izven samega skupnega podgrafa. Za definicijo Hausdorffove razdalje potrebujemo naslednje definicije iz [4].

Definicija 2. *[4] Naj bo G poljuben graf. Hausdorffov graf grafa G, označimo ga z 2^G, ima za množico vozlišč $V(2^G)$ množico vseh nepraznih podgrafov grafa G. Sosednost vozlišč grafa 2^G je definirana na naslednji način. Za vsaki $H_1, H_2 \in V(2^G)$, $H_1 \neq H_2$ velja, da je $H_1H_2 \in E(2^G)$ natanko tedaj, ko*

1. *za vsako vozlišče $v \in V(H_1)$ obstaja tako vozlišče $v' \in V(H_2)$, da je $d_G(v, v') \leq 1$ in*
2. *za vsako vozlišče $v' \in V(H_2)$ obstaja tako vozlišče $v \in V(H_1)$, da je $d_G(v', v) \leq 1$.*

Hausdorffova metrika h_G med dvema podgrafoma grafa G je definirana v sledeči definiciji. Pove nam, v kolikšni meri dva podgrafa sovpadata.

Definicija 3. *[4] Naj bo G poljuben graf.* Razdalja med dvema podgrafoma H_1 in H_2 grafa G, *označimo jo s $h_G(H_1, H_2)$, je razdalja med njunima pripadajočima vozliščema v 2^G. Z drugimi besedami,*

$$h_G(H_1, H_2) := d_{2^G}(H_1, H_2).$$

Razdalji h_G bomo rekli Hausdorffova metrika na 2^G.

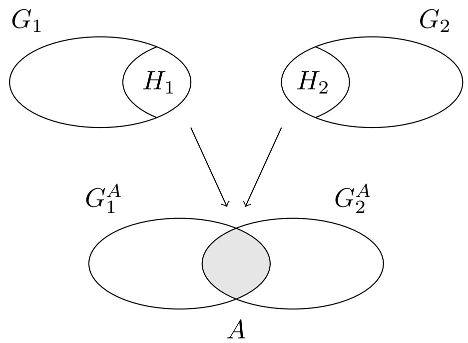
V [4] je dokazana tudi naslednja posledica.

Posledica 4. *[4] Če je G povezan graf, potem je h_G metrika na $V(2^G)$.*

Za definicijo Hausdorffove razdalje na razredu vseh enostavnih povezanih grafov kot mero podobnosti dveh takšnih grafov, moramo vpeljati tudi amalgame [3, 37].

Definicija 5. *Naj bosta H_1 (konveksen) podgraf grafa G_1 in H_2 (konveksen) podgraf grafa G_2. Če sta H_1 in H_2 izomorfna, potem poljubnemu grafu A, ki ga dobimo iz G_1 in G_2 z identifikacijo njunih podgrafov H_1 in H_2, rečemo (konveksen) amalgam grafov G_1 in G_2. Izomorfnima kopijama grafov G_1 in G_2 v A rečemo pokritji amalgama A in ju označimo z G_1^A in G_2^A.*

Na sliki 6.2 je shematsko prikazano, kako iz dveh grafov naredimo amalgam.



Slika 6.2: Amalgam A od G_1 in G_2.

Označimo z $\mathcal{A}(G_1, G_2)$ množico vseh amalgamov in z $\mathcal{X}(G_1, G_2)$ množico vseh konveksnih amalgamov grafov G_1 in G_2.

Naj bo $\mathcal{G}$ razred vseh enostavnih povezanih grafov.

Izrek 6. *[4, Izrek 4.10] Naj bosta $G_1, G_2 \in \mathcal{G}$. Naj bo d nenegativno celo število in A amalgam grafov G_1 in G_2. Potem je $h_A(G_1^A, G_2^A) = d$ natanko tedaj, ko*

(i) za vsak $u \in V(G_1^A)$ obstaja tako vozlišče $v \in V(G_2^A)$, da je $d_A(u, v) \le d$,

(ii) za vsak $u \in V(G_2^A)$ obstaja tako vozlišče $v \in V(G_1^A)$, da je $d_A(u, v) \le d$ in

(iii) obstaja tak $u \in V(G_1^A)$, da je za vsak $v \in V(G_1^A \cap G_2^A)$ razdalja $d_A(u, v) \ge d$ ali obstaja tak $u \in V(G_2^A)$, da je za vsak $v \in V(G_1^A \cap G_2^A)$ razdalja $d_A(u, v) \ge d$.

Iz Izreka 6 dobimo naslednjo posledico.

Posledica 7. *Naj bosta $G_1, G_2 \in \mathcal{G}$. Naj bo A amalgam grafov G_1 in G_2. Potem velja*

$$h_A(G_1^A, G_2^A) = \max_{u \in V(A)} \left\{ d_A(u, G_1^A \cap G_2^A \right\}.$$

Posledica 7 nam pove, da je za določitev $h_A(G_1^A, G_2^A)$ dovolj najti vozlišče $v \in V(A)$, ki ima največjo razdaljo do preseka amalgama $G_1^A \cap G_2^A$, saj velja $h_A(G_1^A, G_2^A) = d_A(v, G_1^A \cap G_2^A)$.

Hausdorffovo razdaljo $\mathcal{H} : \mathcal{G} \times \mathcal{G} \to \mathbb{R}$ na $\mathcal{G}$ definiramo na naslednji način:

Definicija 8. *[4] Za poljubna grafa $G_1, G_2 \in \mathcal{G}$ je*

$$\mathcal{H}(G_1, G_2) = \begin{cases} \min\left\{h_A(G_1^A, G_2^A) \mid A \in \mathcal{X}(G_1, G_2)\right\}, & \text{če } G_1 \ncong G_2 \\ 0, & \text{če } G_1 \cong G2. \end{cases}$$

Oznaki $\mathcal{H}$ rečemo Hausdorffova razdalja *na $\mathcal{G}$.*

Sledi predstavitev izvirnih rezultatov objavljenih v [33, 35].

Konveksnemu amalgamu A dveh enostavnih povezanih grafov G_1 in G_2, za katera je $h_A(G_1^A, G_2^A) = \mathcal{H}(G_1, G_2)$, rečemo *optimalni amalgam*.

Za določitev Hausdorffove razdalje med grafoma G_1 in G_2 iz $\mathcal{G}$ moramo poiskati optimalni amalgam. Če imamo konveksni skupni podgraf grafov G_1 and G_2, potem lahko tvorimo amalgam grafov G_1 in G_2. Torej, poiskati moramo takšen konveksni skupni podgraf grafov G_1 in G_2, da je razdalja med pokritjema G_1^A in G_2^A pripadajočega amalgama A najmanjša možna.

Za dva poljubna enostavna povezana grafa lahko zgornjo mejo za njuno Hausdorffovo razdaljo izrazimo s pomočjo radija grafov.

Izrek 9. *Naj bosta G_1 in G_2 poljubna enostavna povezana grafa. Potem je*

$$\mathcal{H}(G_1, G_2) \leq \max\{\mathrm{rad}(G_1), \mathrm{rad}(G_2)\}.$$

To mejo dosežemo, če je en graf trivialen (graf na enem vozlišču).

Med grafi, ki se pogosto pojavljajo v kemijski teoriji grafov, so tudi poti in cikli. Za te grafe lahko izrazimo Hausdorffovo razdaljo z njihovim številom vozlišč.

Trditev 10. *Če je P_n pot na n vozliščih in P_m pot na m vozliščih, kjer je $n \geq m \geq 1$, potem je $\mathcal{H}(P_n, P_m) = \left\lceil \frac{n-m}{2} \right\rceil$.*

Če je C_n cikel na n vozliščih, pri čemer je $n \geq 3$, potem je največji konveksni podgraf cikla C_n pot na $\lceil \frac{n}{2} \rceil$ vozliščih.

Trditev 11. *Če je P_n pot na n vozliščih in C_m cikel na m vozliščih, kjer je $n \geq 1$ in $m \geq 3$, potem je*

$$\mathcal{H}(P_n, C_m) = \begin{cases} \left\lceil \frac{m-n}{2} \right\rceil, & \textit{če je } n \leq \frac{m}{2} \\ \left\lceil \frac{m-1}{4} \right\rceil, & \textit{če je } \frac{m}{2} < n \leq m \\ \left\lceil \frac{n - \lceil \frac{m}{2} \rceil}{2} \right\rceil, & \textit{če je } n > m. \end{cases}$$

Hausdorffova razdalja med dvema izomorfnima cikloma je po definiciji enaka 0. Za neizomorfne cikle pa velja naslednja trditev.

Trditev 12. *Če je C_n cikel na n vozliščih in C_m cikel na m vozliščih, kjer je $n > m \geq 3$, potem je $\mathcal{H}(C_n, C_m) = \left\lceil \frac{n - \lceil \frac{m}{2} \rceil}{2} \right\rceil$.*

Drevesa se pogosto pojavljajo v kemijski teoriji grafov. Veliko organskih molekul lahko predstavimo z grafi, ki so drevesa. S Hausdorffovo razdaljo med drevesoma lahko merimo podobnost dveh dreves glede na njuno strukturo in s tem tudi podobnost dveh molekul. To nam daje motivacijo, da podamo učinkovit algoritem za računanje Hausdorffove razdalje med dvema drevesoma.

Najprej predstavimo nekaj lastnosti povezanih s Hausdorffovo razdaljo med drevesi.

Izrek 13. *Naj bosta T_1 in T_2 poljubni netrivialni drevesi, za kateri velja $\mathrm{diam}(T_1) \geq \mathrm{diam}(T_2)$. Naj bo $c \in \mathrm{center}(T_1)$. Potem za vsak optimalni amalgam $A \in \mathcal{X}(T_1, T_2)$ velja, da je $\{c^A\} \subseteq V(T_1^A \cap T_2^A)$.*

Naj bo G graf in H njegov podgraf z lastnostjo P. Grafu H rečemo *minimalni podgraf* z lastnostjo P, če ne obstaja pravi podgraf grafa H z lastnostjo P.

Izrek 14. *Naj bosta T_1 in T_2 poljubni netrivialni drevesi, za kateri velja $\mathrm{diam}(T_1) \geq \mathrm{diam}(T_2)$. Naj bo $0 \leq k \leq rad(T_1)$ fiksirano celo število. Naj bo H minimalno poddrevo drevesa T_1, ki vsebuje centralno vozlišče drevesa T_1 in ima lastnost $\max_{u \in V(T_1) \setminus V(H)} \{d_{T_1}(u, H)\} \leq k$. Če T_2 ne premore podgrafa izomorfnega grafu H, potem je $\mathcal{H}(T_1, T_2) > k$.*

Naslednja trditev pove, kako daleč narazen sta lahko centra obeh dreves, ki ju primerjamo.

Trditev 15. *Naj bosta T_1 in T_2 poljubni netrivialni drevesi, za kateri velja* $\mathrm{diam}(T_1) \geq \mathrm{diam}(T_2)$*. Naj bo $A \in \mathcal{X}(T_1, T_2)$ optimalni amalgam dreves T_1 in T_2. Potem obstajata takšna centra $c_1 \in \mathrm{center}(T_1)$ in $c_2 \in \mathrm{center}(T_2)$, da je $d_A(c_1^A, c_2^A) \leq \mathcal{H}(T_1, T_2)$.*

Meja iz Trditve 15 je dosegljiva, kar zapišemo v obliki naslednje trditve.

Trditev 16. *Za poljubno nenegativno celo število k obstajata drevesi T_1 in T_2 s takšnima lastnostima* $\mathrm{diam}(T_1) \geq \mathrm{diam}(T_2)$ *in $\mathcal{H}(T_1, T_2) = k$, da za vsak optimalni amalgam A dreves T_1 in T_2 velja $d_A(c_1^A, c_2^A) = \mathcal{H}(T_1, T_2)$, kjer je $c_1 \in \mathrm{center}(T_1)$ in $c_2 \in \mathrm{center}(T_2)$.*

Algoritem za izračun Hausdorffove razdalje med drevesoma deluje s pomočjo tako imenovanih *skupnih poddreves od zgoraj navzdol*, in zato potrebujemo naslednje definicije povzete po [49].

Definicija 17. *Naj bo $T = (V(T), E(T))$ drevo s korenom.* Poddrevo *drevesa T je povezan podgraf drevesa T.* Poddrevo od zgoraj navzdol *$S = (V(S), E(S))$ je poddrevo drevesa T, ki ima koren in za katerega velja $parent[v] \in V(S)$ za vsa nekorenska vozlišča $v \in V(S)$. Korensko vozlišče poddrevesa od zgoraj navzdol S je vedno korensko vozlišče drevesa T. Naj bo $u \in V(T)$. Poddrevo drevesa T, ki je inducirano na množici vozlišč $\{u\} \cup descendants[u]$, imenujemo* poddrevo s korenom *u.*

Definicija 18. *Drevesi s korenom $T_1 = (V(T_1), E(T_1))$ in $T_2 = (V(T_2), E(T_2))$ sta* izomorfni, *če med njima obstaja bijekcija $M \subseteq V(T_1) \times V(T_2)$, za katero velja $(koren[T_1], koren[T_2]) \in M$ in $(parent[v], parent[u]) \in M$ za vsa nekorenska vozlišča $v \in V(T_1)$ in $u \in V(T_2)$ z lastnostjo $(v, u) \in M$. Množici M rečemo* drevesni izomorfizem s korenom.

Definicija 19. Skupno poddrevo od zgoraj navzdol *drevesa s korenom $T_1 = (V(T_1), E(T_1))$ in drevesa s korenom $T_2 = (V(T_2), E(T_2))$ je struktura (S_1, S_2, M), kjer je $S_1 = (V(S_1), E(S_1))$ poddrevo od zgoraj navzdol drevesa T_1, $S_2 = (V(S_2), E(S_2))$ poddrevo od zgoraj navzdol drevesa T_2 in $M \subseteq V(S_1) \times V(S_2)$ drevesni izomorfizem s korenom dreves S_1 in S_2.*

Spomnimo se, da je za določitev Hausdorffove razdalje med dvema drevesoma potrebno poiskati takšen konveksen skupni podgraf oziroma poddrevo, da je razdalja med pokritjema pripadajočega amalgama najmanjša možna.

Konveksni amalgam dreves T_1 in T_2 je drevo. Če amalgamu določimo korensko vozlišče v nekem vozlišču iz preseka amalgama $v^A \in V(T_1^A \cap T_2^A)$, potem bo presek amalgama njegovo poddrevo od zgoraj navzdol. Poddrevesi dreves T_1 in T_2, ki določata amalgam A, pa sta poddrevesi od zgoraj navzdol dreves T_1 in T_2 s korenskim vozliščem v^A.

Vsak optimalni amalgam lahko dobimo tako, da poiščemo ustrezni poddrevesi od zgoraj navzdol vhodnih dreves. Iz tega razloga naš algoritem deluje s pomočjo skupnih poddreves od zgoraj navzdol in zato je potrebno obema vhodnima drevesoma določiti korensko vozlišče.

Optimalni amalgam od zgoraj nazvdol je amalgam, ki je optimalen glede na korensko vozlišče, kar pomeni, da je pripadajoči izomorfizem drevesni izomorfizem s korenom. Skupnemu poddrevesu od zgoraj navzdol rečemo optimalno, če je pripadajoči amalgam optimalni amalgam od zgoraj navzdol.

Zdaj smo definirali vse potrebno za predstavitev Algoritma 1. Algoritem določi Hausdorffovo razdaljo med dvema poljubnima drevesoma v polinomskem času. Zraven vrednosti za razdaljo algoritem vrača tudi skupno poddrevo, ki določa optimalni amalgam.

Algoritem za svoje delovanje uporablja dve pomembnejši funkciji.

Prva funkcija se imenuje `OptimalnoSkupnoPoddrevoOdZgorajNavzdol`. Z njo izračunamo razdaljo med pokritjema optimalnega amalgama od zgoraj navzdol za dve drevesi s korenom. To funkcijo pokličemo večkrat z različnimi drevesi s korenom v vhodnih podatkih. Funkcija dela po principu "deli in vladaj". Optimalno skupno poddrevo med vhodnima drevesoma s korenom konstruiramo tako, da razbijemo originalni drevesi s korenom na manjša poddrevesa s korenom in potem med njimi iščemo optimalna skupna poddrevesa s korenom. Funkcija začne delovati v korenskih vozliščih vhodnih dreves in nato s pomočjo rekurzije razbija drevesa, dokler ne pridemo do elementarnih poddreves, za katera znamo določiti optimalno skupno poddrevo s korenom. Na poti nazaj združujemo delne

Algoritem 1: HausdorffovaRazdaljaMedDrevesoma

Vhod: Poljubni drevesi T_1 in T_2, kjer je $\mathrm{diam}(T_1) \geq \mathrm{diam}(T_2)$.
Izhod: Hausdorffova razdalja med T_1 in T_2 shranjena v hd in pripadajoče skupno poddrevo shranjeno v M.

$hd \leftarrow \infty$
$O \leftarrow \emptyset$
$r_1 \in center(T_1)$
Izračunaj višine vozlišč drevesa T_1 s korenom v r_1
foreach $u \in V(T_2)$ **do**
    $M' \leftarrow \emptyset$
    Izračunaj višine vozlišč drevesa T_2 s korenom v u
    $d \leftarrow$ `OptimalnoSkupnoPoddrevoOdZgorajNavzdol`(T_1,r_1,T_2,u,M')
    if $d < hd$ **then**
        $hd \leftarrow d$
        $r_2 \leftarrow u$
        $O \leftarrow M'$
$M \leftarrow \emptyset$
`RekonstrukcijaPreslikave`(T_1,r_1,r_2,O,M)

rešitve. To naredimo s pomočjo polnih dvodelnih grafov in največjih prirejanj v dvodelnih grafih.

Druga pomembnejša funkcija algoritma je `RekonstrukcijaPreslikave`. Ta funkcija rekonstruira izomorfizem poddreves, ki pripada optimalnemu amalgamu. Rekonstrukcija poteka s pomočjo informacij, ki jih dobimo tekom prve funkcije, ko združujemo delne rešitve v večje.

V zvezi z Algoritmom 1 sta v disertaciji dokazana naslednja izreka.

Izrek 20. *Algoritem 1 določi Hausdorffovo razdaljo med vhodnima drevesoma in poišče pripadajoči izomorfizem skupnega poddrevesa M.*

Izrek 21. *Naj bosta $T_1 = (V(T_1), E(T_1))$ in $T_2 = (V(T_2), E(T_2))$ vhodni drevesi Algoritma 1, za kateri velja $\mathrm{diam}(T_1) \geq \mathrm{diam}(T_2)$. Časovna zahtevnost Algoritma 1 je omejena z*

$$\mathcal{O}\left(|V(T_1)|^2 \cdot |V(T_2)|^2 \cdot \left(|V(T_1)|^{\frac{3}{2}} + |V(T_2)|^{\frac{3}{2}}\right)\right).$$

Povezavna metrična dimenzija

Naj bo podan graf $G = (V(G), E(G))$ z vsaj dvema vozliščema. Razdalja med vozliščem $v \in V(G)$ in povezavo $e = uw \in E(G)$ je definirana kot $d_G(e, v) = \min\{d_G(u, v), d_G(w, v)\}$. Pravimo, da vozlišče $w \in V(G)$ *razlikuje* povezavi $e_1, e_2 \in E(G)$, če $d_G(w, e_1) \neq d_G(w, e_2)$. Neprazna množica vozlišč S povezanega grafa G je *povezavni metrični generator* grafa G, če neko vozlišče iz množice S razlikuje vsaki dve različni povezavi grafa G. Moči najmanjšega povezavnega metričnega generatorja grafa G rečemo *povezavna metrična dimenzija* in jo označimo z $\dim_{\mathrm{e}}(G)$. *Povezavna metrična baza* grafa G je povezavni metrični generator grafa G, ki ima moč $\dim_{\mathrm{e}}(G)$.

Za poljubno vozlišče v grafa G je množica $V(G)\setminus\{v\}$ povezavni metrični generator. Po drugi strani moramo v vsakem povezavnem metričnem generatorju imeti vsaj eno vozlišče. Iz tega sledita naravni meji za povezavno metrično dimenzijo grafa.

Trditev 22. *Za poljuben graf G na n vozliščih velja*

$$1 \leq \dim_{\mathrm{e}}(G) \leq n - 1.$$

Grafi, ki dosežejo spodnjo mejo, so poti. Za zgornjo mejo so karakterizacijo naredili drugi avtorji. Zubrilina je te grafe karakterizirala v [54]. Zhu in drugi [53] so neodvisno naredili karakterizacijo grafov, ki dosežejo zgornjo mejo za povezavno metrično dimenzijo s pomočjo komplementa grafa.

Tudi vse vrednosti med obema mejama so zavzete ne glede na red grafa.

Trditev 23. *Za poljubni celi števili n in r, kjer je $1 \leq r \leq n - 1$, obstaja povezan graf G na n vozliščih, za katerega je $\dim_{\mathrm{e}}(G) = r$.*

Problem obstoja grafa z določeno vrednostjo za povezavno metrično dimenzijo postane težji, če zraven dodamo še vrednost za metrično dimenzijo.

Vprašanje 24. *Naj bodo r, t in n poljubna cela števila, za katera velja $1 \leq r, t \leq n - 1$. Ali obstaja povezan graf G na n vozliščih, za katerega je $\dim(G) = r$ in $\dim_{\mathrm{e}}(G) = t$?*

Podajmo nekatere družine grafov, v katerih sta $\dim(G)$ in $\dim_e(G)$ v različnih razmerjih, da dobimo občutek o obstoju grafov iz vprašanja 24. Najprej bomo podali družine grafov G, za katere velja $\dim(G) = \dim_e(G)$.

Trditev 25. *Za poljubno celo število $n \geq 2$, $\dim_e(P_n) = \dim(P_n) = 1$, $\dim_e(C_n) = \dim(C_n) = 2$ in $\dim_e(K_n) = \dim(K_n) = n - 1$. Še več, $\dim_e(G) = 1$ natanko tedaj, ko je G pot na n vozliščih.*

Trditev 26. *Za poljuben polni dvodelni graf $K_{r,t}$, različen od $K_{1,1}$, velja $\dim_e(K_{r,t}) = \dim(K_{r,t}) = r + t - 2$.*

Še ena družina grafov, za katero sta metrična dimenzija in povezavna metrična dimenzija enaki, so drevesa. Za predstavitev tega rezultata potrebujemo naslednjo terminologijo iz [32].

Naj bo $T = (V(T), E(T))$ drevo in naj bo $v \in V(T)$. Ekvivalenčno relacijo R_v na množici $E(T)$ definirajmo na naslednji način: za vsaki dve povezavi e, f naj bo eR_vf natanko tedaj, ko obstaja pot v T, ki vsebuje povezavi e, f in ne vsebuje vozlišča v, razen morda na krajišču poti. Podgrafe, inducirane s povezavami ekvivalenčnih razredov množice $E(T)$, imenujemo *mostovi* drevesa T glede na v. Mostovom drevesa T glede na vozlišče v, ki so poti, rečemo *noge* vozlišča v. Z oznako l_v označimo število nog glede na v.

Trditev 27. *Naj bo $T = (V(T), E(T))$ drevo. Če T ni pot, potem je*

$$\dim_e(T) = \dim(T) = \sum_{v \in V,\ l_v > 1} (l_v - 1).$$

Mreža je kartezični produkt dveh poti P_r in P_t.

Trditev 28. *Če je graf G mreža $P_r \square P_t$, kjer je $r \geq t \geq 2$, potem je $\dim_e(G) = \dim(G) = 2$.*

Obstajajo tudi družine grafov z neenakostjo med $\dim(G)$ in $\dim_e(G)$. Neenakosti $\dim(G) < \dim_e(G)$ med drugimi zadoščata naslednji dve družini.

Graf *kolo* $W_{1,n}$ je izomorfen $C_n \vee K_1$, kjer operator $(\vee)$ predstavlja stik grafov. Za metrično dimenzijo velja (glej [7])

$$\dim(W_{1,n}) = \begin{cases} 3, & \text{če je } n = 3, 6, \\ 2, & \text{če je } n = 4, 5, \\ \left\lfloor \frac{2n+2}{5} \right\rfloor, & \text{če je } n \geq 6. \end{cases}$$

Povezavna metrična dimenzija kolesa je strogo večja od metrične dimenzije, razen v primeru $W_{1,3}$.

Trditev 29. *Če je $W_{1,n}$ kolo, potem je*

$$\dim_e(W_{1,n}) = \begin{cases} n, & \textit{če je } n = 3, 4, \\ n - 1, & \textit{če je } n \geq 5. \end{cases}$$

Podobno kot kolo definiramo tudi *pahljačo* $F_{1,n}$, ki je izomorfna $P_n \vee K_1$. Za pahljače velja (glej [11])

$$\dim(F_{1,n}) = \begin{cases} 1, & \text{če je } n = 1, \\ 2, & \text{če je } n = 2, 3, \\ 3, & \text{če je } n = 6, \\ \left\lfloor \frac{2n+2}{5} \right\rfloor, & \text{sicer.} \end{cases}$$

Povezavna metrična dimenzija pahljače je prav tako strogo večja od metrične dimenzije pahljače z izjemama $F_{1,n}$ za vsak $n \in \{1, 2\}$.

Trditev 30. *Če je $F_{1,n}$ pahljača, potem je*

$$\dim_e(F_{1,n}) = \begin{cases} n, & \textit{če je } n = 1, 2, 3, \\ n - 1, & \textit{če je } n \geq 4. \end{cases}$$

Najtežja izmed vseh možnosti je neenakost $\dim_e(G) < \dim(G)$.

Metrična dimenzija kartezičnega produkta več družin je bila izračunana v [12]. V tem članku je tudi rezultat za kartezični produkt dveh ciklov

$$\dim(C_r \Box C_t) = \begin{cases} 4, & \text{če sta } r \text{ in } t \text{ soda}, \\ 3, & \text{sicer.} \end{cases}$$

V posebnih primerih kartezičnega produkta $C_r \Box C_t$ velja, da je $\dim_e(C_r \Box C_t) < \dim(C_r \Box C_t)$.

Izrek 31. *Za poljuben par pozitivnih celih števil r in t je* $\dim_e(C_{4r} \square C_{4t}) = 3$.

S tem smo pokazali, da za vse tri možnosti $\dim(G) = \dim_e(G)$, $\dim(G) < \dim_e(G)$ ali $\dim_e(G) < \dim(G)$ obstajajo grafi G, in zato je potrebno Vprašanje 24 (ki obravnava trojico r, t, n: metrično dimenzijo, povezavno metrično dimenzijo in red grafa) obravnavati ločeno glede na te tri možnosti.

Primer $\dim(G) = \dim_e(G)$ realiziramo s polnimi grafi ali drevesi. Trojico $n-1, n-1, n$ realiziramo s polnim grafom K_n, trojico r, r, n, kjer je $1 \leq r \leq n-2$, pa realiziramo z drevesom T z $r+1$ listi, ki ga dobimo iz zvezde $S_{1,n-1}$ tako, da odstranimo $n-1-r$ povezav grafa $S_{1,n-1}$ in subdividiramo eno od preostalih povezav z $n-1-r$ vozlišči. Red drevesa T je n in po Trditvi 27 je $\dim(T) = \dim_e(T) = r$.

Nadaljujemo s primerom, ko je $\dim(G) < \dim_e(G)$. Najprej opazimo, da za trojico $1, t, n$, kjer je $t \geq 2$, ne obstaja graf G, saj je $\dim(G) = 1$ natanko tedaj, ko je G pot P_n in za pot velja $\dim_e(P_n) = 1$. V naslednjem izreku predpostavimo, da je $2r \leq n-2$.

Izrek 32. *Za poljubna cela števila r, t in n, kjer velja $2 \leq r < t \leq 2r \leq n-2$, obstaja povezan graf G na n vozliščih, za katerega je* $\dim(G) = r$ *in* $\dim_e(G) = t$.

Še več, razlika med povezavno metrično dimenzijo in metrično dimenzijo je lahko poljubno velika.

Trditev 33. *Za poljubno celo število $q \geq 1$ obstaja povezan graf G, tak, da je* $\dim_e(G) - \dim(G) \geq q$.

Ostane nam še del primera $\dim(G) < \dim_e(G)$, ko je $2r < t \leq n-2$. Ta del primera puščamo kot odprt problem.

Problem 34. *Ali obstaja povezan graf G na n vozliščih, za katerega je* $\dim(G) = r$ *in* $\dim_e(G) = t$, *kjer so r, t, n poljubna cela števila, za katere je $r \geq 2$ in $2r < t \leq n-2$?*

Na koncu še ostane primer, ko je $\dim_e(G) < \dim(G)$. Za to neenakost nismo našli drugega primera kot kartezični produkt dveh ciklov, katerih število vozlišč je deljivo s štiri. Zaradi tega podajamo naslednji odprt problem.

Problem 35. *Naj bodo podana tri poljubna cela števila r, t in n, kjer je $2 \leq t < r \leq n-2$. Ali obstaja povezan graf G na n vozliščih, za katerega je* $\dim(G) = r$ *in* $\dim_e(G) = t$*?*

Druga možnost bi bila, da poiščemo mejo za $\dim_e(G)$ glede na $\dim(G)$ za vsak povezan graf G ob predpostavki, da je $\dim_e(G) < \dim(G)$. Na primer, če je G kartezični produkt dveh ciklov $C_{4r} \Box C_{4t}$, potem je $3 = \dim_e(G) = 4 - 1 = \dim(G) - 1$.

Problem 36. *Ali obstaja konstanta c, takšna, da je $\dim_e(G) \geq \dim(G) - c$ za vsak povezan graf G?*

Zanima nas tudi, kakšna je zahtevnost problema izračuna povezavne metrične dimenzije grafa. Odločitveni problem za metrično dimenzijo grafa je eden od klasičnih NP-polnih problemov, ki so predstavljeni v knjigi [25].

Odločitveni problem za povezavno metrično dimenzijo definiramo na naslednji način:

PROBLEM POVEZAVNE METRIČNE DIMENZIJE
INSTANCA: Povezan graf G na $n \geq 3$ vozliščih in celo število $1 \leq r \leq n-1$.
VPRAŠANJE: Ali je $\dim_e(G) \leq r$?

Za njega dokažemo naslednji izrek.

Izrek 37. *PROBLEM POVEZAVNE METRIČNE DIMENZIJE je NP-poln.*

Neposredno iz Izreka 37 dobimo spodnji rezultat.

Posledica 38. *Problem iskanja povezavne metrične dimenzije povezanega grafa je NP-težek.*

Zaradi zahtevnosti problema iskanja povezavne metrične dimenzije je zanj smiselno poiskati tudi aproksimacijski algoritem. S podobnim pristopom kot v [32] lahko v polinomskem času naredimo aproksimacijo s faktorjem $O(\log m)$, kjer je m število povezav grafa.

Izrek 39. *Če je G poljuben povezan graf z m povezavami, potem lahko v polinomskem času aproksimiramo $\dim_e(G)$ s faktorjem $O(\log m)$.*

Mešana metrična dimenzija

Pravimo, da vozlišče v povezanega grafa G *razlikuje* dva elementa $x, y \in V(G) \cup E(G)$ grafa G, če velja $d_G(x,v) \neq d_G(y,v)$. Množici vozlišč S grafa G pravimo *mešani metrični generator*, če za vsaka dva elementa $x, y \in V(G) \cup E(G)$ grafa G, kjer $x \neq y$, obstaja vozlišče iz S, ki ju razlikuje. Moči najmanjšega mešanega metričnega generatorja grafa G rečemo *mešana metrična dimenzija* in jo označimo z $\dim_{\mathrm{m}}(G)$. *Mešana metrična baza* grafa G je mešani metrični generator grafa G, ki ima moč $\dim_{\mathrm{m}}(G)$.

Problem določitve mešane metrične dimenzije grafa lahko predstavimo tudi kot optimizacijski problem. Matematični model za izračun mešane metrične dimenzije ali za iskanje mešane metrične baze je podoben modelu za metrično dimenzijo, opisanem v [13].

Naj bo G graf z n vozlišči in m povezavami. Naj bo $V = \{v_1, v_2, \ldots, v_n\}$ množica vozlišč in $E = \{e_1, e_2, \ldots, e_m\}$ množica povezav. V $n \times (n+m)$ dimenzionalni matriki $D = [d_{ij}]$ so elementi matrike enaki razdaljam med elementi grafa $d_{ij} = d_G(x_i, x_j)$, kjer je $x_i \in V$ in $x_j \in V \cup E$. S pomočjo spremenljivk $y_i \in \{0,1\}$ za $i \in \{1, 2, \ldots, n\}$, definiramo naslednjo funkcijo:

$$\mathcal{F}(y_1, y_2, \ldots, y_n) = y_1 + y_2 + \cdots + y_n.$$

Določitev minimuma funkcije $\mathcal{F}$ glede na omejitve

$$\sum_{i=1}^{n} |d_{ij} - d_{il}| y_i \geq 1, \quad \text{za vse } 1 \leq j < l \leq n+m,$$

je ekvivalentno iskanju mešane metrične baze grafa G. Rešitev $y_1, y_2, \ldots, y_n$ predstavlja množico vrednosti, za katero funkcija $\mathcal{F}$ doseže najmanjšo možno vrednost. To je ekvivaletno trditvi, da je množica $W = \{v_i \in V \mid y_i = 1\}$ mešana metrična baza za G.

Iz definicije sledi, da je mešani metrični generator tudi metrični generator in povezavni metrični generator. Iz tega takoj sledi naslednja zveza. Za poljuben graf

G,

$$\dim_{\mathrm{m}}(G) \geq \max\{\dim(G), \dim_{\mathrm{e}}(G)\}.$$

Hitro vidimo, da je celotna množica vozlišč kateregakoli grafa G mešani metrični generator. Poljubno vozlišče in vsaka povezava grafa G, ki je incidenčna s tem vozliščem, imata enako razdaljo do tega vozlišča. Eno samo vozlišče torej ne more predstavljati mešanega metričnega generatorja grafa G. Torej velja:

Opomba 40. *Za poljuben povezan graf G na n vozliščih velja $2 \leq \dim_{\mathrm{m}}(G) \leq n$.*

Naslednje trditve nam povedo, kdaj neka vozlišča pripadajo mešanemu metričnemu generatorju.

Trditev 41. *Če sta u in v prava dvojčka grafa G, potem u in v pripadata vsakemu mešanemu metričnemu generatorju grafa G.*

Trditev 42. *Če sta u in v neprava dvojčka grafa G in je S mešani metrični generator grafa G, potem velja $\{u, v\} \cap S \neq \emptyset$.*

Trditev 43. *Če je u simplicialno vozlišče grafa G, potem u pripada vsakemu mešanemu metričnemu generatorju grafa G.*

Neposredna posledica Trditve 43 je naslednji rezultat.

Posledica 44. *Če je u vozlišče grafa G stopnje ena, potem u pripada vsakemu mešanemu metričnemu generatorju grafa G.*

V opombi 40 sta podani spodnja in zgornja meja za mešano metrično dimenzijo. Obe meji sta dosegljivi. Še več, v naslednjih dveh izrekih karakteriziramo družine grafov, ki dosežejo meje iz opombe 40.

Izrek 45. *Naj bo G graf na n vozliščih. Velja, da je $\dim_{\mathrm{m}}(G) = 2$ natanko tedaj, ko je G pot.*

Naj bo v vozlišče grafa G. Vozlišču $u \in N(v)$ rečemo *maksimalni sosed* vozlišča v, če so vsi sosedi vozlišča v (in tudi sam v) v zaprti soseščini vozlišča u.

Izrek 46. *Naj bo G graf na n vozliščih. Potem je $\dim_{\mathrm{m}}(G) = n$ natanko tedaj, ko vsako vozlišče grafa G premore maksimalnega soseda.*

Za cikle, polne dvodelne grafe, drevesa in mreže veljajo naslednji rezultati, povezani z njihovo mešano metrično dimenzijo.

Trditev 47. *Za poljubno pozitivno celo število* $n \geq 4$ *je* $\dim_{\mathrm{m}}(C_n) = 3$.

Trditev 48. *Za poljubni pozitivni celi števili* $r, t \geq 2$ *je*

$$\dim_{\mathrm{m}}(K_{r,t}) = \begin{cases} r+t-1, & \textit{če je } r = 2 \textit{ ali } t = 2, \\ r+t-2, & \textit{sicer.} \end{cases}$$

Izrek 49. *Za poljubno drevo* T*, ki ima* $l(T)$ *listov, je* $\dim_{\mathrm{m}}(T) = l(T)$.

Trditev 50. *Če je* G *mreža* $P_r \square P_t$*, kjer je* $r \geq t \geq 2$*, potem je* $\dim_{\mathrm{m}}(G) = 3$.

V naslednjem izreku je podana zgornja meja za mešano metrično dimenzijo grafa G glede na ožino grafa G.

Izrek 51. *Naj bo* G *graf na* n *vozliščih. Če* G *premore cikel, potem je* $\dim_{\mathrm{m}}(G) \leq n - g(G) + 3$.

Meja iz izreka 51 je dosegljiva, kar je vidno tudi v naslednjih primerih. Za vsak cikel C_n je $\dim_{\mathrm{m}}(C_n) = n - g(C_n) + 3 = 3$. Za vsak polni graf je $\dim_{\mathrm{m}}(K_n) = n - g(K_n) + 3 = n$. Za vsak polni dvodelni graf $K_{2,t}$ velja $\dim_{\mathrm{m}}(K_{2,t}) = t + 2 - g(K_{2,t}) + 3 = t + 1$. Za vsak graf G, v katerem ima vsako vozlišče maksimalnega soseda in premore cikel, je ožina $g(G) = 3$, in zato po izreku 46 velja $\dim_{\mathrm{m}}(G) = n - g(G) + 3$.

Tudi za problem izračuna mešane metrične dimenzije se izkaže, da je NP-težek. Odločitveni problem za mešano metrično dimenzijo je definiran na naslednji način:

PROBLEM MEŠANE METRIČNE DIMENZIJE
INSTANCA: Povezan graf G na $n \geq 3$ vozliščih in celo število $2 \leq r \leq n$.
VPRAŠANJE: Ali je $\dim_{\mathrm{m}}(G) \leq r$?

Za ta odločitveni problem, podobno kot pri povezavni metrični dimenziji, velja naslednji izrek.

Izrek 52. *PROBLEM MEŠANE METRIČNE DIMENZIJE je NP-poln.*

Neposredno iz Izreka 52 dobimo naslednji rezultat.

Posledica 53. *Problem iskanja mešane metrične dimenzije povezanega grafa je NP-težek.*

Curriculum Vitae

Aleksander Kelenc
Date of birth: 24/2/1989
Nationality: Slovenian

Education:

- 2008-2011: B.S. in Mathematics, Faculty of Natural Sciences and Mathematics, University of Maribor, Slovenia
- 2011-2013: M.S. in Mathematics, Faculty of Natural Sciences and Mathematics, University of Maribor, Slovenia
- 2014-Present: Ph.D. in Mathematics, Faculty of Natural Sciences and Mathematics, University of Maribor, Slovenia

Teaching experience:

- 2013-Present: teaching assistant at University of Maribor, Slovenia

Working experience:

- 2011-2013: programmer at National Education Institute Slovenia (contract work)
- 2013-2016: Independent expert associate at Faculty of Natural Sciences and Mathematics, University of Maribor, Slovenia

- 2017: Assistant for Mathematics at Faculty of Electrical Engineering and Computer Science, University of Maribor, Slovenia
- 2018-2019: Independent expert associate at Faculty of Natural Sciences and Mathematics, University of Maribor, Slovenia
- 2019-Present: Assistant for Mathematics at Faculty of Electrical Engineering and Computer Science, University of Maribor, Slovenia

Personal bibliography:

[1] D. Božović, A. Kelenc, I. Peterin, I. G. Yero, Incidence dimension and 2-packing number in graphs, (2018) arXiv:1811.03156v1.

[2] A. Kelenc, Determining the Hausdorff Distance Between Trees in Polynomial Time, (2019) arXiv:1907.01299v1.

[3] A. Kelenc, D. Kuziak, A. Taranenko, I. G. Yero, Mixed metric dimension of graphs, *Appl. Math. Comput.* 314 (1) (2017) 429–438.

[4] A. Kelenc, A. Taranenko, On the Hausdorff Distance between Some Families of Chemical Graphs, *MATCH Commun. Math. Comput. Chem.* 74 (2) (2015) 223–246.

[5] A. Kelenc, S. Klavžar, N. Tratnik, The edge-Wiener index of benzenoid systems in linear time, *MATCH Commun. Math. Comput. Chem.* 74 (3) (2015) 521–532.

[6] A. Kelenc, N. Tratnik, I. G. Yero, Uniquely identifying the edges of a graph: the edge metric dimension, *Discrete Appl. Math.* 251 (2018) 204–220.

UNIVERZA V MARIBORU

FAKULTETA ZA NARAVOSLOVJE IN MATEMATIKO

IZJAVA O AVTORSTVU IN ISTOVETNOSTI TISKANE IN ELEKTRONSKE OBLIKE DOKTORSKE DISERTACIJE

Ime in priimek študenta: Aleksander Kelenc

Študijski program: Matematika

Naslov doktorske disertacije: Na razdaljah osnovane invariante in mere v grafih

Mentor: doc. dr. Andrej Taranenko

Somentor: izr. prof. dr. Ismael Gonzalez Yero

Podpisani študent Aleksander Kelenc

- izjavljam, da je zaključno delo rezultat mojega znanstvenoraziskovalnega dela;
- izjavljam, da sem pridobil vsa potrebna soglasja za uporabo podatkov in avtorskih del v zaključnem delu in jih v zaključnem delu jasno in ustrezno označil;
- na Univerzo v Mariboru neodplačno, neizključno, prostorsko in časovno neomejeno prenašam pravico shranitve avtorskega dela v elektronski obliki, pravico reproduciranja ter pravico ponuditi zaključno delo javnosti na svetovnem spletu preko DKUM in drugih informacijskih zbirk in ponudnikov; sem seznanjen, da bodo dela deponirana/objavljena v DKUM dostopna široki javnosti pod pogoji licence Creative Commons BY-NC-D, kar vključuje tudi avtomatizirano indeksiranje preko spleta in obdelavo besedil za potrebe tekstovnega in podatkovnega rudarjenja in ekstrakcije znanja iz vsebin; uporabnikom se dovoli reproduciranje brez predelave avtorskega dela, distribuiranje, dajanje v najem in priobčitev javnosti samega izvirnega avtorskega dela, in sicer pod pogojem, da navedejo avtorja in da ne gre za komercialno uporabo;
- dovoljujem objavo svojih osebnih podatkov, vezanih na zaključek študija (ime, priimek, leto zaključka študija, naslov zaključnega dela) na spletnih straneh Univerze v Mariboru in v publikacijah Univerze v Mariboru;
- izjavljam, da je tiskana oblika zaključnega dela istovetna elektronski obliki zaključnega dela, ki sem jo oddal za objavo v DKUM;
- izjavljam, da sem seznanjen s pogoji Proquest-a za oddajo in javno objavo doktorske disertacije v podatkovno zbirko ProQuest Dissertations & Theses Global (http://contentz.mkt5049.com/lp/43888/382619/PQDTauthoragreement.pdf).

Kraj in datum: Podpis študenta:

Maribor, 10. 12. 2019