Measuring closeness of graphs – the Hausdorff distance

Iztok Banič[1] and Andrej Taranenko[2]
[1,2]Faculty of Natural Sciences and Mathematics,
University of Maribor,
Koroška 160, SI-2000 Maribor, Slovenia
[1]Andrej Marušič Insitute, University of Primorska,
Muzejski trg 2, SI-6000 Koper, Slovenia
[2]Institute of Mathematics, Physics and Mechanics,
Jadranska 19, SI-1000 Ljubljana, Slovenia
[1]E-mail: iztok.banic@uni-mb.si
[2]E-mail: andrej.taranenko@uni-mb.si

April 28, 2014

Abstract

We introduce an apparatus to measure closeness or relationship of two given objects. This is a topology based apparatus that uses graph representations of the compared objects. In more detail, we obtain a metric on the class of all pairwise non-isomorphic connected simple graphs to measure closeness of two such graphs. To obtain such a measure, we use the theory of hyperspaces from topology to introduce the notion of the Hausdorff graph 2^G of any graph G. Then, using this new concept of Hausdorff graphs combined with the notion of graph amalgams, we present the Hausdorff distance, which proves to be useful when examining the closeness of any two connected simple graphs. We also present many possible applications of these concepts in various areas.

2010 *Mathematics Subject Classification*: Primary 54B20, 05C99; Secondary 05C40, 05C12.

Key words and phrases: Graphs, Hausdorff graphs, Hausdorff metric, Hausdorff distance

1 Introduction

Given any two objects, A and B (e.g. two hand prints, two tree leaves, two pictures, two faces, two DNA codes...), many questions regarding their closeness may appear. For example:

(i) How close (similar) is A to B and how to measure this closeness?

(ii) Are A and B somehow related?

(iii) How much do A and B look alike?

(iv) How much do A and B fit to each other?

In topology, one way to measure closeness of two sets is to use the so-called Hausdorff metric. To explain it in more details, suppose that (X, d) is a non-empty compact metric space. The family of all non-empty closed subsets of X is usually denoted by 2^X. For each $A \in 2^X$ and each $r > 0$, denote by $N(A, r) = \bigcup_{a \in A} K(a, r)$ the union of open r-balls $K(a, r)$ in X with centres $a \in A$. Think of $N(A, r)$ as A being inflated by factor r, see Figure 1.

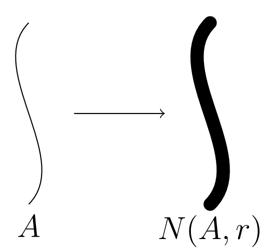

Figure 1: Inflated A.

Then the *Hausdorff metric* h_X is defined on 2^X by

$$h_X(A, B) = \inf\{\varepsilon > 0 \mid A \subseteq N(B, \varepsilon), B \subseteq N(A, \varepsilon)\}, \tag{1.1}$$

for any $A, B \in 2^X$, to measure closeness between A and B. In other words, h_X is defined in such a way that A and B are close to each other, if for each point $a \in A$ there is a point $b \in B$ that is close to a, and for each point $b \in B$ there is a point $a \in A$ that is close to b.

It is a well-known fact that h_X is a metric on 2^X. The pair (X, h_X) is called the hyperspace of (X, d). More details about hyperspaces and the introduced Hausdorff metric can be found in [6, 13].

In the present paper we use this idea of closeness from topology and apply it into the language of graph theory. We realize this idea by introducing the new concept of so-called Hausdorff graphs (in graph theory the term hypergraph is already used to define another object). We combine them with the notion of amalgams (cf. [1, 9]) to define the Hausdorff distance on the class of all connected simple graphs as a measure of similarity of two such graphs. We also discuss about possible applications of this new concept and give an easy example of a useful application in biology.

Simon Romero [10, 11, 12] also uses the idea of hyperspaces $C(X) \subseteq 2^X$ of connected compact subspaces of X and applies it into graph theory by defining so called hyperspace graphs of connected subgraphs. He also poses the question of how to define graphs 2^G that are analogous to the topological hyperspaces 2^X [10, p. 91, Question 1]. Among other things, we answer the question in Section 3.

In graph theory the distance between two graphs has been defined in various ways, for examples see [2, 3, 4, 5, 8]. One common way is to define the distance as the minimum number of some operations (on vertices or edges) one needs to transform one graph into the other. Under the assumption that the graphs compared are of the same order and size, the operations defined were edge move [2], edge rotation [4] and edge slide [2, 8], among others.

A graph G is said to be a common subgraph of the graphs G_1 and G_2 if it holds that $G \subseteq G_1$ and $G \subseteq G_2$. We say that a common subgraph G of G_1 and G_2 is a maximal common subgraph if there does not exist a common subgraph H with $|V(H)| > |V(G)|$ and $G \subseteq H$. In [3], the authors use the notion of the maximal common subgraph to define the distance between two non-empty graphs, where the metric they define uses only the order of a maximal common subgraph and the order of the graphs compared.

We proceed as follows. In the next section we introduce the basic definitions and notations that are used throughout the paper. In Section 3 we define the Hausdorff graph 2^G of a given graph G, study some of its properties and use it to define the Hausdorff metric on subgraphs of G. We use the ideas of Section 3 and graph amalgams to define the Hausdorff distance as a measure of closeness of two connected simple graphs. We conclude the paper by presenting possible applications of the concept and by stating some open problems.

2 Definitions and Notations

A *graph* $G = (V(G), E(G))$ is determined by a non-empty *vertex set* $V(G)$ and a set $E(G)$ of unordered pairs of vertices $\{u, v\}$, called the set of *edges*. We will use the short notation uv for edge $\{u, v\}$. We say that a vertex u is adjacent to a vertex v if $uv \in E(G)$.

Let $G = (V(G), E(G))$ and $H = (V(H), E(H))$ be any graphs. If $V(H) \subseteq V(G)$ and $E(H) \subseteq E(G)$, then we say that H is a subgraph of G and write $H \subseteq G$.

All graphs considered in the paper are simple graphs, i.e. the graphs without multiple edges and without 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 $uv \in E(G)$.

Two graphs are *isomorphic*, if there is a bijection between their vertex sets that preserves adjacency and non-adjacency of the vertices.

A walk W from a vertex x to a vertex y in a graph G is a sequence

$$x = v_0 v_1 v_2 \dots v_{k-1} v_k = y$$

of vertices of G, where $v_i v_{i+1} \in E(G)$ or $v_i = v_{i+1}$, for each $i \in \{0, 1, \dots, k-1\}$. We will also denote it by $(x = v_0, v_1, v_2, \dots, v_{k-1}, v_k = y)$. The *length* of a walk W, denoted by $\ell(W)$, is the number of edges in W.

A path P from a vertex x to a vertex y in G is a sequence

$$x = v_0 v_1 v_2 \dots v_{k-1} v_k = y$$

of pairwise different vertices of G, where $v_i v_{i+1} \in E(G)$, for each $i \in \{0, \dots, k-1\}$. The vertices x and y are called the *endpoints* of the path. The path P will also be denoted by $(x = v_0, v_1, v_2, \dots, v_{k-1}, v_k = y)$. The *length* of a path P, denoted by $\ell(P)$, is the number of edges in P. The *distance* between vertices x and y, denoted by $d_G(x, y)$ is the length of a shortest path between x and y in G.

Note that every walk W from x to y in a graph G gives rise to a path P from x to y in G, such that $\ell(P) \leq \ell(W)$.

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

A subgraph H of a graph G is isometric in G if for any $u, v \in V(H)$, it holds that $d_H(u, v) = d_G(u, v)$.

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

3 Hausdorff Graphs

In this section we introduce the new notion of the Hausdorff graph 2^G of a graph G and define a metric on 2^G to measure closeness of any two subgraphs of G.

In the following definition we apply (1.1) in such a way that for a graph G, two of its subgraphs H_1 and H_2 are close to each other in G, if for each vertex $v \in V(H_1)$ there is a vertex $v' \in V(H_2)$ that is close to v, and for each vertex $v' \in V(H_2)$ there is a vertex $v \in V(H_1)$ that is close to v'.

Definition 3.1. *Let G be an arbitrary graph. The Hausdorff graph of the 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*

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

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

Example 3.2. *Let G be a trivial graph with $V(G) = \{1\}$. Then 2^G is the trivial graph, $V(2^G) = \{\{1\}\}$.*

Example 3.3. *Let P_3 be a path on three vertices, $V(P_3) = \{1, 2, 3\}$. Then 2^{P_3} is the graph depicted in Figure 2, where the vertices are subgraphs of P_3 (in rounded rectangles), and edges between vertices are black and gray lines.*

Definition 3.4. *The i^{th} level of 2^G, denoted by $[2^G]_i$, is the induced subgraph of 2^G, where the vertices of $[2^G]_i$ represent all subgraphs of G on exactly i vertices.*

This means that the graph $[2^G]_1$ contains as vertices all subgraphs of G on exactly one vertex, the graph $[2^G]_2$ contains as vertices all subgraphs of G on exactly two vertices, $\ldots$, $[2^G]_{|V(G)|}$ contains as vertices all subgraphs of G on exactly $|V(G)|$ vertices. Therefore also $G \in V([2^G]_{|V(G)|}) \subseteq V(2^G)$. In Figure 2, the edges drawn in black depict edges between vertices in the same level.

Proposition 3.5. *For an arbitrary graph G, the graph $[2^G]_1$ is isomorphic to G.*

Proof. Let u be an arbitrary vertex of G, denote by H_u the corresponding vertex in $[2^G]_1$. Explicitly, H_u represents the subgraph $H_u \subseteq G$ for which

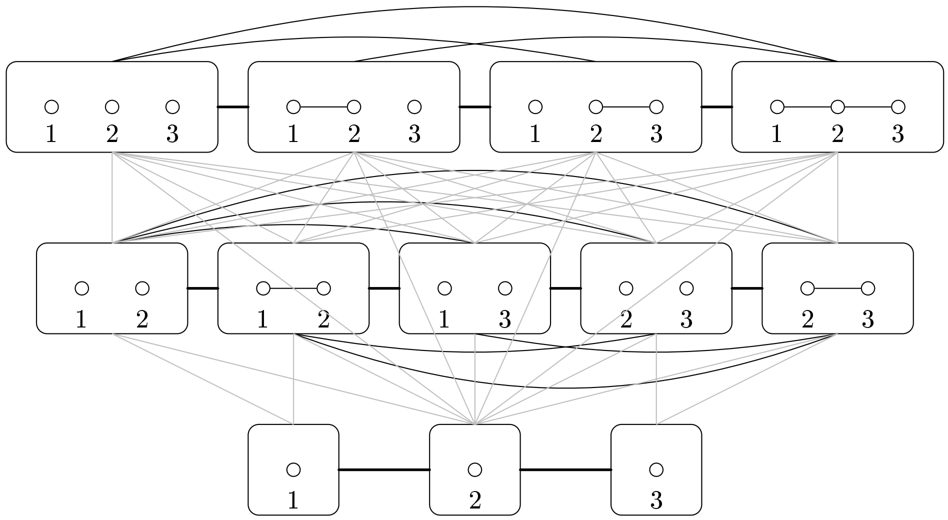

Figure 2: The Hausdorff graph 2^{P_3} of P_3.

$V(H_u) = \{u\}$ and $E(H_u) = \emptyset$. It is clear, that the mapping $V(G) \to V([2^G]_1)$, $u \mapsto H_u$, is a bijection, hence the vertex sets of graphs G and $[2^G]_1$ are of the same size.

We need to prove that $uv \in E(G)$ if and only if $H_uH_v \in E([2^G]_1)$. Let $uv \in E(G)$. Since $d_G(u, v) = 1$ the vertices H_u and H_v are adjacent in $[2^G]_1$. For the converse, suppose that $H_uH_v \in E([2^G]_1)$. Since H_u and H_v correspond to disjoint graphs (each represents a trivial graph), the distance $d_G(u, v) = 1$. □

From the previous proposition we immediately obtain the following corollary.

Corollary 3.6. *Let G be an arbitrary graph. Then 2^G contains an isomorphic copy of G as an induced subgraph.*

The Hausdorff metric h_G between two subgraphs of a graph G will be defined in the following definition. It will tell us how much two subgraphs of G coincide. Namely, the smaller the Hausdorff metric between the two subgraphs is, more they coincide.

Definition 3.7. *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.

The following example shows that the Hausdorff metric h_G may not be a metric on $V(2^G)$.

Example 3.8. *Let G be a totally disconnected graph on two vertices, i.e. $V(G) = \{1,2\}$ and $E(G) = \emptyset$. Then 2^G is the totally disconnected graph on three vertices, $V(2^G) = \{H_1, H_2, H_3\}$, where $V(H_1) = \{1\}$, $V(H_2) = \{2\}$, and $H_3 = G$. Therefore $h_G(H_1, H_3)$ is not defined.*

We will show in Corollary 3.18, that if G is a connected graph, then h_G is in fact a metric on $V(2^G)$.

First we prove the following easy lemmas.

Lemma 3.9. *Let G be an arbitrary graph and $H_1, H_2 \subseteq G$, such that $V(H_1) = V(H_2)$. Then $h_G(H_1, H_2) = 1$ if and only if $H_1 \neq H_2$.*

Proof. If $H_1 \neq H_2$ then for each $v \in V(H_1)$, obviously $v \in V(H_2)$, and vice versa. Therefore H_1 and H_2 are adjacent, since $H_1 \neq H_2$. The converse follows immediately from Definition 3.1. □

From Lemma 3.9 we obtain the following corollary.

Corollary 3.10. *Let $H \subseteq G$ and let $\mathcal{K} = \{H_i \subseteq G | V(H_i) = V(H)\}$. Then $\langle \mathcal{K} \rangle \subseteq 2^G$ is a complete graph.*

Next, we define and introduce some important subgraphs of a Hausdorff graph 2^G.

Definition 3.11. *Let G be a graph. Then*

(*i*) $C(G) = \left\langle \{H \in V(2^G) \mid H \textit{ is connected}\} \right\rangle$

(*ii*) $CI(G) = \langle \{H \in V(C(G)) \mid H \textit{ is induced in } G\} \rangle$

(*iii*) $K(G) = \langle \{H \in V(CI(G)) \mid H \textit{ is a complete graph}\} \rangle$

Remark 3.12. *Note that $K(G) = \left\langle \{H \in V(2^G) \mid H \textit{ is a complete graph}\} \right\rangle$.*

Example 3.13. *Observe that for a totally disconnected graph on n vertices, say S_n, the following hold true:*

(*i*) $2^{S_n} = S_{2^n - 1}$ *and*

(*ii*) $C(S_n) = CI(S_n) = K(S_n) = S_n$.

Example 3.14. *Observe that for a complete graph on n vertices, say K_n, the following hold true:*

(i) $2^{K_n} = K_m$, *where* $m = \sum_{i=1}^{n}\left(\binom{n}{i}\sum_{j=0}^{\frac{i(i-1)}{2}}\binom{\frac{i(i-1)}{2}}{j}\right)$.

(ii) $CI(K_n) = K(K_n) = K_m$, *where* $m = \sum_{i=1}^{n}\binom{n}{i}$.

Example 3.15. *Figure 3 shows the graphs $C(P_3), CI(P_3)$ and $K(P_3)$. The whole graph depicted is $C(P_3) = CI(P_3)$, while $K(P_3)$ is the graph depicted inside the dashed area.*

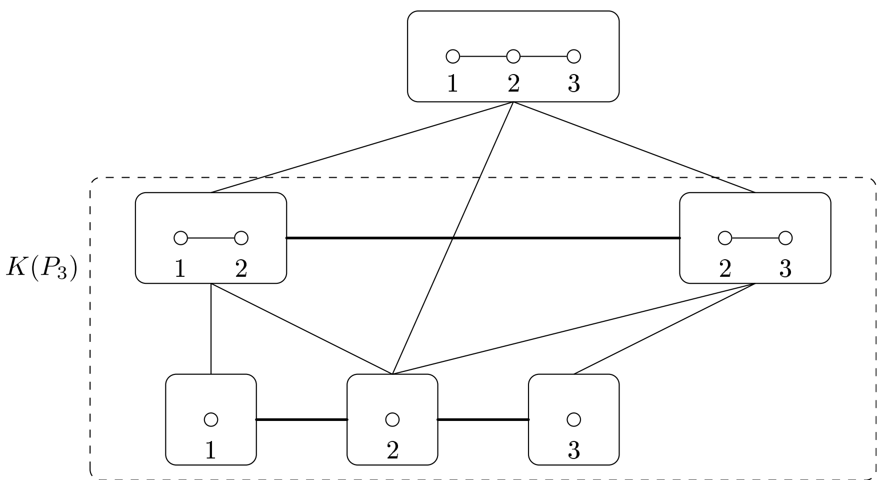

Figure 3: The graphs $C(P_3), CI(P_3)$ and $K(P_3)$.

Example 3.16. *Figure 4 shows graphs $C(C_3)$, $CI(C_3)$ and $K(C_3)$. The whole graph depicted is $C(C_3)$, while $CI(C_3) = K(C_3)$ is the graph depicted inside the dashed area.*

Let $H \subseteq 2^G$. As in Definition 3.4 we use $[H]_i$ to denote the i^{th} level of H.

Theorem 3.17. *Let G be a graph. The following statements are equivalent.*

(i) *G is connected.*

(ii) *2^G is connected.*

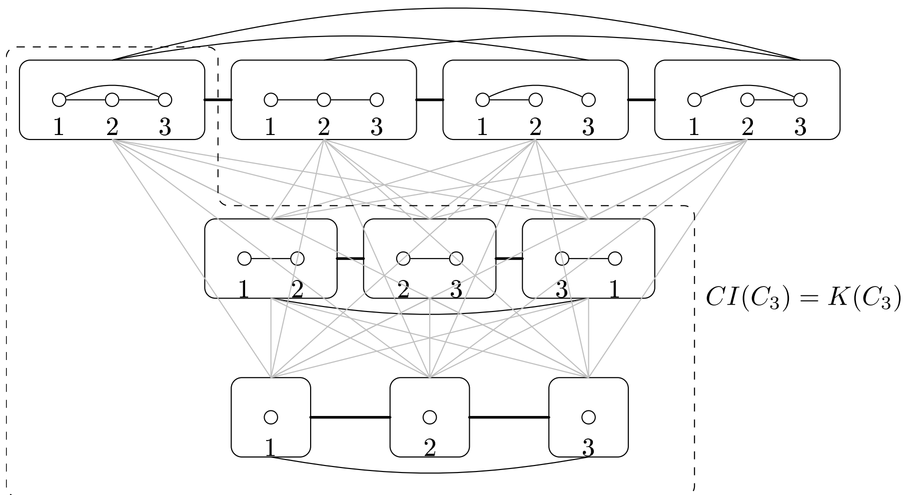

Figure 4: The graphs $C(C_3), CI(C_3)$ and $K(C_3)$.

(iii) $C(G)$ *is connected.*

(iv) $CI(G)$ *is connected.*

(v) $K(G)$ *is connected.*

Proof. (i)$\Leftrightarrow$(ii). Let G be connected. Let $H \in V(2^G)$ be arbitrarily chosen. We will construct a path $H = H_1 H_2 \dots H_n = G$ in 2^G connecting vertices H and G.

If H is not the induced subgraph of G on the vertex set $V(H)$, then let H_2 be the induced subgraph of G on $V(H)$. By Lemma 3.9 the vertices $H_1 = H$ and H_2 are adjacent in 2^G. If H_2 equals G, we are done. Otherwise, let $u_1 \in V(G) \setminus V(H_2)$ such that u_1 is adjacent to a vertex in H_2 (such a vertex exists, since G is connected). Let H_3 be the induced subgraph of G on the vertex set $V(H_2) \cup \{u_1\}$. It follows from Definition 3.1 that $H_2 H_3 \in E(2^G)$. If H_3 equals G, we are done. Otherwise we continue with constructing graphs $H_4, H_5, \dots$ such that for each $i \geq 4$ it holds that $V(H_{i+1}) \setminus V(H_i)$ consists of a single vertex u_{i-1} with a neighbour in H_i, hence $H_i H_{i+1} \in E(2^G)$. Since $V(G)$ is finite, the procedure ends at step a n, with $V(H_n) = V(G)$.

If H is the induced subgraph of G on the vertex set $V(H)$, then the desired path is $H = H_2 H_3 \dots H_n = G$, where $H_2, H_3, \dots$ are as described in the previous case.

For the converse, suppose that G is not connected. We will prove that 2^G is not connected. Let G_1 and G_2 be distinct connected components of G. Assume that 2^G is connected. There exists a path $G_1 = H_1 H_2 \dots H_n = G_2$

in 2^G. Since, G_1 and G_2 are disjoint, there exists an index $i \in \{2, \dots, n\}$ such that $H_i \not\subseteq G_1$ (at least $i = n$ satisfies this condition). Let i be the smallest such that $H_i \not\subseteq G_1$. Let $u \in V(H_i) \setminus V(G_1)$. Since, u has no neighbour in G_1 (since u is not in the connected component G_1), it has no neighbour in H_{i-1}. Hence, $h_G(H_{i-1}, H_i) > 1$. This contradicts the assumption that 2^G is connected.

To prove (i)$\Leftrightarrow$(iii) and (i)$\Leftrightarrow$(iv) it suffices to follow the proof of (i)$\Leftrightarrow$(ii).

(i)$\Leftrightarrow$(v). Suppose G is connected. Observe that $[2^G]_1 = [K(G)]_1$. Then by Proposition 3.5 the level $[K(G)]_1$ is also connected. Let $H \in V(K(G))$ be arbitrarily chosen, with $|V(H)| > 1$, and $u \in V(H)$. Since H is a complete graph it is adjacent in $K(G)$ to the trivial graph corresponding to vertex u in $[K(G)]_1$. Therefore $K(G)$ is connected.

For the converse, suppose that G is not connected. We will prove that $K(G)$ is not connected. Let G_1 and G_2 be distinct connected components of G, also, let A_1 and A_2 be any subgraphs of G_1 and G_2, respectively, each isomorphic to a complete graph. Assume that $K(G)$ is connected. There exists a path $A_1 = H_1 H_2 \dots H_n = A_2$ in $K(G)$. Since, A_1 and A_2 are disjoint, there exists H_i such that $H_i \not\subseteq G_1$, for some $i \in \{2, \dots, n\}$. Let i be the smallest such that $H_i \not\subseteq G_1$. Let $u \in V(H_i) \setminus V(G_1)$. Since, u has no neighbour in A_1, it has no neighbour in H_{i-1}. Hence, $h_G(H_{i-1}, H_i) > 1$. This contradicts the assumption that $K(G)$ is connected. □

Corollary 3.18. *If G is connected, then h_G is a metric on $V(2^G)$.*

Proof. Let G be a connected graph. Then 2^G is connected by Theorem 3.17. Therefore $h_G = d_{2^G}$ is a well-defined metric on $V(2^G)$. □

The graph $C(G)$ plays an important role in the next section, where we define the Hausdorff distance between arbitrary connected simple graphs. Since the property of connectedness is defined through paths, we describe $C(P_n)$ of an arbitrary path P_n in the following results.

Proposition 3.19. *Let P_n be a path on n vertices. Then $[C(P_n)]_i$ is isomorphic to P_{n-i+1}, for $i \in \{1, 2, \dots, n\}$.*

Proof. Let $P_n = v_1 v_2 \dots v_n$. Let $i \in \{1, 2, \dots, n\}$ be arbitrary. Note that the only connected induced subgraphs of P_n on i vertices are paths of length $i - 1$. It is easy to see, that in P_n there are exactly $n - i + 1$ different paths of length $i - 1$. So the i^{th} level, $[C(P_n)]_i$, has $n - i + 1$ vertices. Let $H_1, H_2 \subseteq P_n$ be two different induced connected paths of order i. Let $H_1 = v_j v_{j+1} \dots v_{j+i-1}$, $j \in \{1, \dots, n - i + 1\}$, and $H_2 = v_k v_{k+1} \dots v_{k+i-1}$,

$k \in \{1, \ldots, n-i+1\}$, with $j \neq k$. Then by Definition 3.1, H_1 and H_2 are adjacent, if every vertex of H_1 not in the intersection of the two paths, has a neighbour in H_2, and vice versa. In other words, the endpoints of H_1 have as a neighbour one (the closest one) of the endpoints of H_2, otherwise they are not adjacent. So, H_1 and H_2 are adjacent if and only if $|j-k| = 1$. Since $j \neq k$ the assertion follows. □

Note, that $C(P_n) = CI(P_n)$, since connected subgraphs of a path are exactly the induced connected subgraphs of the path.

Proposition 3.20. *Let $P_n = u_1u_2 \ldots u_n$ be a path on $n \geq 2$ vertices. Let $P \in V([C(P_n)]_i)$ and $Q \in V([C(P_n)]_{i+1})$, for some $i \in \{1, 2, \ldots, n-1\}$. Moreover, let $P = u_ju_{j+1} \ldots u_{j+i-1}$, $j \in \{1, 2, \ldots, n-i+1\}$, and $Q = u_ku_{k+1} \ldots u_{k+i}$, $k \in \{1, 2, \ldots, n-i\}$. Then $PQ \in E(C(P_n))$ if and only if $j = k$ or $j = k+1$.*

Proof. Let $PQ \in E(C(P_n))$. By Definition 3.1 every vertex of P is either in Q or it is adjacent to a vertex in Q, and vice-versa. Since the endpoints of a path are of degree 1, the endpoints of Q must either be in P or have a neighbour in P. Since $\ell(Q) - \ell(P) = 1$, both endpoints of Q cannot be in P and cannot both be disjoint with P. It follows that exactly one endpoint of Q is in P, this implies that $j = k$ or $j + i - 1 = k + i$ ($j = k+1$).

For the converse, suppose $j = k$, then $P \subseteq Q$ and the vertex $u_{k+i} \in V(Q)$ is the only vertex in $V(Q) \setminus V(P)$. Since it is adjacent to $u_{k+i-1} \in V(P)$, the paths P and Q are adjacent in $C(P_n)$. Also suppose $j = k+1$. Again $P \subseteq Q$ and the vertex $u_k \in V(Q)$ is the only vertex in $V(Q) \setminus V(P)$. Since it is adjacent to $u_{k+1} = u_j \in V(P)$, the paths P and Q are adjacent in $C(P_n)$. □

Proposition 3.21. *Let $P_n = u_1u_2 \ldots u_n$ be a path on $n \geq 3$ vertices. Let $P \in V([C(P_n)]_i)$ and $Q \in V([C(P_n)]_{i+2})$, for some $i \in \{1, 2, \ldots, n-2\}$. Moreover, let $P = u_ju_{j+1} \ldots u_{j+i-1}$, $j \in \{1, 2, \ldots, n-i+1\}$, and $Q = u_ku_{k+1} \ldots u_{k+i+1}$, $k \in \{1, 2, \ldots, n-i-1\}$. Then $PQ \in E(C(P_n))$ if and only if $j = k+1$.*

Proof. Let $PQ \in E(C(P_n))$. Since the endpoints of a path are of degree 1, the endpoints of Q must either be in P or have a neighbour in P. Since $\ell(Q) - \ell(P) = 2$, none of the endpoints of Q is in P. It follows that both endpoints of Q are adjacent to a vertex (an endpoint) in P, this implies that $j = k+1$ and $j + i - 1 = k + i$ ($j = k+1$).

For the converse, suppose $j = k + 1$, then $P \subseteq Q$ and the vertices $u_k, u_{k+i+1} \in V(Q)$ are the only vertices in $V(Q) \setminus V(P)$. Since u_k, u_{k+i+1} are adjacent to $u_j = u_{k+1}, u_{j+i-1} = u_{k+i} \in V(P)$, respectively, the paths P and Q are adjacent in $C(P_n)$. □

Proposition 3.22. *Let P_n be a path on n vertices. Let $P \in V([C(P_n)]_i)$ and $Q \in V([C(P_n)]_j)$, for some $i, j \in \{1, 2, \ldots, n\}$. If $|i - j| > 2$ then $PQ \notin E(C(P_n))$.*

Proof. Suppose $|i-j| \geq 3$ (this implies $n \geq 4$) and $j > i$. Since $\ell(Q)-\ell(P) \geq 3$ there exists an endpoint u in Q such that none of its neighbours are in P. This means that $d_{P_n}(u, v) > 1$, for all $v \in V(P)$. Hence, the assertion follows. □

4 Closeness of graphs

In this section we apply the notion of Hausdorff graphs to define a measure, called the Hausdorff distance, for closeness of any two connected simple graphs. First, we present some auxiliary definitions and results.

Definition 4.1. *Let H_1 be a subgraph of G_1 and H_2 a subgraph of G_2. If H_1 and H_2 are isomorphic graphs, then an 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.*

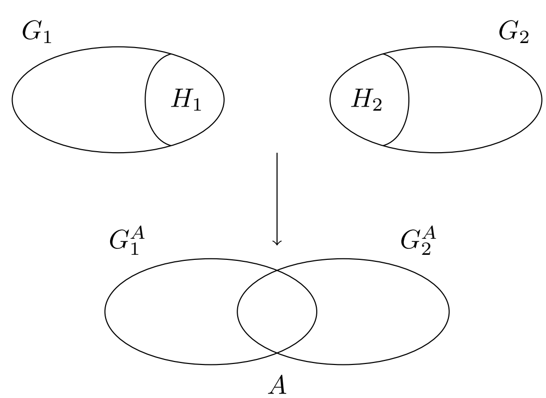

Figure 5: An amalgam A of G_1 and G_2.

Remark 4.2. *Let A be an amalgam of G_1 and G_2. We will always denote the covers of A by G_1^A and G_2^A. If $H \subseteq G_i$ ($u \in V(G_i)$), the corresponding graph (vertex) in G_i^A will also be denoted by H^A (u^A), $i \in \{1, 2\}$.*

Remark 4.3. *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.*

Remark 4.4. *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 4.5).*

Example 4.5. *Let G_1 and G_2 be the graphs depicted in Figure 6, 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 6 they are depicted by 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.*

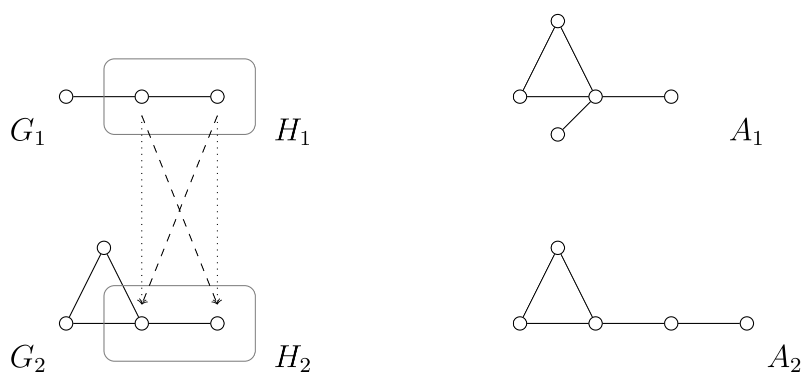

Figure 6: The amalgams A_1 and A_2 from Example 4.5.

In the rest of the paper, $\mathcal{G}$ will always denote the class of all connected simple graphs. We will be interested in the distance between the covers G_1^A and G_2^A in an amalgam A of G_1 and G_2. Moreover, we use the Hausdorff metric h_A on 2^A to determine this distance and express it via distances between vertices in A.

Lemma 4.6. *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 $h_A(G_1^A, G_2^A) \leq d$ 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 $d_A(u, v) \leq d$ and

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

Proof. Suppose, $h_A(G_1^A, G_2^A) \leq d$. Assume that (i) does not hold. Then there is a vertex $u \in V(G_1^A)$ such that for each $v \in V(G_2^A)$ it holds that $d_A(u,v) > d$. It follows that $u \notin V(G_1^A) \cap V(G_2^A)$ (otherwise, for $v = u$, $d_A(u,v) = 0 \not> d$) and

$$k = d_A(u, G_2^A) = \min\{d_A(u,v) \mid v \in V(G_2^A)\} > d. \tag{4.1}$$

Let Q be a shortest path of length k connecting u to a vertex in G_2^A.

On the other hand, since $h_A(G_1^A, G_2^A) \leq d$, there is a shortest path

$$P = (G_1^A = K_1, K_2, \ldots, K_j, K_{j+1} = G_2^A)$$

of length $j \leq d$ in 2^A between G_1^A and G_2^A. Next, we construct a walk from u to a vertex in G_2^A of length at most j. Let $u_1 = u \in V(K_1)$. Since $K_1K_2 \in E(2^A)$, there is a vertex $u_2 \in V(K_2)$ such that $d_A(u_1, u_2) \leq 1$. Say that we have already chosen vertices $u_1, u_2, \ldots, u_n$, $n < j+1$, such that for each $i \in \{1, 2, \ldots, n-1\}$ it holds that $d_A(u_i, u_{i+1}) \leq 1$, and $u_i \in V(K_i)$. Since $K_nK_{n+1} \in E(2^A)$, there is a vertex $u_{n+1} \in V(K_{n+1})$ such that $d_A(u_n, u_{n+1}) \leq 1$.

The chosen vertices $u_1, u_2, \ldots, u_{j+1}$ define a walk W of length at most j from $u = u_1$ to the vertex $u_{j+1} \in V(G_2^A)$. Since Q is the shortest path from u to a vertex from G_2^A, it follows that $k = \ell(Q) \leq \ell(W) \leq j \leq d$. This is a contradiction with (4.1).

Assuming that (ii) does not hold, we can obtain a contradiction in a similar way.

For the converse, assume (i) and (ii). We will construct a path P from G_1^A to G_2^A in 2^A of length $n \leq d$. For each $i \in \{0, 1, \ldots, d\}$ let $\mathcal{A}_i = \{v \in V(G_1^A) \mid d_A(v, G_2^A) = i\}$ and $\mathcal{B}_i = \{v \in V(G_2^A) \mid d_A(v, G_1^A) = i\}$. The sets $\mathcal{A}_i$ and $\mathcal{B}_i$ may be empty. Note also, that $\bigcup_{i=0}^{d}(\mathcal{A}_i \cup \mathcal{B}_i) = V(A)$. Say, $K_1 = G_1^A$. Suppose, K_i has already been constructed. Then let K_{i+1} be the induced graph $\langle (V(K_i) \setminus \mathcal{A}_{d-i+1}) \cup \mathcal{B}_i \rangle$ in A. It follows from (i) and (ii), as well as from the construction of K_i's that

(a) $h_A(K_i, K_{i+1}) \leq 1$, for each i,

(b) $K_{d+1} = G_2^A$ and

(c) $W = (K_1, K_2, \ldots, K_{d+1})$ is a walk from G_1^A to G_2^A in 2^A.

Hence there is a path from G_1^A to G_2^A in 2^A of length at most d. □

Remark 4.7. *Let $G_1, G_2 \in \mathcal{G}$ and A an amalgam of G_1 and G_2. Note, that in the proof of Lemma 4.6, all of the constructed paths in 2^A are also the paths in $C(A) = \big\langle \{H \in V(2^A) \mid H \text{ is a connected subgraph of } A\}\big\rangle$, if $G_1^A \cap G_2^A$ is a connected subgraph of A. Hence, following the same proof as the proof of Lemma 4.6, one can get the same result by replacing h_A with $d_{C(A)}$ for such amalgams A.*

Lemma 4.8. *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 $h_A(G_1^A, G_2^A) \geq d$ if and only if*

(*i*) *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*

(*ii*) *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 $d_A(u,v) \geq d$.*

Proof. We begin the proof by the following simple reasoning.

$$\begin{aligned}
& h_A(G_1^A, G_2^A) \geq d \quad \Leftrightarrow \\
& \forall d' < d : h_A(G_1^A, G_2^A) \not\leq d' \quad \Leftrightarrow \\
& \forall d' < d : \neg \big((\forall u \in V(G_1^A)\ \exists v \in V(G_2^A) : d_A(u,v) \leq d')\ \& \\
& \qquad (\forall u \in V(G_2^A)\ \exists v \in V(G_1^A) : d_A(u,v) \leq d') \big) \quad \Leftrightarrow \\
& \forall d' < d : \big((\exists u \in V(G_1^A) \text{ such that } \forall v \in V(G_2^A) : d_A(u,v) \not\leq d') \text{ or} \\
& \qquad (\exists u \in V(G_2^A) \text{ such that } \forall v \in V(G_1^A) : d_A(u,v) \not\leq d') \big) \quad \Leftrightarrow \\
& \big((\exists u \in V(G_1^A) \text{ such that } \forall v \in V(G_2^A) : d_A(u,v) \geq d) \text{ or} \\
& \qquad (\exists u \in V(G_2^A) \text{ such that } \forall v \in V(G_1^A) : d_A(u,v) \geq d) \big).
\end{aligned} \tag{4.2}$$

Obviously (4.2) implies ((i) or (ii)). Also, since there are no edges between $G_1^A \setminus G_2^A$ and $G_2^A \setminus G_1^A$, the assumption ((i) or (ii)) implies (4.2).

□

Remark 4.9. *Let $G_1, G_2 \in \mathcal{G}$ and A an amalgam of G_1 and G_2. Note, that following the proof of Lemma 4.8 one can get the same result by replacing h_A with $d_{C(A)}$, if $G_1^A \cap G_2^A$ is a connected subgraph of A.*

As an immediate consequence of Lemmas 4.6 and 4.8 we obtain the following theorem and corollary.

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* $d_A(u,v) \geq d$*.*

Remark 4.11. *Following Remarks 4.7 and 4.9, one can easily see that*

$$h_A(G_1^A, G_2^A) = d_{C(A)}(G_1^A, G_2^A),$$

for arbitrary $G_1, G_2 \in \mathcal{G}$ *and an amalgam* A *of* G_1 *and* G_2*, with* $G_1^A \cap G_2^A$ *being a connected subgraph of* A*.*

Corollary 4.12. *Let* $G_1, G_2 \in \mathcal{G}$*. Let* A *be an amalgam of* G_1 *and* G_2*. Then there is* $i \in \{1,2\}$ *such that there are vertices* $u \in V(G_i^A)$ *and* $v \in V(G_1^A \cap G_2^A)$ *with the distance* $d_A(u,v) = h_A(G_1^A, G_2^A)$*. Moreover, for each* $w \in V(A)$ *it holds that* $d_A(u,v) \geq d_A(w, G_1^A \cap G_2^A)$*.*

Proof. Let $d = h_A(G_1^A, G_2^A)$. By (iii) of Theorem 4.10 there is $i \in \{1,2\}$ such that there is a vertex $u \in V(G_i^A)$ such that for all vertices $v \in V(G_1^A \cap G_2^A)$ the distance $d_A(u,v) \geq d$. Without loss of generality, suppose $i = 1$. Using (i) of Theorem 4.10 there is a vertex $w_u \in V(G_2^A)$ such that the distance $d_A(u, w_u) \leq d$. Since, for each vertex $w \in V(G_2^A \setminus G_1^A)$ there is a vertex $v_w \in V(G_1^A \cap G_2^A)$ such that $d_A(u,w) = d_A(u,v_w) + d_A(v_w, w)$, such vertex v_{w_u} exists also for w_u. Therefore, $d_A(u, v_{w_u}) \leq d_A(u, w_u) \leq d$. Since $d_A(u, v_{w_u}) \geq d$, it follows that $d_A(u, v_{w_u}) = d$.

Let us now prove that for each $w \in V(A)$ it holds that $d_A(u,v) \geq d_A(w, G_1^A \cap G_2^A)$. Suppose there is a $w \in V(A)$, such that $d_A(w, G_1^A \cap G_2^A) > d_A(u,v)$. By Lemma 4.6 there is a vertex $w' \in V(A)$ such that $d_A(w,w') \leq d_A(u,v)$. Since w and w' belong to different covers of A, any shortest path between these vertices intersects $G_1^A \cap G_2^A$, meaning that $d_A(w, G_1^A \cap G_2^A) \leq d_A(u,v)$, a contradiction. □

We will define a measure called *the Hausdorff distance* on $\mathcal{G}$ which will serve as a measure of closeness of two connected simple graphs, i.e. how much two graphs coincide in such a way that two isomorphic graphs have Hausdorff distance 0.

We define on $\mathcal{G}$ a binary relation $\sim$ as follows:

$$G_1 \sim G_2 \Longleftrightarrow G_1 \text{ is isomorphic to } G_2.$$

Clearly, the relation $\sim$ is an equivalence relation on $\mathcal{G}$.

Definition 4.13. *Let $\mathcal{G}/_\sim = \{[G] \mid G \in \mathcal{G}\}$ be the family of all equivalence classes of the relation $\sim$ on $\mathcal{G}$. We define the function*

$$H : \mathcal{G}/_\sim \times \mathcal{G}/_\sim \to \mathbb{R}$$

as

$$H([G_1],[G_2]) = \min\{h_A(G_1^A, G_2^A) \mid A \text{ is an amalgam of } G_1 \text{ and } G_2\},$$

for any graphs $G_1, G_2 \in \mathcal{G}$.

The function H is obviously well-defined, since its definition does not depend on the representatives of the equivalence classes.

For the function H the following holds true.

Theorem 4.14. *Let $G_1, G_2 \in \mathcal{G}$ be arbitrary graphs. Then*

(i) $H([G_1],[G_2]) \geq 0$,

(ii) $H([G_1],[G_2]) = 0$ *if and only if* $[G_1] = [G_2]$, *and*

(iii) $H([G_1],[G_2]) = H([G_2],[G_1])$.

Proof. (i) Obviously $H([G_1],[G_2]) \geq 0$ for any $G_1, G_2 \in \mathcal{G}$.

(ii) Let $G_1, G_2 \in \mathcal{G}$ be arbitrarily chosen and suppose $[G_1] = [G_2]$. Then, since G_1 and G_2 are isomorphic, the minimum

$$\min\{h_A(G_1^A, G_2^A) \mid A \text{ is an amalgam of } G_1 \text{ and } G_2\}$$

is achieved when $A = G_1^A = G_2^A$. Therefore $h_A(G_1^A, G_2^A) = 0$ for an amalgam A of G_1 and G_2, and hence $H([G_1],[G_2]) = 0$. For the converse, let $G_1, G_2 \in \mathcal{G}$ be arbitrarily chosen and suppose $H([G_1],[G_2]) = 0$. Then there is an amalgam A of G_1 and G_2, such that $h_A(G_1^A, G_2^A) = 0$. Therefore G_1^A and G_2^A represent the same vertex in 2^A. This means that $G_1^A = G_2^A$ and therefore $[G_1] = [G_2]$.

(iii) Let $G_1, G_2 \in \mathcal{G}$ be arbitrarily chosen. Then

$$H([G_1],[G_2]) = \min\{h_A(G_1^A, G_2^A) \mid A \text{ is an amalgam of } G_1 \text{ and } G_2\} =$$

$$\min\{h_A(G_2^A, G_1^A) \mid A \text{ is an amalgam of } G_2 \text{ and } G_1\} = H([G_2],[G_1]).$$

$\square$

However, H is not a metric on $\mathcal{G}/_\sim$, see Example 4.15.

Example 4.15. *Let K_1, P_7 and W_7 be the graphs in Figure 7. Then $H([W_7],[K_1]) = 1$, $H([W_7],[P_7]) = 1$ and $H([P_7],[K_1]) = 3$. Therefore*

$$H([P_7],[K_1]) \leq H([P_7],[W_7]) + H([W_7],[K_1])$$

does not hold.

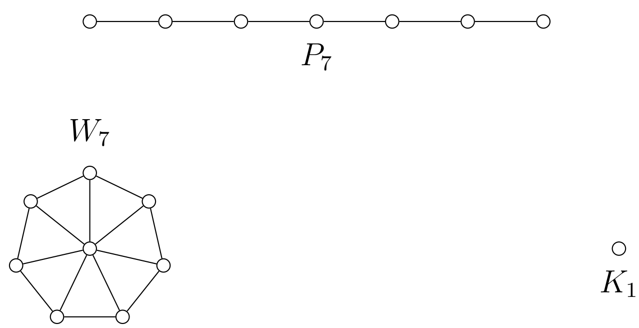

Figure 7: Graphs K_1, P_7 and W_7.

We will prove that for convex amalgams (defined below) the triangle inequality holds true.

Definition 4.16. *Let $G_1, G_2 \in \mathcal{G}$, let H_1 be a subgraph of G_1, and let H_2 a subgraph of G_2, where H_1 and H_2 are isomorphic graphs. If H_1 and H_2 are both convex, then any amalgam of G_1 and G_2 obtained by identifying H_1 and H_2 is called a convex amalgam of G_1 and G_2.*

In convex amalgams the intersection of covers is also convex and by definition connected. It follows from Remark 4.11 that h_A can be obtained by determining $d_{C(A)}$, which is easier.

Theorem 4.17. *Let $H_X([G_1],[G_2]) = \min\{h_A(G_1^A, G_2^A) \mid A$ is a convex amalgam of G_1 and $G_2\}$, for arbitrary $G_1, G_2 \in \mathcal{G}$. Then H_X is a metric on $\mathcal{G}/_\sim$.*

Proof. For

(i) $H_X([G_1],[G_2]) \geq 0$,

(ii) $H_X([G_1],[G_2]) = 0$ if and only if $[G_1] = [G_2]$, and

(iii) $H_X([G_1],[G_2]) = H_X([G_2],[G_1])$.

we follow the same line of thought as in Theorem 4.14.

Let $G_1, G_2, G_3 \in \mathcal{G}$ be arbitrary graphs. We prove that

$$H_X([G_1],[G_3]) \leq H_X([G_1],[G_2]) + H_X([G_2],[G_3]).$$

Let $i, j \in \{1,2,3\}$, where $i < j$. Then let $A_{i,j}$ be a convex amalgam of G_i and G_j with $d_{i,j} := d_{C(A_{i,j})}(G_i^{A_{i,j}}, G_j^{A_{i,j}}) = H_X([G_i],[G_j])$. In other words, $A_{i,j}$ is the graph which gives rise to the minimum for $H_X([G_i],[G_j])$. Denote by $H_{i,j}$ the convex graph $(G_i^{A_{i,j}}) \cap (G_j^{A_{i,j}})$.

Now, create an amalgam of $A_{1,2}$ and $A_{2,3}$ by identifying the vertices in the covers $G_2^{A_{1,2}}$ and $G_2^{A_{2,3}}$, denote the resulting graph by A. Note, this amalgam may not be the one giving rise to the minimum for $H_X([A_{1,2}],[A_{2,3}])$, but it clearly is a convex amalgam.

First, assume that the graphs G_1^A and G_3^A corresponding to G_1 and G_3, respectively, in the graph A have a non-empty intersection, denote the set of vertices in this intersection by S. Since the graph $\langle S \rangle = H_{1,2}^A \cap H_{2,3}^A$ is the intersection of two convex subgraphs of G_2^A, $\langle S \rangle$ is also convex in G_2^A (as well as in G_1^A and G_3^A). Therefore, clearly $\langle S \rangle$ is convex in A. Let $d' = d_{C(A')}(G_1^{A'}, G_3^{A'})$, where A' is the graph obtained from A by removing all vertices of graph G_2^A, which are not in the graphs G_1^A and G_3^A. Note that A' is a convex amalgam of graphs G_1 and G_3. Therefore $H_X([G_1],[G_3]) \leq d'$. It follows from Corollary 4.12 that there exist vertices $u \in S$ and $v \in V(G_i^{A'})$ for an index $i \in \{1,3\}$, such that $d_{A'}(u,v) = d'$. Without loss of generality suppose that $i = 3$. Let $P = (u = u_1, u_2, \ldots, u_k, u_{k+1}, \ldots, u_{d+1} = v)$ be a shortest path from u to v in A', where for each $j \leq k$, $u_j \in H_{2,3}^A$ and for each $j > k$, $u_j \notin H_{2,3}^A$. Clearly, P is also a shortest path from u to v in A. Then $d' = d_{A'}(u,u_k) + d_{A'}(u_{k+1},v) = d_{A'}(u,u_k) + d_{A_{2,3}}(u_{k+1},v) \leq d_{1,2} + d_{2,3}$.

Second, assume that the graphs G_1^A and G_3^A corresponding to G_1 and G_3, respectively, in the graph A have an empty intersection. Following Corollary 4.12 we choose the following vertices: $u_1 \in V(A_{1,2})$, $u_2 \in V(A_{2,3})$, $u_3 \in V(A_{1,3})$ and $v_1 \in V(H_{1,2}^{A_{1,2}})$, $v_2 \in V(H_{2,3}^{A_{2,3}})$, $v_3 \in V(H_{1,3}^{A_{1,3}})$, such that $d_{1,2} = d_{A_{1,2}}(u_1,v_1)$, $d_{2,3} = d_{A_{2,3}}(u_2,v_2)$ and $d_{1,3} = d_{A_{1,3}}(u_3,v_3)$. Without loss of generality assume that $u_1 \in V(G_1^{A_{1,2}})$, $u_2 \in V(G_2^{A_{2,3}})$, $u_3 \in V(G_1^{A_{1,3}})$. Let u_1^A,v_1^A, u_2^A, v_2^A be the vertices in A corresponding to vertices u_1, v_1, u_2, v_2, respectively. Next let u_3' and v_3' be the vertices in G_1^A corresponding to u_3 and v_3 in $G_1^{A_{1,3}}$. Finally, let v_3'' be the vertex in G_3^A corresponding to v_3 in $G_3^{A_{1,3}}$. See Figure 8 for reference.

Next we define the graph G as the graph obtained from A by identifying the vertices v_1^A and v_2^A. Denote the resulting vertex in G by x, also denote by

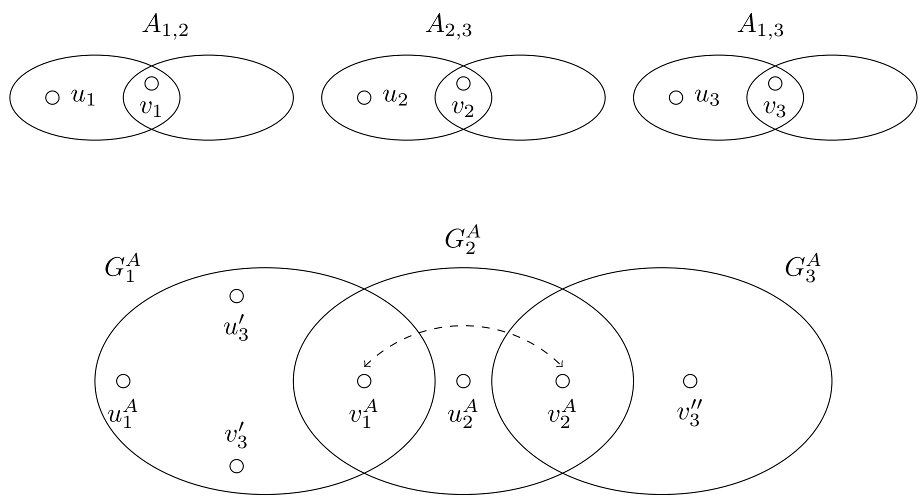

Figure 8: Graphs, vertices and notation from the proof of Theorem 4.17.

u_3^G and v_3^G the vertices in G corresponding to u_3' and v_3' in A, respectively. Note that distances between two vertices of G_1 (G_3) remain the same when observed in A or in G. Hence,

$$\begin{aligned} d_{1,3} = d_{A_{1,3}}(u_3, v_3) = d_A(u_3', v_3') = d_G(u_3^G, v_3^G) &\leq \\ d_G(u_3^G, x) + d_G(x, v_3^G) = d_A(u_3', v_1^A) + d_A(v_2^A, v_3'') &\leq \\ d_A(u_1^A, v_1^A) + d_A(v_2^A, u_2^A) = d_{A_{1,2}}(u_1, v_1) + d_{A_{2,3}}(u_2, v_2) &= \\ d_{1,2} + d_{2,3}. & \end{aligned}$$

□

Finally we define the Hausdorff distance $\mathcal{H} : \mathcal{G} \times \mathcal{G} \to \mathbb{R}$ on $\mathcal{G}$.

Definition 4.18. *For any graphs $G_1, G_2 \in \mathcal{G}$, we define*

$$\mathcal{H}(G_1, G_2) = H_X([G_1], [G_2]).$$

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

Let us point out that the Hausdorff distance is not a metric on $\mathcal{G}$, since from $\mathcal{H}(G_1, G_2) = 0$ it follows that $G_1 \sim G_2$ and not necessarily $G_1 = G_2$. The following theorem follows directly from Theorem 4.17.

Theorem 4.19. *Let $G_1, G_2, G_3 \in \mathcal{G}$ be arbitrary graphs. Then*

(*i*) $\mathcal{H}(G_1, G_2) \geq 0$,

(*ii*) $\mathcal{H}(G_1, G_2) = 0$ *if and only if* $G_1 \sim G_2$, (*iii*) $\mathcal{H}(G_1, G_2) = \mathcal{H}(G_2, G_1)$, *and*

(*iv*) $\mathcal{H}(G_1, G_3) \leq \mathcal{H}(G_1, G_2) + \mathcal{H}(G_2, G_3)$.

Example 4.20. *On Figure 9 there are all non-isomorphic convex amalgams of P_2 and P_3, which are denoted by A_1, A_2 and A_3. Moreover there are all non-isomorphic convex amalgams of Q_3 and P_3, which are denoted by B_1, B_2 and B_3.*

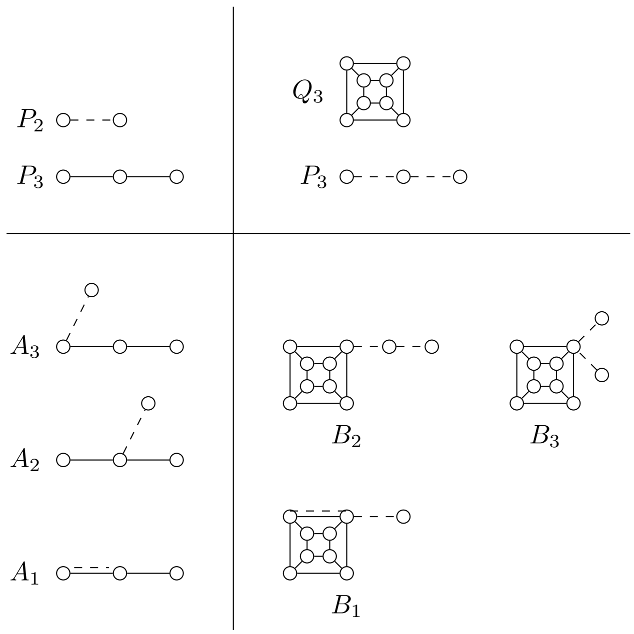

Figure 9: Graphs from Example 4.20.

Since $d_{C(A_1)}(P_2^{A_1}, P_3^{A_1}) = d_{C(A_2)}(P_2^{A_2}, P_3^{A_2}) = 1$ and $d_{C(A_3)}(P_2^{A_3}, P_3^{A_3}) = 2$, it follows that the Hausdorff distance between P_2 and P_3 is $\mathcal{H}(P_2, P_3) = 1$. Similarly, it follows from $d_{C(B_2)}(Q_3^{B_2}, P_3^{B_2}) = d_{C(B_3)}(Q_3^{B_3}, P_3^{B_3}) = 3$ and $d_{C(B_1)}(Q_3^{B_1}, P_3^{B_1}) = 2$, that the Hausdorff distance between Q_3 and P_3 is $\mathcal{H}(Q_3, P_3) = 2$.

5 Applications

We see applications of our method of measuring closeness of two graphs in all areas where the objects in question can be represented as graphs. Among others, such applications may be found in

(i) computer science (e.g. representations of networks and their comparisons);

(ii) chemistry (e.g. representations of molecules and their comparisons);

(iii) linguistics (e.g. representations of phrase structures and their comparisons);

(iv) physics (e.g. representations of complicated simulated atomic structures in condensed matter physics and their comparisons);

(v) sociology (e.g. representations of social networks and their comparisons);

(vi) biology (e.g. representations of species habitats and their comparisons).

Here we present one possible application of our method in biology, where similarity of two species is often studied by observing and comparing various parameters of two specimens (e.g. skull features, teeth positions, vein systems in leaves). One such method, recognized by biologists, is called landmark-based geometric morphometrics, where landmark (special points, e.g. intersection points of veins) coordinates are used as reference points to determine similarity of two objects compared (for an example see [7]). Unfortunately, this method fails when two such representations of objects differ immensely. Our method has no such limits.

Next we present an easy application of the Hausdorff distance in biology. We compare three different trees by using their leaves; two of them from the same tree species and one from a different tree species. We use the vein systems of the leaves to represent them as graphs, see Figures 10 and 11.

It can be easily checked that the Hausdorff distance between the graphs of leaves in Figure 10 is $\mathcal{H}(T_1, T_2) = 1$, so they are very close - related - with respect to the meaning of the Hausdorff distance.

Let us now compare graphs T_1 and T_3. Since T_1 is a convex subgraph of T_3, one can easily see that $\mathcal{H}(T_1, T_3) \geq 5$, therefore the two trees corresponding to T_1 and T_3 are not as related (w.r.t. the Hausdorff distance) as those corresponding to T_1 and T_2.

This example shows that Hausdorff distance can be used to determine a relationship between the three trees compared. Namely, with respect to the Hausdorff distance, the first two trees are more related than the first and the third.

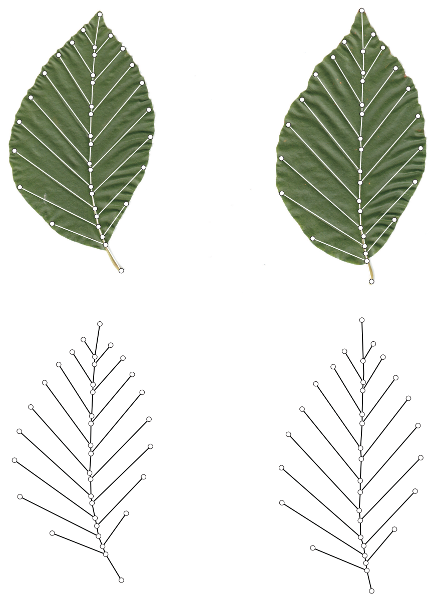

Figure 10: Two leaves from the same tree species and their graph representations T_1 and T_2.

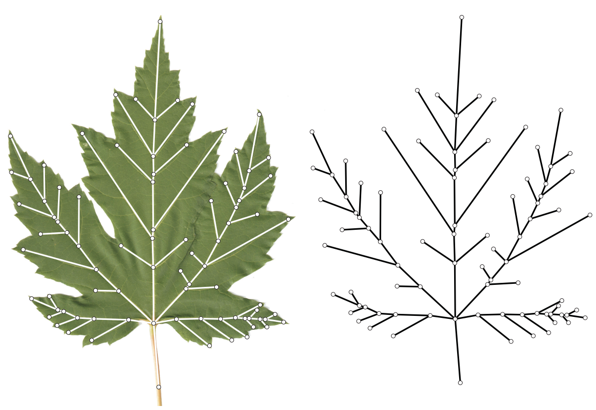

Figure 11: A leaf from a different tree species than those in Figure 10 and its graph representation T_3.

6 Open problems

In the last section we introduce some open problems about Hausdorff graphs and the Hausdorff distance. First we introduce a natural question that arises when constructing Hausdorff graphs and the introduced families of their subgraphs.

Question 6.1. *Let G and H be arbitrary graphs. Are the following statements equivalent?*

(*i*) *G is isomorphic to H.*

(*ii*) *2^G is isomorphic to 2^H.*

(*iii*) *$C(G)$ is isomorphic to $C(H)$.*

(*iv*) *$CI(G)$ is isomorphic to $CI(H)$.*

(*v*) *$K(G)$ is isomorphic to $K(H)$.*

It is obvious that (ii), (iii), (iv) and (v) follow from (i).

We have shown in Example 4.15 that H is not a metric on $\mathcal{G}/_\sim$. Then we obtained a metric on $\mathcal{G}/_\sim$ by applying the convex amalgams. In some applications, other types of amalgams may give better results about comparison of two objects (it is all up to the structure of studied objects and the properties to be compared). This is why we conclude the paper with two questions about obtaining new metrics on $\mathcal{G}/_\sim$ by applying so-called induced (or isometric) amalgams instead of the convex amalgams.

Definition 6.2. *Let $G_1, G_2 \in \mathcal{G}$, let H_1 be a subgraph of G_1, and let H_2 a subgraph of G_2, where H_1 and H_2 are isomorphic graphs. If H_1 and H_2 are both induced (isometric), then any amalgam of G_1 and G_2 obtained by identifying H_1 and H_2 is called an induced (isometric) amalgam of G_1 and G_2.*

Question 6.3. *For arbitrary graphs $G_1, G_2 \in \mathcal{G}$ let*
$H_I([G_1],[G_2]) = \min\{h_A(G_1^A, G_2^A) \mid A \textit{ is an induced amalgam of } G_1 \textit{ and } G_2\}$.
Is then H_I a metric on $\mathcal{G}/_\sim$?

Question 6.4. *For arbitrary graphs $G_1, G_2 \in \mathcal{G}$ let*
$H_M([G_1],[G_2]) = \min\{h_A(G_1^A, G_2^A) \mid A \textit{ is an isometric amalgam of } G_1 \textit{ and } G_2\}$.
Is then H_M a metric on $\mathcal{G}/_\sim$?

Acknowledgments

The authors thank Sandi Klavžar and Uroš Milutinović for constructive suggestions and discussions. We also thank the anonymous referees for their remarks and suggestions that helped us to correct several mistakes in the original version of the paper and to improve its presentation.

References

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

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

[3] H. Bunke, K. Shearer, *A graph distance metric based on the maximal common subgraph*, Pattern Recognit. Lett. 19, 1998, 255–259.

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

[5] M. Gorše Pihler, J. Žerovnik, *Partial cubes are distance graphs*, Discrete Math. 308, 2008, 820–826.

[6] A. Illanes, S. B. Nadler, *Hyperspaces. Fundamentals and recent advances*, Marcel Dekker, Inc., New York, 1999.

[7] F. Janžekovič, B. Kryštufek, *Geometric morphometry of the upper molars in European wood mice Apodemus*, Folia Zool. 53 (1), 2004, 47–55.

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

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

[10] L. C. Simon Romero, *The Hyperspace Graph of Connected Subgraphs*, Doctoral Dissertation, Eberly College of Arts and Sciences, Morgantown, West Virginia, 2005.

[11] L. C. Simon Romero, *Uniqueness of the hyperspace graph of connected subgraphs*, Topology Proc. 31 (1), 2007, 283–294.

[12] L. C. Simon Romero, *Relations between the hyperspace graph of connected subgraphs and the topological hyperspace of subcontinua*, Continuum Theory: in honor of Professor David P. Bellamy on the ocasion of his 60th bithday, Aportaciones Mat. Investig., 19, Soc. Mat. Mexicana, México, 2007, 123–133.

[13] S. B. Nadler, *Continuum theory. An introduction*, Marcel Dekker, Inc., New York, 1992.