universität wien

MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's Thesis

"Residual Finiteness In Hyperbolic Groups"

verfasst von / submitted by

Alexandra Edletzberger, BSc

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Master of Science (MSc)

Wien, 2019 / Vienna, 2019

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	UA 066 821
Studienrichtung lt. Studienblatt / degree programme as it appears on the student record sheet:	Mathematik
Betreut von / Supervisor:	Univ.-Prof. Goulnara Arzhantseva, PhD
Mitbetreut von / Co-Supervisor:	Christopher Cashen, PhD

Abstract

The question whether all hyperbolic groups are residually finite has remained open since Mikael Gromov posed it in 1987 in [14], although it has received a lot of attention. Trusting Daniel Wise's expertise in [50], most workers in the field expect a negative answer. Thus the search for a counterexample is in progress. One candidate for a non-residually finite, but hyperbolic group is the Heineken group.

In order to study it, we first need to understand the two fundamental properties of hyperbolicity and resiudal finiteness and explore equivalent definitions and (non-)examples. Furthermore the Hopf-property of groups is closely connected, as every finitely generated residually finite group is Hopfian.

A toolkit for determining whether a finitely generated group is residually finite is needed: It contains the concept of Tietze transformations, whose value is illustrated with a handful of examples. Then the classical result of Malcev affirms the residual finiteness of finitely generated linear groups. As the third tool serves Bass-Serre theory, which studies groups acting on trees. We work towards its fundamental result, the Structure Theorem.

The third Chapter is dedicated to the application and benefit of these tools by means of well-studied examples. Different arguments proving that free groups are residually finite are explored. The family of Baumslag-Solitar groups $BS(m,n)$ for the integer pair (m,n) is studied. We do not only work through the conditions on m and n that make $BS(m,n)$ a (non-)residually finite and/or (non-)Hopfian group, but also the Isomorphism Problem is tackled. Lastly, examples of groups that are non-residually finite by absence of finite quotients such as the famous Higman group are introduced.

This builds the foundation for focusing on the Heineken group. For it to be a valid candidate, we need to make sure that it is not finite (and thus trivially residually finite) and that it is hyperbolic. In order to do so the concept of automatic groups is crucial. The algorithm to find a shortlex automatic structure is explained. After trying to check the Heineken group for residual finiteness with the tools at hand, we try to make a first step towards finding an answer by focusing on finding finite quotients. Again, computational group theory is of use: The Coset Enumeration and the Low Index Method are explained. Further investigations with computational methods are suggested.

Die Frage, ob alle hyperbolischen Gruppen residuell endlich sind, ist offen, seitdem sie von Mikael Gromov in 1987 in [14] gestellt wurde, obwohl ihr schon viel Aufmerksamkeit gewidmet wurde. Vertraut man Daniel Wise's Expertise in [50], erwarten die meisten Forscherinnen und Forscher dieses Fachgebietes eine negative Antwort. Darum ist die Suche nach einem Gegenbeispiel im Gange. Ein Kandidat für eine nicht residuell endliche, aber hyperbolische Gruppe ist die Heineken-Gruppe.
Um sie zu erforschen müssen wir zuerst die beiden fundamentalen Eigenschaften der Hyperbolie und der residuellen Endlichkeit verstehen und equivalente Definitionen und Beispiele untersuchen. Außerdem ist die Hopf-Eigenschaft von Gruppen eng mit residueller Endlichkeit verbunden, da jede endlich erzeugte, residuell endliche Gruppe Hopf ist.
Hilfsmittel um festzustellen, ob eine endlich erzeugte Gruppe residuell endlich ist, werden benötigt: Das erste sind die Tietze-Transformationen, deren Nutzen anhand von einigen Beispielen illustriert wird. Darauf folgt das klassische Resultat von Malcev, das bestätigt, dass alle endlich erzeugten, linearen Gruppen residuell endlich sind. Als drittes Hilfsmittel dient die Bass-Serre-Theorie, in der es um Gruppen geht, die auf Bäume wirken. Wir erarbeiten ihr fundamentales Resultat, den Struktursatz.
Das dritte Kapitel widmet sich anhand von wohlbekannten Beispielen den Anwendungen und dem Nutzen dieser Hilfsmittel. Unterschiedliche Argumente, die beweisen, dass freie Gruppen residuell endlich sind, werden vorgestellt. Die Familie der Baumslag-Solitar-Gruppen $BS(m,n)$ für das Paar von ganzen Zahlen (m,n) wird untersucht. Nicht nur alle Bedingungen an m und n, die $BS(m,n)$ zu einer (nicht) residuell endlichen und/oder (nicht) Hopfschen Gruppe machen, werden ausgearbeitet, sondern auch das Isomorphismus-Problem wird gelöst. Abschließend werden Gruppen wie die berühmte Higman-Gruppe vorgestellt, die aufgrund des Fehlens von endlichen Quotienten nicht residuell endlich sind.
Damit ist die Basis geschaffen, um den Fokus auf die Heineken-Gruppe zu legen. Um zu zeigen, dass sie ein valider Kandidat ist, müssen wir uns versichern, dass sie nicht endlich (und damit trivialerweise residuell endlich) und hyperbolisch ist. Dafür ist das Konzept von automatischen Gruppen essentiell. Der Algorithmus, um eine shortlex-automatische Gruppe zu finden, wird erklärt. Nach dem Versuch mit den entwickelten Hilfsmitteln zu prüfen, ob die Heineken-Gruppe residuell endlich ist, versuchen wir einen ersten Schritt in Richtung einer Antwort zu machen, indem wir uns auf die Suche nach endlichen Quotienten fokussieren. Auch hier ist computerunterstützte Gruppentheorie von Bedeutung: Nebenklassen-Numerierung und die Niedrige-Index-Methode werden erklärt und weitere Forschungsansätze vorgeschlagen.

Contents

1 Introduction

Ever since in [14], Gromov suggested to treat groups as geometric objects and gave the definition of a hyperbolic group, finding a connection of hyperbolicity and other group theoretic properties was of great interest. Although many properties have been understood, one intriguing question, already posed in Gromov's original work [14], remains open: Is every hyperbolic group resiudally finite?
While many mathematicians like Wise in [50] and in [25] with I. Kapovich, Agol, Groves and Manning in [1] or Arzhantseva in [3] work on exploring equivalent statements and connected conjectures, it is also tempting to answer the question according to common beliefs with "no" by giving a counterexample.
In order to be able to work in this direction, two developments need to be retraced: On the one hand, results essential for hyperbolic groups and residually finite groups need to be understood. It is important to work extensively with equivalent definitions and well-known examples to get a feeling for the properties and methods to determine them. One the other hand, a suitable candidate for a counterexample needs to be found. One might be the Heineken group. It has been in the center of attention since the 1980's, when Heineken asked in private communication with Neubüser, whether a certain finitely generated group - now called Heineken group - is trivial or at least finite, since its presentation can be viewed as the generalization of a presentation of a group that turned out to be trivial. As often seen, it can be hard to find an answer to the question of finiteness of a group.
A lot of computational effort was put to finding finite quotients of the Heineken group in order to prove that it is finite. But only a handful of finite quotients were determined. From 1988 until the introduction of automatic structures in the 1990's no progress at all was made - but the implementation of the algorithm to find an automatic structure was a big game changer as discussed in [20]. It proved that the Heineken group is infinite - in contrast to Heineken's conjecture. Also, with automaticity it can be proven that the Heineken group is hyperbolic. So the research focus shifted to answering whether the Heineken group is residually finite. Since it has turned out to be hard to find finite quotients, it might not have a lot of them. That could make proving non-residual finiteness easier than in other groups. We aim to explore both research direction from scratch.

We start with a reminder of two definitions that are crucial for everything forthcoming. All further important fundamental group theoretical concepts can be looked up in [5].

Definition 1.1. Let F be a group and S a set with a map $i : S \to F$. F is called *free over* S if for every group H and for every map $j : S \to H$ there is a unique homomorphism $f : F \to H$ such that $f \circ i = j$.

It is easily checked that equivalently we can use the following:

Definition 1.2. A group F is called a *free group generated by* S if the set S generates F freely and there are no non-trivial relations between elements of F.

Definition 1.3. Let F be a free group over S. The cardinality of S is called the rank r of the group, i.e. $r = rank(F) \coloneqq |S|$.

Remark 1.1. We often denote a free group of rank $r \geq 2$ by F_r and the free group generated by the set S by F_S.

Example 1.1.

(i). The free group generated by the empty set is the trivial group.

(ii). The free group generated by $S = \{s\}$ is isomorphic to $\mathbb{Z}$.

Definition 1.4. Let G be a group generated by the set S. Then the *Cayley graph* $Cay(G, S)$ of G with respect to S is a (often coloured,) directed graph. Its vertex set $V(Cay(G, S))$ is identified with G, so every element $g \in G$ has an assigned vertex $v \in V(Cay(G, S))$. For every $g \in G$ and $s \in S$, there is a directed edge $(g, gs) \in E(Cay(G, S))$, connecting the vertices g and gs.

Example 1.2. Since a free group has no non-trivial relations, its Cayley graph with respect to a free generating set cannot contain any circuits and is therefore a tree. Also, the converse is true: A group whose Cayley graph is a tree, has to be free, as shown in [5, p. 65, Proposition 8.1].

1.1 Hyperbolic Groups

Unless stated otherwise, all proofs in this chapter are guided by Bowditch's lecture notes [6] and Bridson's and Haeflinger's standard reference [7].

1.1.1 Hyperbolic spaces

Definition 1.5. Let (X, d) be a metric space and let $I \subseteq \mathbb{R}$ be a closed interval with $\gamma : I = [a, b] \to X$ a map.

(i). The map γ is called a *geodesic* if

$$d(\gamma(t), \gamma(u)) = |t - u|$$

for any $t, u \in I$.

(ii). The map γ is called (λ, h)-*quasi-geodesic* for $\lambda \geq 1$ and $h \geq 0$ if

$$\tfrac{1}{\lambda}\,|t-u| - h \leq d(\gamma(t), \gamma(u)) \leq \lambda\,|t-u| + h$$

for any $t, u \in I$. The map γ is called *quasi-geodesic* if there is a pair (λ, h) such that γ is (λ, h)-quasi-geodesic.

(iii). The metric space (X, d) is called a *geodesic space* if for any pair of points $x, y \in X$ there are real numbers $a, b \in \mathbb{R}$ and a geodesic $\gamma : I \coloneqq [a, b] \to X$ such that $\gamma(a) = x$ and $\gamma(b) = y$.

(iv). The metric space (X, d) is called a *quasi-geodesic space* if there is a pair (λ, h) such that for any pair of points $x, y \in X$ there are real numbers $a, b \in \mathbb{R}$ and a (λ, h)-quasi-geodesic $\gamma : I \coloneqq [a, b] \to X$ such that $\gamma(a) = x$ and $\gamma(b) = y$.

Remark 1.2. In general a geodesic does not have to be unique. Furthermore quasi-geodesics need not be continuous.

Example 1.3. Since every word in a group $G = \langle\, S \mid R \,\rangle$ has a minimal reduced presentation in elements of S, the Cayley graph $Cay(G, S)$ of G equipped with the word metric is always a geodesic space.

Remark 1.3. Forthcoming we may often abuse the notation in a sense that we simply write γ for the image of the geodesic $\gamma : I \to X$ in the metric space (X, d). Moreover we may often write simply X for the metric space (X, d) when it is clear which metric is being used.

Remark 1.4. For a geodesic $\alpha \subseteq X$ in the geodesic space X, we denote a segment of α starting at $x \in \alpha \subseteq X$ and ending at $y \in \alpha \subseteq X$ with $\alpha[x, y] \subseteq X$. If the name of the map is not specified, we denote the geodesic segment between $x \in X$ and $y \in X$ simply by $[x, y] \subseteq X$.

Definition 1.6. Let (X, d) be a metric space and let $\gamma : [a, b] \to X$ be a curve with $[a, b] \subseteq \mathbb{R}$ an interval. Then the *length* $l(\gamma)$ *of the curve* γ is defined as

$$l(\gamma) = \sup_{a=t_0 \leq t_1 \leq \cdots \leq t_n = b} \sum_{i=0}^{n-1} d(\gamma(t_i), \gamma(t_{i+1}))\,,$$

where the supremum is taken over all possible partitions of the interval $[a, b]$ of the form $\{a = t_0, t_1, \ldots, t_n \;=\; b\} \subseteq [a, b]$ without any bound on $n \in \mathbb{N}$. If a curve γ is of finite length $l(\gamma) < \infty$, then it is called *rectifiable*.

Definition 1.7. Let (X, d) be a metric space. Let $x \in X$, $Q \subseteq X$ and $r \geq 0$.

- The *closed r-neighborhood of x* is defined as

$$\mathcal{N}(x,r) \coloneqq \{y \in X \mid d(x,y) \leq r\}\,.$$

- The *closed $r-$neighborhood of Q* is defined as

$$\mathcal{N}(Q,r) \coloneqq \bigcup_{x\in Q} \mathcal{N}(x,r)\,.$$

- The set $Q \subseteq X$ is called *r-dense* if $\mathcal{N}(Q,r) = X$.
- The set $Q \subseteq X$ is called *cobounded* if there is an $r' \geq 0$ such that Q is r'-dense.

Definition 1.8. Let (X,d) be a metric space.

(i). A triangle in X, whose edges lie on (quasi-)geodesics, i.e. they are (quasi-)geodesic segments, is called a *(quasi-)geodesic triangle.*

Let (α,β,γ) be such a (quasi-)geodesic triangle.

(ii). The (geodesic) triangle (α,β,γ) is called *δ-slim* for $\delta \in \mathbb{R}$ if

$$\alpha \subseteq \mathcal{N}(\beta,\delta) \cup \mathcal{N}(\gamma,\delta),\ \beta \subseteq \mathcal{N}(\alpha,\delta) \cup \mathcal{N}(\gamma,\delta) \text{ and } \gamma \subseteq \mathcal{N}(\alpha,\delta) \cup \mathcal{N}(\beta,\delta)\,.$$

(iii). A point $p \in X$ is called a *k-centre* of the triangle (α,β,γ) for $k \geq 0$ if

$$max\{d(p,\alpha), d(p,\beta), d(p,\gamma)\} \leq k\,.$$

Remark 1.5. If we want to specify the vertices $x,y,z \in X$ of a (quasi-)geodesic triangle, we often write $([x,y],[y,z],[z,x]) \subseteq X$ or $(\alpha[x,y],\beta[y,z],\gamma[z,x]) \subseteq (\alpha,\beta,\gamma) \subseteq X$.

Definition 1.9. Let (X,d) be a metric space and let $P,Q \subseteq X$ be subsets. The *Hausdorff distance* between P and Q is defined as

$$\mathcal{D}(P,Q) \coloneqq inf\{r \in [0,\infty) \mid P \subseteq \mathcal{N}(Q,r), Q \subseteq \mathcal{N}(P,r)\}\,.$$

Finally, we are ready to define a hyperbolic space. It turns out that we can use three equivalent definitions:

Definition 1.10 (Rips-condition)**.** The geodesic space (X,d) is called a *δ-hyperbolic* or *δ-hyperbolic space* if any (geodesic) triangle $(\alpha,\beta,\gamma) \subseteq X$ is δ-slim. (X,d) is called *hyperbolic* or a *hyperbolic space* if there is a $\delta \geq 0$ such that every (geodesic) triangle is δ-slim. The constant $\delta \geq 0$ is called a *hyperbolicity constant* of X.

Definition 1.11. The geodesic space (X, d) is called *k-hyperbolic* or *k-hyperbolic space* if every (geodesic) triangle $(\alpha, \beta, \gamma) \subseteq X$ has a k-centre. (X, d) is called *hyperbolic* or *hyperbolic space* if there is a $k \geq 0$ such that every (geodesic) triangle has a k-centre. The constant $k \geq 0$ is called *hyperbolicity constant* of X.

Less often we use the concept of thinness: To introduce it, let (X, d) be a geodesic space and let $([x, y], [y, z], [z, x]) \subseteq X$ be a geodesic triangle. Since by the triangle inequality there are positive integers $a, b, c \in \mathbb{N}$ such that

$$d(x, y) = a + b, \quad d(y, z) = b + c \quad \text{and} \quad d(z, x) = c + a\,,$$

we can find an isometry from the vertex set $\{x, y, z\}$ of the triangle to the vertices of a tripod T, which has edges of length a, b and c. This isometry extends uniquely to a map $\chi : ([x, y], [y, z], [z, x]) \to T$ that is an isometry when restricted to an edge of the geodesic triangle. By definition for every $t \in T$ its fibre $\chi^{-1}(t)$ in X

$$\chi^{-1}(t) \coloneqq \{a \in ([x, y], [y, z], [z, x]) \mid \chi(a) = t\} \subseteq X$$

consists of at most two elements, except for the center $o_T \in T$ of the the tripod. Its fiber consists of three elements. This leads to the definition:

Definition 1.12. Let (X, d) be a geodesic space and let $([x, y], [y, z], [z, x]) \subseteq X$ be a geodesic triangle. Let $\chi : ([x, y], [y, z], [z, x]) \to T$ be the map defined above.

- The (three) elements of the fibre $\chi^{-1}(o_T) \coloneqq \{i_x, i_y, i_z\}$ of $o_T \in T$ are called *internal points* and are indexed by the vertex of the triangle opposite to it. The diameter of the fiber $\chi^{-1}(o_T)$ of $o_T \in T$ is called the *insize* $\mathcal{I}$ of the triangle $([x, y], [y, z], [z, x]) \subseteq X$.
- The triangle $([x, y], [y, z], [z, x]) \subseteq X$ is said to be *δ-thin* for some $\delta \geq 0$ if for all $t \in T$ with $\{p_t, q_t\} \subseteq \chi^{-1}(t)$ it holds that

$$d(p_t, q_t) \leq \delta\,.$$

- If there is one $\delta \geq 0$ such that every geodesic triangle is δ-thin, X is called δ-hyperbolic.

Remark 1.6.

(i). The definition of the internal points $i_x, i_y, i_z \in X$ of a geodesic triangle $([x, y], [y, z], [z, x]) \subseteq X$ implies

$$d(x, i_y) = d(x, i_z)\,, \qquad d(y, i_x) = d(y, i_z) \quad \text{and} \quad d(z, i_x) = d(z, i_y)\,.$$

(ii). Note that in the Euclidean space $X = \mathbb{E}^2$ the internal points of a geodesic triangle are precisely the points where the edges meet the inscribed circle.

Hyperbolicity is a well-defined property, i.e. it does not matter whether Definition 1.10, Definition 1.11 or Definition 1.12 is used:

Lemma 1.1. *[7, p. 408, Proposition 1.17] Let (X, d) be a geodesic space. The following conditions are equivalent:*

(i). There is a $\delta_0 \geq 0$ such that every geodesic triangle in X is δ_0-slim.

(ii). There is a $\delta_1 \geq 0$ such that every geodesic triangle in X is δ_1-thin.

(iii). There is a $k \geq 0$ such that every geodesic triangle has a k-centre.

Remark 1.7. Forthcoming we will apply Definition 1.10, Definition 1.11 or Definition 1.12 dependent on whichever is more convenient in dealing with the current problem. If necessary, indication of which definition is used gives the name of the variable of the hyperbolicity constant.

Example 1.4. Any tree T is 0-hyperbolic: If we choose any three distinct points $x, y, \rho \in V(T)$ in the vertex set of T and connect them by geodesics to obtain a geodesic triangle $([x, y], [y, \rho], [\rho, x]) \subseteq T$, they will always form a (maybe degenerated) tripod as illustrated in Figure 1 adopted from [47] below, since T does not have any circuits.

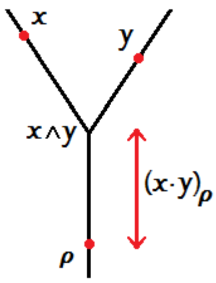


Figure 1: *Three geodesics connecting three vertices $x, y, \rho \in V(T)$ in a tree always have to pass through one common vertex, often denoted by $x \wedge y$, turning the geodesic triangle into a tripod. The distance $d(x \wedge y, \rho)$ between $x \wedge y$ and ρ is precisely $(x \cdot y)_\rho$ [c.f. Equation (1) in Definition 1.14].*

Hence the geodesics can be written as unions

$$
\begin{aligned}
[x, y] &= [x, x \wedge y] \cup [x \wedge y, y] \\
[y, \rho] &= [y, x \wedge y] \cup [x \wedge y, \rho] \\
[\rho, x] &= [\rho, x \wedge y] \cup [x \wedge y, x]\,,
\end{aligned}
$$

where $x \wedge y \in V(T)$ denotes center of the tripod, i.e. the vertex where all three geodesics intersect. So by taking the union of two geodesics, for example

$$\begin{aligned}[x,y] \cup [y,\rho] &= [x, x\wedge y] \cup [x \wedge y, y] \cup [y, x\wedge y] \cup [x\wedge y, \rho] \\ &= [x, x\wedge y] \cup [x\wedge y, y] \cup [x \wedge y, \rho]\,,\end{aligned}$$

the third geodesic is already contained. Hence the triangle is 0-slim.

Remark 1.8. Because of Example 1.4 we often think of δ-slim triangles as thickened tripods and of δ-hyperbolic spaces as thickened trees.

Definition 1.13. Let G be a finitely generated group. Then G is called a *hyperbolic group* if it has some finite generating set S such that the Cayley graph $Cay(G,S)$ is a hyperbolic space.

Remark 1.9. We will see in Chapter 1.1.2 that the choice of Cayley graph does not matter for the property of hyperbolicity of a group, i.e. if there is one finite generating set S of a group G such that the Cayley graph $Cay(G,S)$ is a hyperbolic space, the Cayley graphs for all finite generating sets are hyperbolic spaces.

Example 1.5. Any free group is hyperbolic by Example 1.4, since by Example 1.2 the Cayley graph with respect to a free generating set is always a tree.

Example 1.6. Any finite group is hyperbolic, since it has a finite generating set such that the Cayley graph is finite, hence there is a path of longest length l, so the group is l-hyperbolic.

In order to give a complete description of hyperbolicity, we should not omit its fourth definition, due to Mikhail Gromov [14]:

Definition 1.14. Let (X,d) be a metric space and let $x \in X$. The *Gromov product* of $y,z \in X$ with respect to x is defined to be

$$(y \cdot z)_x = \tfrac{1}{2}\left[d(y,x) + d(z,x) - d(y,z)\right]. \tag{1}$$

Remark 1.10. It is apparent from the definition that the Gromov product is commutative.

Remark 1.11. Let $T = ([x,y],[y,z],[z,x]) \subseteq X$ be a geodesic triangle in X.

(i). By definition of the internal points,

$$(y \cdot z)_x = d(x, i_y) = d(x, i_z)\,.$$

(ii). Furthermore

$$d(x,[y,z]) \leq (y \cdot z)_x + \mathcal{I}\,,$$

where $\mathcal{I} \geq 0$ denotes the insize of T.

(iii). If T is δ-thin for $\delta \geq 0$, then

$$|d(x,[y,z]) - (y \cdot z)_x| \leq \delta\,.$$

Definition 1.15 (Gromov)**.** Let (X,d) be a metric space and let $\delta \geq 0$ be a constant. X is called (δ)-*hyperbolic* if

$$(x \cdot y)_w \geq min\{(x \cdot z)_w, (y \cdot z)_w\} - \delta \tag{2}$$

for all points $x, y, z, w \in X$.

Equivalently we can use:

Definition 1.16. (Four-Point Condition) Let (X,d) be a metric space and let $\delta \geq 0$ be a constant. The condition

$$d(x,w) + d(y,z) \leq \max\{d(x,y) + d(z,w), d(y,w) + d(x,z)\} + 2\delta\,, \tag{3}$$

for $x, y, z, w \in X$ is called the *(Gromov-)Four-Point Condition.* X is called (δ)-*hyperbolic* if the the Four-Point Condition is satisfied for all $x, y, z, w \in X$.

Again, we want to check that the property of hyperbolicity stays coherent:

Lemma 1.2. *[2, p. 6, Theorem 2.3] Let (X,d) be a metric space.*

- *X is (δ)-hyperbolic for some $\delta \geq 0$ in the sense of Definition 1.15 if and only if it is (δ')-hyperbolic in the sense of Definition 1.16 for some $\delta' \geq 0$.*
- *Every geodesic triangle in X is δ-slim for some $\delta \geq 0$, i.e. X is δ-hyperbolic in the sense of Definition 1.10 if and only if X is (δ')-hyperbolic for some $\delta' \geq 0$ in the sense of Definition 1.15.*

1.1.2 Well-definedness of Hyperbolic Groups

To see that the hyperbolicity of a group is a well-defined property, i.e. independent of the choice of the generating set S, we have to do some work and broaden first our knowledge about (quasi-)geodesics:

Proposition 1.3 (Stability of Quasi-Geodesics)**.** *Let (X,d) be a δ-hyperbolic space and let $p \in X$ and $q \in X$ be two points connected by a geodesic segment $\alpha = [p,q] \subseteq X$. Let β be a (λ, h)-quasi-geodesic connecting the same points $p, q \in X$ as α. Then there is a constant $r \geq 0$ dependent only on λ, h and δ bounding the Hausdorff distance of α and β, i.e.*

$$\mathcal{D}(\alpha, \beta) < r\,.$$

Proof. To tackle this proof we need to introduce two lemmata:

Lemma 1.4. *[7, p. 400, Proposition 1.6] Let (X, d) be a δ-hyperbolic geodesic space with $\gamma \subseteq X$ a continuous, rectifiable path. Let $\alpha = [p, q] \subseteq X$ be the geodesic segment connecting the endpoints $p, q \in X$ of γ. Then for every $x \in \alpha$*

$$d(x, \gamma) \leq \delta\, |\log_2 l(\gamma)| {+} 1\,, \tag{4}$$

where $l(\gamma)$ denotes the length of γ.

Lemma 1.5. *[7, p. 403, Lemma 1.11] Let (X, d) be a geodesic space with a (λ, h)-quasi-geodesic $\gamma: [a, b] \to X$ for $[a, b] \subseteq \mathbb{R}$. Then there is a continuous (λ, h')-quasi-geodesic $\gamma' : [a, b] \to X$ satisfying the following conditions:*

(1). $\gamma'(a) = \gamma(a)$ *and* $\gamma'(b) = \gamma(b)$.

(2). $h' = 2\,(\lambda + h)$.

(3). $l(\gamma'|_{[t,t']}) \leq k_1\, d(\gamma'(t), \gamma'(t')) + k_2$

$$\textit{with} \quad k_1 = \lambda\,(\lambda + h) \quad \textit{and} \quad k_2 = (\lambda\, h' + 3)(\lambda + h)\,.$$

(4). The Hausdorff distance between γ and γ' is bounded by

$$\mathcal{D}(\gamma, \gamma') < (\lambda + h)\,.$$

First replace the (λ, h)-quasi-geodesic β via Lemma 1.5 by the continuous (λ, h')-quasi-geodesic β' satisfying the conditions *(1). - (4).* . Define the constant

$$D \coloneqq sup\{d(x, \beta') \mid x \in [p, q] = \alpha\}$$

and let $x_o \in \alpha \subseteq X$ be the point where the supremum is attained. Hence, the open ball around x_o with radius D does not meet β'.
Next define $y, z \in \alpha$:

If $d(x_o, p) > 2D$ define $y \in [p, x_o] \subseteq \alpha$ such that $d(y, x_o) = 2D\,,$
if $d(x_o, p) \leq 2D$ define $y \coloneqq p \in [p, x_o] \subseteq \alpha\,,$

and

if $d(q, x_o) > 2D$ define $z \in [x_o, q] \subseteq \alpha$ such that $d(x_o, z) = 2D\,,$
if $d(q, x_o) \leq 2D$ define $z \coloneqq q \in [x_o, q] \subseteq \alpha\,.$

Furthermore we fix elements $y', z' \in \beta'$ such that

$$d(y, y') \leq D \quad \text{and} \quad d(z, z') \leq D\,,$$

which exist by the definition of D and choose geodesic segments $[y, y']$ and $[z, z']$. Now consider the path $\tau \subseteq X$ connecting y and z that goes along $[y, y']$, $\beta'|_{[y',z']}$ and $[z', z]$ as illustrated in Figure 2 below.

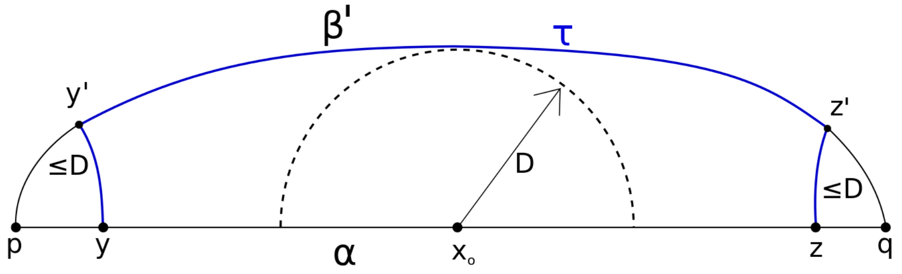


Figure 2: *The points p and q are connected by the geodesic α and the quasi-geodesic β'. The points y and z on α are connected via the path τ that passes through y' and z' on β'.*

By construction, τ lies outside of the ball around x_o with radius D. Since

$$d(y', z') \leq d(y', y) + d(y, z) + d(z, z') \leq d(y', y) + d(y, x_o) + d(x_o, z) + d(z, z') \leq 6D$$

we obtain

$$\begin{aligned} l(\tau) &= l([y, y']) + l(\beta'|_{[y',z']}) + l([z, z']) \\ &\overset{(3).}{\leq} D + k_1\, d(\beta'(y'), \beta'(z')) + k_2 + D \\ &\leq D + k_1 6D + k_2 + D \\ &= 2D + 6Dk_1 + k_2\,. \end{aligned}$$

With the fact that $d(x_o, \beta) = D$ and Lemma 1.4 we deduce

$$\begin{aligned} D - 1 &\leq \delta\,|\log_2 l(\tau)| \\ &\leq \delta\,|\log_2(2D + 6Dk_1 + k_2)|\,. \end{aligned}$$

Thus we have found an upper bound D_o of D in terms of λ, h and δ.
Next, we define

$$r' := D_o\,(1 + k_1) + \tfrac{k_2}{2}$$

and show that $\beta' \subseteq \mathcal{N}(\alpha, r')$.

In order to do so consider the maximal subinterval $[a', b'] \subseteq \mathbb{R}$ in the preimage $[a, b]$ of β' such that

$$\beta'([a', b']) \subsetneqq \mathcal{N}(\alpha, D_o)\,.$$

By connectedness there is a $w \in \alpha \subseteq \mathcal{N}(\beta', D_o)$ and there are two elements $t \in [a, a']$ and $t' \in [b', b]$ such that

$$d(w, \beta'(t)) \leq D_o \quad \text{ and } \quad d(w, \beta'(t')) \leq D_o\,.$$

Thus, $d(\beta'(t), \beta'(t')) \leq 2D$. Hence, by condition *(3).* of Lemma 1.5 we obtain

$$l(\beta'|_{[t,t']}) \leq 2k_1 D_o + k_2$$

and thus

$$\beta' \subseteq \mathcal{N}(\alpha, r')\,.$$

Lastly, we can define $r := r' + \lambda + h$, since then by *(4).* of Lemma 1.5 the claim holds. □

Corollary 1.6. *Let (X, d) be a k-hyperbolic geodesic space. Any (λ, h)-quasi-geodesic triangle $(\bar{\alpha}, \bar{\beta}, \bar{\gamma}) \subseteq X$ with vertices $x, y, z \in X$ has a t-centre, where $t \geq 0$ is only dependent on λ, h and k.*

Proof. Let $(\alpha, \beta, \gamma) \subseteq X$ be a geodesic triangle connecting the same vertices $x, y, z \in X$ as $(\bar{\alpha}, \bar{\beta}, \bar{\gamma}) \subseteq X$. Then any k-centre of (α, β, γ), which exists since X is k-hyperbolic, by Lemma 1.1 and by Proposition 1.3 has to be a $(k + r)$-centre of $(\bar{\alpha}, \bar{\beta}, \bar{\gamma})$, with $r \geq 0$ depending only on λ, h and k. □

Next, we introduce a map, that turns out to pass hyperbolicity between spaces:

Definition 1.17. Let (X, d) and (X', d') be two metric spaces. A map $\phi : X \to X'$ is called a $(\lambda, h)-$*quasi-isometric embedding* if there are constants $\lambda \geq 1$ and $h \geq 0$ such that for any $x, y \in X$

$$\tfrac{1}{\lambda}\, d(x, y) - h \leq d'(\phi(x), \phi(y)) \leq \lambda\, d(x, y) + h\,. \tag{5}$$

If in addition there is a constant $c \geq 0$ such that for every $x' \in X'$ there is an $x \in X$ such that

$$d'(\phi(x), x') \leq c\,, \tag{6}$$

then the map ϕ is called (λ, h, c)-*quasi-isometry.* If there exists a quasi-isometry between them, the metric spaces (X, d) and (X', d') are called *quasi-isometric.*

Definition 1.18. Let (X, d) and (X', d') be two metric spaces with a map $\phi : X \to X'$. A map $\psi : X' \to X$ is called a *quasi-inverse* of ϕ if there is a constant $c \geq 0$ such that

$$d'(\phi(\psi(x')), x') \leq c \quad \text{and} \quad d(\psi(\phi(x)), x) \leq c$$

for all $x \in X$ and $x' \in X'$.

Lemma 1.7. *If $\phi : X \to X'$ is a quasi-isometry between the metric spaces (X, d) and (X', d'), then there exists a map $\psi : X' \to X$, that is a quasi-isometric quasi-inverse of ϕ.*

Proof. [45, p. 16f, Proposition 3.0.3.] Let $\phi : X \to X'$ be a (λ, h, c)-quasi-isometry. Define the map $\psi : X' \to X$ such that the image of $x' \in X'$ is some $x \in X$ such that the property

$$d'(\phi(x), x') \leq c$$

holds. Also, by the property of ϕ, we know that for every $x' \in X'$

$$d'(\phi(\psi(x')), x') \leq c\,.$$

Furthermore, for any $x \in X$ we have

$$d(\psi(\phi(x)), x) \leq \lambda\, d'(\phi(\psi(\phi(x))), \phi(x)) + \lambda\, h \leq \lambda\,(h + c)\,.$$

Thus, $d'(\phi(\psi(x'), x')$ and $d(\psi(\phi(x)), x)$ are both bounded, say by a constant $c' \geq 0$, i.e. ψ is a quasi-inverse.
Since that fact already ensures condition (6) of a quasi-isometry, it is left to check that ψ is a quasi-embedding: For any $x'_1, x'_2 \in X'$ we have

$$\begin{aligned} d(\psi(x'_1), \psi(x'_2)) &\leq \lambda\, d'(\phi(\psi(x'_1)), \phi(\psi(x'_2))) + \lambda\, h \\ &\leq \lambda\, d'(x'_1, x'_2) + 2\,\lambda\, c' + \lambda\, h\,, \end{aligned}$$

which shows the right inequality of the condition (5) for a quasi-embedding and analogously with

$$\begin{aligned} d(\psi(x'_1), \psi(x'_2)) &\geq \tfrac{1}{\lambda}\, d'(\phi(\psi(x'_1)), \phi(\psi(x'_2))) - \tfrac{h}{\lambda} \\ &\geq \tfrac{1}{\lambda}\, d(x'_1, x'_2) - \tfrac{2c'}{\lambda} - \tfrac{h}{\lambda}\,, \end{aligned}$$

we ensure the left one. Thus, ψ is a quasi-isometry. $\square$

Proposition 1.8. *Let (X, d) and (X', d') be two quasi-isometric geodesic metric spaces. Then X is hyperbolic if and only if X' is hyberbolic.*

Proof. Let $\phi : X' \to X$ be a (λ, h, c)-quasi-isometry and let $(\alpha, \beta, \gamma) \subseteq X'$ be a geodesic triangle in X'. Then $(\phi \circ \alpha, \phi \circ \beta, \phi \circ \gamma) \subseteq X$ is a quasi-geodesic triangle in the hyperbolic space X and by Corollary 1.6 it has a t-centre. Thus, by Lemma 1.1 it is also δ-slim for some $\delta \geq 0$. That means that for all $x \in \alpha$ there is a $y \in \beta \cup \gamma$ such that

$$d(\phi(x), \phi(y)) \leq \delta \,.$$

Moreover, since ϕ is a (λ, h, c)-quasi-isometry,

$$d(x, y) \leq \lambda \, d(\phi(x), \phi(y)) + \lambda \, h \leq \lambda \, \delta + \lambda \, h \,.$$

Analogously, we can repeat this argument for the other edges of the triangle β and γ. Hence, the triangle $(\alpha, \beta, \gamma) \subseteq X'$ is δ'-slim with $\delta' := \lambda \, \delta + \lambda \, h$, which makes X' a hyperbolic space as well. The converse is completely analogous since by Lemma 1.7 with the quasi-isometry $\phi : X' \to X$ there exists a quasi-isometric inverse $\psi : X \to X'$. □

Lastly, we show that we can find a quasi-isometry between Cayley-graphs:

Proposition 1.9. *Let G be a group with two finite generating sets S, S'. Then there is a quasi-isometry $\phi : Cay(G, S) \to Cay(G, S')$.*

Proof. Denote $Cay(G, S) =: X$ and $Cay(G, S') =: X'$ with word metrics d and d' respectively. Since $V(X) = G = V(X')$ we can extend the identity map $id : G \to G$ to

$$\begin{aligned} \phi : X &\to X' & \\ v &\mapsto v & \text{for any } v \in V(X) \\ e = (x, y) &\mapsto [x, y] \subseteq X' & \text{for any } e = (x, y) \in E(X) \end{aligned}$$

such that $g \, \phi(p) = \phi(gp)$ for every $p \in X$ and $g \in G$.
Now define $r := max\{d'(1, a) \mid a \in S\}$. So ϕ maps every edge $e \in E(T)$ to a geodesic of at most length r in X'. Hence for any $x, y \in X$

$$d'(\phi(x), \phi(y)) \leq r \cdot d(x, y)$$

In the same way we can extend the identity map to $\psi : X' \to X$ and obtain for $s := max\{d(1, b) \mid b \in S'\}$

$$d(\psi(u), \psi(v)) \leq s \cdot d'(u, v) \,,$$

which lets us conclude that ϕ and ψ are both quasi-isometries and quasi-inverse to each other: Let $x \in X$ and $x' \in X'$. Since ϕ and ψ are both the identity map on vertices $x \in V(X)$ and $x' \in V(X')$ respectively, we have

$$d'(\phi(x), x) = d'(x, x) = 0 \quad \text{and} \quad d(\psi(x'), x') = d(x', x') = 0\,.$$

For an edge $e' \in E(X')$ it suffices to consider its midpoint $m' \in X'$. We do not know, whether there is any point in X that gets mapped to m' via ϕ. But we can simply take one of the vertices of e', which is definitely the image of an $x \in V(X)$ under ϕ. Hence the distance $d'(\phi(x), m')$ is at most $\frac{1}{2}$. The same argument holds for an edge $e \in E(X)$. Thus condition (6) is satisfied and ϕ and ψ are quasi-isometric embeddings for $c = \frac{1}{2}$.
Furthermore we get that

$$d'(\phi(\psi(x')), x') = d'(x', x') = 0 \quad \text{and} \quad d(\psi(\phi(x)), x) = d(x, x) = 0$$

for $x \in V(X)$ and $x' \in V(X')$. Again, we need to consider also the midpoint m' of an edge $e' \in E(X')$. It can get mapped under ψ to a point in X that is at most distance $\frac{s}{2}$ away. Mapping this point back to X' via ϕ, this distance can get stretched at most by r. Hence we have

$$d'(\phi(\psi(m')), m') \leq \tfrac{sr}{2}\,.$$

Analogously we obtain for the midpoint m of an edge $e \in E(X)$

$$d(\psi(\phi(m)), m) \leq \tfrac{rs}{2}\,.$$

Since with the midpoints this condition has to hold for all points lying on an edge, ϕ and ψ are quasi-inverses of each other with constant $c = \frac{rs}{2}$.
Lastly we get for $g, h \in V(X) = V(X')$

$$d'(g, h) \leq r\, d(\psi(g), \psi(h)) \leq s\, r\, d'(g, h)$$

and

$$d(g, h) \leq s\, d'(\phi(g), \phi(h)) \leq r\, s\, d(g, h)\,.$$

Thus, ϕ and ψ are both $(\lambda, 0)$-quasi-isometries for $\lambda = rs$, i.e. condition (5) holds. $\square$

Corollary 1.10. *Let G be a group generated by two distinct finite sets S and S'. Then the Cayley graph $Cay(G, S)$ is hyperbolic if and only if the Cayley graph $Cay(G, S')$ is hyperbolic.*

Remark 1.12. By Corollary 1.10, hyperbolicity of groups is a well-defined property, since it is independent of the choice of the generating set of the group.

Proof. This is immediate from Proposition 1.8 and Proposition 1.9. □

1.1.3 Švarc-Milnor-Lemma

When we have a hyperbolic space X given, we want to "extract" a hyperbolic group from it by finding an appropriate group action on it:

Definition 1.19. Let G be a group acting on the metric space (X, d) by isometry, i.e.

$$d(x, y) = d(gx, gy) \qquad \text{for every } x, y \in X \text{ and } g \in G\,.$$

- The action is called *properly discontinuous* if for all $r \geq 0$ and for all $x \in X$

$$|\{g \in G \mid d(x, gx) \leq r\}| < \infty\,. \tag{7}$$

- An action is called *cocompact* if the quotient space ${}_G\backslash^X$ is compact in the quotient topology.

Definition 1.20. A metric space (X, d) is called *proper* if it is complete and locally compact.

Remark 1.13. Some authors call an action satisfying condition (7) *metrically proper*. They use the term *properly discontinous* only if the space (X, d) is a proper geodesic space, where the property is equivalent to the following: For all compact subsets $K \subseteq X$ we have

$$|\{g \in G \mid gK \cap K \neq \emptyset\}| < \infty\,. \tag{8}$$

To show that, observe that in any proper geodesic space (X, d) the neighborhood $\mathcal{N}(x, r)$ is compact for all $x \in X$ and all $r \geq 0$: Fix an $x \in X$ and consider the set

$$A = \{r \in [0, \infty) \mid \mathcal{N}(x, r) \text{ is compact}\}\,. \tag{9}$$

Since X is proper, hence locally compact, A is not empty. So we can consider the supremum of A. Supposing then that $sup(A) < \infty$, gives a contradiction. With that fact at hand, we know now that in a proper geodesic space, the condition (7) for the property of proper discontinuity is equivalent to condition (8).

The following result will come in handy in the next chapter:

Lemma 1.11. *Let G be a group acting cocompactly on a locally compact graph X. Let $H \leq G$ be a finite index subgroup of G. Then H acts cocompactly on X as well.*

Proof. By assumption, the quotient $Y =_G \backslash^X$ is compact. First, we want to show that this is equivalent to saying that there is a compact subset $K \subseteq X$, whose image under the action of G covers X. It is evident that the existence of such a K ensures compactness of Y. For the converse, choose (by Axiom of Choice) for every $y \in Y$ an element $y' \in X$ that projects to y. Now for every $y' \in X$ there is an open neighborhood $U_{y'}$ with compact closure in X, since X is assumed to be locally compact. Projecting all $U_{y'}$ to Y and calling the images U_y yields an open covering of Y by the sets U_y. But Y is compact, hence there is a finite set $\{y_1, \ldots, y_n\} \subseteq Y$ such that the U_{y_i} cover Y. But that means that the set

$$K := \bigcup_{i=1}^{n} \overline{U}_{y_i} \subseteq X$$

is compact and projects onto Y. Hence, the image of K under the action of G covers X.

Now let $\{g_1, \ldots, g_H\} \subseteq G$ be the finite set of right coset representatives of H in G. Define the compact subset $g_1 K \cup \cdots \cup g_H K =: Z$ of X. Since H is a subgroup of G, the image of Z under the action of H covers X. Thus the quotient $Y =_H \backslash^X$ is compact, i.e. H acts cocompactly on X. □

Theorem 1.12 (Švarc-Milnor-Lemma)**.** *Let G be a group acting properly discontinuously and cocompactly on the geodesic metric space (X, d). Then:*

(i). G has a finite generating set A.

(ii). The Cayley graph $Cay(G, A)$ of G with respect to A is quasi-isometric to X.

Proof.

(i). Let $a \in X$. Consider the orbit $G \cdot a \in {}_G\backslash^X$. Since the action is cocompact, hence ${}_G\backslash^X$ is compact, there is an $r \geq 0$ such that ${}_G\backslash^X \subseteq \mathcal{N}(G \cdot a, r)$. Hence for any $x \in X$, the distance $d(G \cdot a, G \cdot x) := \min\{d(a, gx) \mid g \in G\} \leq r$. Thus, for every $x \in X$ there is a $g \in G$ such that $d(x, ga) \leq r$, which implies that $G \cdot a$ is r-dense. Define $k > 0$ to be $k = 2r + 1$.
Define a graph $\Delta = (V(\Delta), E(\Delta))$ that has vertex set $V(\Delta) = G$ and edge set $E(\Delta) = \{(g, h) \mid d(ga, ha) \leq k\}$. Since G acts properly discontinuously on X, the graph Δ is locally finite. Now we want to show that Δ is connected:
Let $g, h \in G$ and let $\alpha = [ga, ha] \subseteq X$ be a geodesic in X connecting

$ga, ha \in X$. Choose a sequence of points $ga = x_0, x_1, \ldots, x_{n-1}, x_n = ha$ in X such that $x_i \in \alpha$ for any $i \in \{0, 1, \ldots, n\}$ and $d(x_i, x_{i+1}) \leq 1$ for any $i \in \{0, 1, \ldots, n-1\}$. Next take $g_i \in G$ such that $d(x_i, g_i a) \leq r$ for any $i \in \{1, \ldots, n-1\}$ and set $g_0 = g$ and $g_n = h$. Then we can conclude

$$d(g_i a, g_{i+1} a) \leq 2r + 1 = k\,,$$

which means that $(g_i, g_{i+1}) \in E(\Delta)$. So $g_0 g_1 \cdots g_n$ connects $g \in G$ to $h \in G$ in Δ, thus Δ is connected.
Now define the set $A := \{g \in G \setminus \{1_G\} \mid d(a, ga) \leq k\}$. Since G acts on X properly discontinuously, A is finite. Furthermore A is symmetric. Choose $g, h \in V(\Delta)$ such that $(g, h) \in E(\Delta)$ then

$$d(ga, ha) \leq k \quad \Leftrightarrow \quad d(a, g^{-1} ha) \leq k \quad \Leftrightarrow \quad g^{-1} h \in A$$

which implies that $\Delta = Cay(G, A)$, hence G is generated by the finite set A.

(ii). We need to refine the construction from the proof of (i) above: Consider the map

$$\begin{aligned} f : \Delta &\to X \\ g &\mapsto ga && \text{for every } g \in V(\Delta) = G \\ (g, h) &\mapsto [ga, ha] \subseteq X && \text{for every } (g, h) \in E(\Delta)\,. \end{aligned}$$

Choose again $x_i \in X$ with $i \in \{0, 1, \ldots, n\}$ as above such that

$$n \leq d(ga, ha) + 1 = d(f(g), f(h)) + 1\,.$$

So if $d_\Delta(g, h) \leq n$, it follows that $d(f(g), f(h)) \leq r \cdot n$. But $G = V(\Delta)$ is cobounded in Δ and Ga is cobounded in X, hence f is a quasi-isometry. Since Δ is the Cayley graph $Cay(G, A)$ of G with respect to A, the theorem is proven.

□

Corollary 1.13. *A group G is finitely generated and hyperbolic if and only if it acts properly discontinuously cocompactly on a proper hyperbolic space X.*

Proof. First let G be generated by the finite set A and let its Cayley graph $Cay(G, A) =: X'$ be hyperbolic. Then G acts on X' by left translation, which is properly discontinuous and cocompact. Furthermore as A is finite, the Cayley graph X' is proper and by assumption hyperbolic, so the first implication is true.

For the converse let G be acting on the proper hyperbolic space X properly discontinuously cocompactly. Then by the first part of the Švarc-Milnor-Lemma 1.12 (i) G is finitely generated and by the second part (ii) the hyperbolic space X is quasi-isomorphic to $X' = Cay(G, A)$. Now Proposition 1.8 implies that X' has to be hyperbolic as well, which means that G is a hyperbolic group. $\square$

Example 1.7. By the Corollary 1.13 of the Švarc-Milnor-Lemma, any group that acts properly discontinuously on a locally finite tree is hyperbolic.

1.1.4 Bigons

Another concept that will help us determining whether a group is hyperbolic and will be crucial for the computational aspect is the following:

Definition 1.21. Let Γ be a graph. A pair of geodesics $\gamma, \gamma' \in \Gamma$ is called a *bigon* if

$$\gamma(0) = \gamma'(0) \quad \text{and} \quad \gamma(l) = \gamma'(l)\,,$$

where l denotes the length of γ.
We say the bigons are *ε-thin* if for any bigon γ, γ', we have

$$d(\gamma(t), \gamma'(t)) < \varepsilon$$

for all $t \in (0, l)$. A bigon γ, γ' is called *M-thick* for some $M \in \mathbb{N}$, if there is a $t \in (0, l)$ such that

$$d(\gamma(t), \gamma'(t)) \geq M\,.$$

Theorem 1.14. *[39, Theorem 1.4] Let $G = \langle X \mid R \rangle$ be a finitely generated group with Cayley graph $Cay(G, X)$. If there is an $\varepsilon \geq 0$ such that the bigons in $Cay(G, X)$ are ε-thin, then G is hyperbolic.*

1.2 Residual Finiteness

The following proofs are guided by [42] and [48].

Definition 1.22. Let P be a property a group can have. We say a group G is *residually P* if for every $g \in G \setminus \{1_G\}$ there is a normal subgroup $N \trianglelefteq G$ such that $g \notin N$ and the quotient G/N has property P.

Using that definition for finiteness gives the following:

Definition 1.23. A group G is *residually finite* if for every $g \in G \setminus \{1_G\}$ there is a normal subgroup $N \trianglelefteq G$ such that $g \notin N$ and the quotient G/N is finite.

Example 1.8. Any finite group is residually finite.

Example 1.9. Any simple infinite group is not residually finite.

We can easily show that residual finiteness has some useful characterizations:

Proposition 1.15. *Let G be a group. Then the following statements are equivalent:*

(i). *G is residually finite.*

(ii). *For every $g \in G \setminus \{1_G\}$ there is a homomorphism $\phi : G \to F$ into some finite group F such that $\phi(g) \neq 1_F$.*

(iii). $\bigcap\limits_{\substack{N \trianglelefteq G \\ |G/N| < \infty}} N = \{1_G\}$

(iv). *For every $g \in G \setminus \{1_G\}$ there is a subgroup $K \leq G$, such that $g \notin K$ and $|G/K| < \infty$.*

(v). $\bigcap\limits_{\substack{K \leq G \\ |G/K| < \infty}} K = \{1_G\}$

Proof. Let G be a group.

- $(i) \Rightarrow (ii)$: Let G be residually finite. Let $g \in G \setminus \{1_G\}$ and let $N \trianglelefteq G$ such that $g \notin N$ and $|G/N| < \infty$. Define the map

$$\begin{aligned} \phi : G &\to G/N \\ h &\mapsto hN\,. \end{aligned}$$

  Clearly, ϕ is a homomorphism and

$$\phi(g) = gN \overset{g \notin N}{\neq} N = 1_{G/N}\,,$$

  so ϕ has the desired properties.

- $(i) \Leftarrow (ii)$: Suppose there is a homomorphism $\phi : G \to F$ into some finite group F with $\phi(g) \neq 1_F$ for every $g \in G \setminus \{1_G\}$. Then $g \notin ker(\phi)$, but $ker(\phi)$ is a normal subgroup of G. By the Isomorphism Theorem we know $G/ker(\phi) \cong Im(\phi) \subseteq F$. But since F is finite we get $|G/ker(\phi)| = |Im(\phi)| \leq |F| < \infty$, so $ker(\phi)$ meets the conditions that make G a residually finite group.

- $(i) \Rightarrow (iii)$: Suppose $\bigcap\limits_{\substack{N \trianglelefteq G \\ |G/N| < \infty}} N \neq \{1_G\}$. That means that there is a $g \in G \setminus \{1_G\}$ such that $g \in N$ for any normal subgroup $N \trianglelefteq G$ that has finite index in G. So for this particular $g \in G$, we can not find a normal subgroup, in which g is not contained. But that means that G can not be residually finite.

- $(i) \Leftarrow (iii)$: Let $\bigcap\limits_{\substack{N \trianglelefteq G \\ |G/|<\infty}} N = \{1_G\}$. That means that we cannot find a $g \in G \setminus \{1_G\}$ that is contained in all finite index normal subgroups of G. So for every $g \in G \setminus \{1_G\}$ we can find a finite index normal subgroup $N_g \trianglelefteq G$ such that $g \notin N_g$. Therefore G is residually finite.

- $(i) \Leftrightarrow (iv)$ and $(i) \Leftrightarrow (v)$: Those equivalences are immediate from *Poincaré's theorem*: Any finite index subgroup $K \leq G$ contains a subgroup $N \leq K$ that is normal and of finite index in G. So those cases reduce to the equivalence of $(i) \Leftrightarrow (ii)$ and $(i) \Leftrightarrow (iii)$.

□

Remark 1.14. From now on we will use whichever characterization of residual finiteness suits our problem best without any further comment.

Example 1.10. The direct product of finitely many residually finite groups is residually finite.

Proof. Let G and H be two residually finite groups. Then for every element $g \in G \setminus \{1_G\}$ there is a finite group F_g and a homomorphism $\phi_g : G \to F_g$ such that $\phi_g(g) \neq 1_{F_g}$ and for every element $h \in G \setminus \{1_H\}$ there is a finite group F_h and a homomorphism $\phi_h : G \to F_h$ such that $\phi_h(h) \neq 1_{F_h}$. Furthermore define the trivial groups $F_{1_G} = \{1_G\}$ and $F_{1_H} = \{1_H\}$ and the trivial homomorphisms $\phi_{1_G} : G \to F_{1_G}$ and $\phi_{1_H} : H \to F_{1_H}$ respectively. Let now $K = G \times H$ define the direct product of G and H and let $F_{x,y} = F_x \times F_y$ for $x \in G$ and $y \in H$ define the direct product of F_x and F_y. Furthermore we can define the homomorphism

$$\begin{aligned} \phi_{x,y} : K &\to F_{x,y} \\ (x, y) &\mapsto (\phi_x(x), \phi_y(y)) . \end{aligned}$$

Let now $(x, y) \in K \setminus \{(1_G, 1_H)\}$ be a non-trivial element in K. This element is either of the form (g, h), $(g, 1_H)$ or $(1_G, h)$. Either way, the image $\phi_{x,y}((x, y))$ is non-trivial, since:

$$\begin{gathered} (\phi_g(g), \phi_h(h)) \neq (1_{F_g}, 1_{F_h}) = 1_{F_{g,h}} \\ (\phi_g(g), \phi_{1_H}(1_H)) = (\phi_g(g), 1_H) \neq (1_{F_g}, 1_{F_{1_H}}) = 1_{F_{g,1_H}} \\ (\phi_{1_G}(1_G), \phi_h(h),) = (1_G, \phi_h(h)) \neq (1_{F_{1_G}}, 1_{F_h}) = 1_{F_{1_G,h}} . \end{gathered}$$

As the finite direct product of finite groups, $F_{g,h}$ has to be finite as well. Hence, K is residually finite. Without any further ado, we can extend this argument to any finite collection of residually finite groups. $\square$

Also, we can widen the definition of residual finiteness to any finite number of group elements:

Proposition 1.16. *Let G be a residual finite group and let $g_1, g_2, \dots, g_n \in G$ for some $n \in \mathbb{N}$. Then there is a subgroup $K \leq G$ such that $[G : K] < \infty$ and $g_i \notin K$ for any $i \in \{1, 2, \dots, n\}$.*

Proof. Let G be residually finite. Then for every $g_i \in G$ with $i \in \{1, 2, \dots, n\}$ there is a normal subgroup $K_i \trianglelefteq G$ such that $g_i \notin K_i$ and $|G/K_i| < \infty$ for every $i \in \{1, 2, \dots, n\}$. Now define $K \coloneqq \bigcap\limits_{i=1}^{n} K_i$. Then clearly $g_i \notin K$ for any $i \in \{1, 2, \dots, n\}$ and as the finite intersection of finite index normal subgroups, K is a finite index normal subgroup as well. $\square$

Under special circumstances, residual finiteness can be passed on to sub- and supergroups:

Proposition 1.17. *Let G be a finitely generated group. Then:*

(i). If G is residually finite and $H \leq G$ is a subgroup, then H is residually finite as well.

(ii). If $H \leq G$ is a residually finite subgroup that has finite index $[G : H] < \infty$, then G is residually finite as well.

Proof. Let G be finitely generated.

(i). Let G be residually finite and let $H \leq G$ be a subgroup. Choose $h \in H \leq G$. Since G is residually finite, there is a homomorphism $\phi : G \to F$ into some finite group F such that $\phi(h) \neq 1_F$. Now restrict ϕ to H to get a homomorphism $\phi|_H : H \to F$ with $\phi|_H(h) \neq 1_F$. Therefore, H is residually finite as well.

(ii). Let $g \in G$ and let H be residually finite and of finite index in G. There are two cases:
Case 1: $g \notin H$. Then we can simply take H itself as the finite index subgroup not containing g.
Case 2: $g \in H$. Since H is residually finite, there is a subgroup $K \leq H$ with finite index $[H : K] < \infty$ such that $g \notin K$. But of course K is a subgroup of G as well and $[G : K] = [G : H] \cdot [H : K] < \infty$, hence K is the subgroup with the desired property.

□

Example 1.11. The special linear group $SL_n(\mathbb{Z})$ is residually finite for any $n \in \mathbb{N}$.

Proof. Consider the congruence groups $SL_n(\mathbb{Z})[m]$ which are defined to be the kernel of the map $\phi : SL_n(\mathbb{Z}) \to SL_n(\mathbb{Z}/m\mathbb{Z})$. Since $SL_n(\mathbb{Z})[m] = \{A = (a_{i,j})_{i,j=1}^n \in SL_n(\mathbb{Z}) \mid a_{i,i} \equiv 1 \pmod m; \quad a_{i,j} \equiv 0 \pmod m \text{ if } i \neq j\}$, we see that by multiplying elements from $SL_n(\mathbb{Z})$ to it we can at most obtain m^{n^2} different elements, therefore $[SL_n(\mathbb{Z}) : SL_n(\mathbb{Z})[m]] < \infty$, so $SL_n(\mathbb{Z})[m]$ is a finite index normal subgroup for any $m \geq 3$. Now we consider $A \in I := \bigcap_{m \geq 3} SL_n(\mathbb{Z})[m]$. Then all diagonal elements of A are equivalent to 1 modulo m for any $m \geq 3$, so therefore they can only be precisely 1. Any off-diagonal element of A is equivalent to 0 modulo m for any $m \geq 3$, so they are all precisely 0. So $A = \mathbb{I} = 1_{SL_n(\mathbb{Z})}$, So by Proposition 1.15 (iv), $SL_n(\mathbb{Z})$ is residually finite. □

We can also find a topological interpretation for residual finiteness:

Remark 1.15. The basic concepts and definitions in algebraic topology that might be used forthcoming can be found in Hatcher's standard textbook [15] and Kriegl's lecture notes [27].

Proposition 1.18. *Let G be a finitely generated group. Let M be a compact manifold such that $\pi_1(M) = G$ and let $\tilde{M}$ be its universal covering. Then G is residually finite if and only if for every compact subset $C \subseteq \tilde{M}$ there is a finite sheeted covering M_C of M such that C embedds homeomorphically into M_C, i.e. there exists an homeomorphism $C \to M_C$.*

Proof. First let G be residually finite. Choose a compact subset $C \subseteq \tilde{M}$. Since $\pi_1(M) = G$, we know G acts freely and properly discontinuously on $\tilde{M}$, so for every compact $K \subseteq \tilde{M}$ the set $\{g \in G \mid gK \cap K \neq \emptyset\}$ is finite. Let us denote this set without $\{1_G\}$ as $T_K := \{g \in G \setminus \{1_G\} \mid gK \cap K \neq \emptyset\}$, i.e. for $K = C$ we consider T_C. Now choose $H_C \leq G$ such that $|G/H_C| < \infty$ and $T_C \cap H_C = \emptyset$. Note that H_C exists by Proposition 1.16, since G is residually finite. Now let M_C be the finite sheeted covering corresponding to H_C, that is $\pi_1(M_C) = H_C$. Then $hC \cap C = \emptyset$ for every $h \in \pi_1(M_C)$. Therefore the covering map $\tilde{M} \to M$ restricts to a homeomorphism on C.
Conversely assume that the topological condition holds. Choose $g \in G \setminus \{1_G\}$, so g corresponds to a loop in M not homotopic to 0, since $G = \pi_1(M)$. Let $p : \tilde{M} \to M$ be the universal covering. Since $\tilde{M}$ is simply connected, $p^{-1}(g) =: a$ is not a loop, so it has two distinct endpoints, say x and y. Now define $C := \{x, y\}$. Clearly C is compact, so there is a finite sheeted cover M_C, in which x and y are distinct points. Now define $\pi_1(M_C) =: H_C$. By the lifting property there is a homomorphism $H_C = \pi_1(M_C) \to \pi_1(M) = G$ with $|G/H_C| < \infty$. But since

$x \neq y$ and g is not a loop, $g \notin H_C$. So we have found a finite index subgroup of G, that does not contain g, so G is residually finite. $\square$

Remark 1.16. In fact, in the proof of Proposition 1.18 we never used any property of the manifold M, except that its fundamental group G acts properly discontinuously and freely on its universal covering. Also, we can loose the condition that the action needs to be free, as that only adds a finite number of elements to T_C. So instead of assuming that M is a manifold it suffices to require M to be Hausdorff and locally compact.

Example 1.12. $\mathbb{Z}$ is residually finite.

Proof. We know that $\mathbb{Z} = \pi_1(S^1)$ where S is the unit circle. Now consider a loop α, m times wrapping around S^1. Take the standard $m+1$-sheeted cover of S^1. Then α lifts to a path that is not closed. So we can again define the compact subset C as the endpoints of α and continue as above in the proof of Proposition 1.18. $\square$

Example 1.13. Finitely generated abelian groups are residually finite.

Proof. Let G be a finitely generated abelian group. By the Structure Theorem we know

$$G \cong \mathbb{Z}^n \oplus \mathbb{Z}_{p_1} \oplus \cdots \oplus \mathbb{Z}_{p_n} .$$

But by Example 1.8 the finite group $\mathbb{Z}_{p_i}$ is residually finite for $i \in \{1, 2, \ldots, n\}$ and $\mathbb{Z}$ is residually finite by Example 1.12. As the direct product of finitely many residually finite groups, G is residually finite by Example 1.10 as well. $\square$

1.3 Hopfian Groups

Definition 1.24. A group G is called *Hopfian* if either one of the following conditions hold:

(i). G is not isomorphic to any of its proper quotients.

(ii). Any epimorphism $\pi : G \to G$ is an isomorphism.

In fact, we only need one of the above conditions:

Lemma 1.19. *The statements (i) and (ii) are equivalent.*

Proof. First consider a short exact sequence

$$1 \to A_1 \overset{\phi_1}{\to} A_2 \overset{\phi_2}{\to} A_3 \to 1$$

where ϕ_1 is injective, ϕ_2 is surjective and $ker(\phi_2) = Im(\phi_1)$. By the Isomorphism Theorem we obtain

$$\begin{aligned}
& Im(\phi_2) \cong A_2/ker(\phi_2) \\
\overset{ker(\phi_2)=Im(\phi_1)}{\Rightarrow} \quad & Im(\phi_2) \cong A_2/Im(\phi_1) \\
\overset{\phi_1 \text{ injective}}{\Rightarrow} \quad & Im(\phi_2) \cong A_2/A_1 \\
\overset{\phi_2 \text{ surjective}}{\Rightarrow} \quad & A_3 \cong A_2/A_1
\end{aligned}$$

Now take $A_1 := K \trianglelefteq G$, $A_2 = A_3 := G$, $\phi_1 = Id_K$ and $\phi_2 = \pi$.
Then we get

$$1 \to K \to G \overset{\pi}{\to} G \to 1 \quad \text{and} \quad G \cong G/K\,.$$

Suppose now that (i) holds, i.e. $G \ncong G/K$, unless $K = \{1_G\}$, so we get the sequence

$$1 \to 1 \to G \overset{\pi}{\to} G \to 1$$

which means that π is bijective, so (ii) holds.
Conversely assume (ii) is true, that means π is bijective. Then $K = \{1_G\}$ and so we cannot find a proper quotient isomorphic to G, i.e. (i) holds. □

Remark 1.17. Since it is hard to check whether every epimorphism is an isomorphism, it is easier to determine that a group G is not Hopfian: We just need to find one epimorphism $\pi : G \to G$ such that $ker(\pi) \neq \{1_G\}$.

Example 1.14. $G = \mathbb{Z}^\infty$ is not Hopfian.

Proof. Consider the map

$$\begin{array}{cccc}
\pi: & G & \to & G \\
 & (n_1, n_2, \dots) & \mapsto & (n_2, n_3, \dots)
\end{array}\,.$$

Clearly, this π is an epimorphism, but $(1, 0, 0, \dots) \in ker(\pi)$ and $(1, 0, 0, \dots) \neq (0, 0, 0, \dots) = 1_{\mathbb{Z}^\infty}$. □

Theorem 1.20 (Malcev)**.** *Let G be a finitely generated, residually finite group. Then G is Hopfian.*

Proof. Let G be finitely generated and residually finite and let $\phi : G \to G$ be an epimorphism with kernel N. Then $N \trianglelefteq G$ is a normal subgroup and $G \cong G/N$. We need to show that $N = \{1_G\}$.
Let $m \in \mathbb{N}$, then we know, since G is finitely generated, by Marshal Hall's theorem, there is only a finite number $r_m \in \mathbb{N}$ of subgroups of index m in G.

Denote them as $\{H_1, \dots, H_{r_m}\}$. Now define $K_i := \phi^{-1}(H_i)$ as the preimage for all $i \in \{1, \dots, r_m\}$. As K_i is not the preimage of the trivial group for any $i \in \{1, \dots, r_m\}$, all K_i contain the kernel N. But since all K_i are also distinct subgroups of G of index m, the set $\{H_i \mid i \in \{1, \dots, r_m\}\}$ is equal to the set $\{K_i \mid i \in \{1, \dots, r_m\}\}$. Hence $N \trianglelefteq H_i$ for all $i \in \{1, \dots, r_m\}$. But m was arbitrarily chosen, thus N has to be contained in any finite index subgroup, i.e.

$$N \leq \bigcap_{\substack{K \leq G: \\ [G:K]<\infty}} K \,.$$

But this intersection is trivial by Proposition 1.15 (iv) as G is residually finite, hence $N = \{1_G\}$ and G is Hopfian. $\square$

Example 1.15. Since any finite group is residually finite by Example 1.8, it is also Hopfian.

Remark 1.18. Malcev's Theorem 1.20 gives us a new tool to prove that a group is not residually finite, by showing it is not Hopfian, i.e. we can find a non-injective epimorphism.

2 Tools to Determine Residual Finiteness

It is obvious that it can become very cumbersome to determine whether a group is residually finite with the original definition. In particular when the group is infinite, we cannot run through all elements to check whether for all of them there is a finite index normal subgroup not containing the element. Although Proposition 1.15 gives some equivalent definitions, the question does not become much more accessible. It remains unclear, how to systematically find finite index subgroups. Hence, we need to develop some machinery that might help to check groups for residual finiteness without simply applying the definition.

2.1 Tietze transformations

Definition 2.1. Let $G = \langle \mathcal{A} \mid \mathcal{R} \rangle = \langle a, b, c, \dots \mid P, Q, R, \dots \rangle$ be a presentation of the group G, where $\mathcal{A}$ is the alphabet and $\mathcal{R}$ is the set of relations.

- For any relation $W \in \mathcal{R}$ of the form $W_1(\mathcal{A}) = W_2(\mathcal{A})$, where $W_1(\mathcal{A})$ and $W_2(\mathcal{A})$ are both words in $\mathcal{A} \cup \mathcal{A}^{-1}$ the word $W_1(\mathcal{A})W_2(\mathcal{A})^{-1}$ is called *relator*.
- For any element $A \in \mathcal{A}$, the word AA^{-1} is called *trivial relator*.

- Let W be a word in G. Then W is *derivable* from $\mathcal{R} = \{P, Q, R, \dots\}$ if the following operations, applied a finite number of times, change W into the empty word:
  (i). Insert one of the relators corresponding to the relations $\{P, Q, R, \dots\}$ or its inverse between any two consecutive symbols of W, before W or after W.
  (ii). Insert any trivial relator between any two consecutive symbols of W, before W or after W.
  (iii). Delete one of the relators corresponding to the relations $\{P, Q, R, \dots\}$ or its inverse forming a block of consecutive symbols in W.
  (iv). Delete any trivial relator forming a block of consecutive symbols in W.

Definition 2.2. Let $G = \langle a, b, c, \cdots \mid P, Q, R, \dots \rangle$ be a group with alphabet $\mathcal{A} = \{a, b, c, \dots\}$ and relations $\mathcal{R} = \{P, Q, R, \dots\}$. The following four transformations to the presentation of G are called *Tietze transformations*:

T_1 Let $\mathcal{W}_1 = \{S, T, \dots\} \subseteq G$ be a set of words in A. If $\mathcal{W}_1$ is derivable from elements in $\mathcal{R}$, add $\{S = 1, T = 1, \dots\}$ to $\mathcal{R}$.

T_2 Let $\mathcal{R}_2 = \{S = 1, T = 1, \dots\} \subseteq \mathcal{R}$ be some relations of the group presentation. If $\{S, T, \dots\}$ is derivable from $\mathcal{R} \setminus \mathcal{R}_2$, delete $\mathcal{R}_2$ from $\mathcal{R}$.

T_3 Let $\mathcal{W}_3 = \{K, M, \dots\} \subseteq G$ be words in $\mathcal{A}$. Adjoin $A_3 = \{x, y, \dots\}$ to $\mathcal{A}$ and $R_3 = \{K = x, M = y, \dots\}$ to $\mathcal{R}$.

T_4 Let $\mathcal{A}_4 = \{V, W, \dots\} \subseteq \mathcal{A}$ be some elements of the alphabet. If there is a subset $\mathcal{A}_4' = \{p, q, \dots\} \subseteq \mathcal{A}$ of A such that $\mathcal{A}_4 \cap \mathcal{A}_4' = \emptyset$ and $\mathcal{R}_4 = \{p = V, q = W, \dots\} \subseteq \mathcal{R}$ is a subset, then delete $\mathcal{A}_4'$ from $\mathcal{A}$ and $\mathcal{R}_4$ from $\mathcal{R}$ and replace every appearance of elements from $\mathcal{A}_4'$ in $\mathcal{R}$ by the corresponding element of $\mathcal{A}_4$.

Theorem 2.1. *[30, Chapter 1.5, p. 51 f.] Let G be a group with two presentations*

$$\langle \{a_i \mid i \in I \subseteq \mathbb{N}\} \mid \{R_j(a_i) \mid j \in J \subseteq \mathbb{N}, i \in I\} \rangle$$

and

$$\langle \{b_k \mid i \in K \subseteq \mathbb{N}\} \mid \{S_l(a_k) \mid l \in L \subseteq \mathbb{N}, k \in K\} \rangle .$$

Call the first presentation (A) with alphabet $\mathcal{A}_A$ and relations $\mathcal{R}_A$ and the second presentation (B) with alphabet $\mathcal{A}_B$ and relations $\mathcal{R}_B$. Then the presentation (B) can be obtained from presentation (A) by repeated application of Tietze transformations.

Corollary 2.2. *Let G be a finitely generated group with two group presentations (A) and (B). Then the presentation (B) can be obtained by the presentation (A) by applying a finite number of Tietze transformations.*

The following examples [30, cf.] illustrate how convenient these Tietze transformations can be, as soon as one knows which generators and relations are suitable to be added.

Example 2.1. Let $G = \langle a, b, c \mid (ab)^2ab^2 = 1\rangle$. Then $G \cong \langle c, x\rangle = F_2$.

Proof.

- Use T_3: Add new generators x, y and relations $x = ab, y = ab^2$ to obtain

$$G = \langle a, b, c, x, y \mid (ab)^2ab^2 = 1, x = ab, y = ab^2\rangle .$$

- Use T_1: Add the new relation $x^2y = 1$, since it is derivable from the relations of the current presentation to obtain

$$G = \langle a, b, c, x, y \mid (ab)^2ab^2 = 1, x^2y = 1, x = ab, y = ab^2\rangle .$$

- Use T_2: Delete the relation $(ab)^2ab^2 = 1$ since it is derivable from the other relations of the current presentation to obtain

$$G = \langle a, b, c, x, y \mid x^2y = 1, x = ab, y = ab^2\rangle .$$

- Use T_1: Add the new relations $b = x^{-1}y$ and $a = xy^{-1}x$, since they are derivable from the relations of the current presentation via $y = ab^2 = abb = xb$ and $x = ab = ayx^{-1}$ to obtain

$$G = \langle a, b, c, x, y \mid x^2y = 1, x = ab, y = ab^2, b = x^{-1}y, a = xy^{-1}x\rangle .$$

- Use T_4: Observe that $a = xy^{-1}x =: V \in \mathcal{W}(\{x, y\}) \subseteq \mathcal{W}(\mathcal{A} \setminus \{a, b\})$ and $b = x^{-1}y =: W \in \mathcal{W}(\{x, y\}) \subseteq \mathcal{W}(\mathcal{A} \setminus \{a, b\})$ and therefore delete a, b from $\mathcal{A}$ and we can replace a by V and b by W in all elements of $\mathcal{R}$ to obtain

$$G = \langle c, x, y \mid x^2y = 1, x = (xy^{-1}x)(x^{-1}y), y = (xy^{-1}x)(x^{-1}y)^2\rangle .$$

- Use T_2: Since $x = xy^{-1}xx^{-1}y = x$ and $y = xy^{-1}xx^{-1}yx^{-1}y = y$, the two latter relations of our current presentations are derivable and can be therefore deleted, so we obtain

$$G = \langle c, x, y \mid x^2y = 1\rangle .$$

- Use T_1: The relation $x^{-2} = y$ can be derived from $x^2y = 1$ and therefore added so we obtain,

$$G = \langle c, x, y \mid x^2y = 1, x^{-2} = y\rangle .$$

- Use T_2: The relation $x^2y = 1$ can be derived from $x^{-2} = y$ and therefore deleted to obtain

$$G = \langle c, x, y \mid x^{-2} = y\rangle .$$

- Use T_4: Observe $y = x^{-2} =: Z \in \mathcal{W}(\{x\}) \subseteq \mathcal{W}(\mathcal{A} \setminus \{y\})$, so we can delete y from $\mathcal{A}$ and the relation $y = X$ from $\mathcal{R}$ and replace every occurrence of y in $\mathcal{R}$ by X. But since $\mathcal{R}$ does not have any other element, the set $\mathcal{R}$ is empty, so we obtain

$$G = \langle c, x\rangle .$$

□

Example 2.2. Let $G = \langle x, y \mid x^3 = 1, y^2 = 1, (xy)^2 = 1\rangle$. Then

$$G = \langle y, z \mid z^2 = 1, y^2 = 1, (zy)^3 = 1\rangle .$$

Proof.

- Use T_3: Add the new generator z and the relation $z = xy = xy^{-1}$ to obtain

$$G = \langle x, y, z \mid x^3 = 1, y^2 = 1, (xy)^2 = 1, z = xy\rangle .$$

- Use T_1 and T_2: Add the derivable relation $x = zy^{-1} = zy$ and delete the derivable relation $z = xy$ to obtain

$$G = \langle x, y, z \mid x^3 = 1, y^2 = 1, (xy)^2 = 1, zy = x\rangle .$$

- Use T_4: Observe $x = zy =: V \in \mathcal{W}(\{y, z\}) \subseteq \mathcal{W}(\mathcal{A} \setminus \{x\})$, hence we can remove x from $\mathcal{A}$ and $x = V$ from $\mathcal{R}$ and replace every occurrence of x in an element of $\mathcal{R}$ by V to obtain

$$\langle y, z \mid y^2 = 1, z^2 = 1, (zy)^3 = 1\rangle .$$

□

Example 2.3. Let $G = \langle a, b, c \mid b^2 = 1, (bc)^2 = 1\rangle$. Then

$$G = \langle x, y, z \mid y^2 = 1, z^2 = 1\rangle .$$

Proof.

- Use T_3: Add the new generators x, y, z with relations $x = a$, $y = b$, $z = bc$ to obtain

$$G = \langle a, b, c, x, y, z \mid b^2 = 1, (bc)^2 = 1, x = a, y = b, z = bc \rangle .$$

- Use T_4: Observe that $a = x =: V \in \mathcal{W}(\{x, y\}) \subseteq \mathcal{W}(\mathcal{A} \setminus \{x, y\})$ and $b = y =: W \in \mathcal{W}(\{x, y\}) \subseteq \mathcal{W}(\mathcal{A} \setminus \{x, y\})$. So we can remove a, b from $\mathcal{A}$ and $a = V$, $b = W$ from $\mathcal{R}$ and replace every occurrence of a and b in elements of $\mathcal{R}$ by V and W respectively to obtain

$$G = \langle c, x, y, z \mid y^2 = 1, (yc)^2 = 1, z = yc \rangle .$$

- Use T_4: Observe that $c = y^{-1}z =: X \in \mathcal{W}(\{y, z\}) \subseteq \mathcal{W}(\mathcal{A} \setminus \{c\})$. So we can remove c from $\mathcal{A}$ and $c = X$ from $\mathcal{R}$ and replace every occurrence of c in elements of $\mathcal{R}$ by X. Since $(yy^{-1}z)^2 = z^2$ we obtain

$$G = \langle x, y, z \mid y^2 = 1, z^2 = 1 \rangle .$$

$\square$

Example 2.4. Let $m, n \in \mathbb{Z}$ such that $gcd(m, n) = 1$ and let $G = \langle a \mid a^{mn} = 1 \rangle$. Then

$$G = \langle b, c \mid b^n = 1, c^m = 1, bc = cb \rangle .$$

Proof.

- Use T_3 to obtain

$$G = \langle a, b, c \mid a^{mn} = 1, b = a^n, c = a^m, bc = cb \rangle .$$

- Use T_1 to obtain

$$G = \langle a, b, c \mid a^{mn} = 1, b = a^m, c = a^n, b^n = 1, c^m = 1, bc = cb \rangle .$$

- Use T_1: Observe that if $gcd(m, n) = 1$ then there are $r, s \in \mathbb{Z}$ such that $1 = rm + sn$, hence $a = a^1 = a^{rm+sn} = (a^m)^r (a^n)^s = b^r c^s$. So we obtain

$$G = \langle a, b, c \mid a^{mn} = 1, b = a^m, c = a^n, b^n = 1, c^m = 1, bc = cb, a = b^r c^s \rangle .$$

- Use T_4 to obtain

$$G = \langle b, c \mid b^n = 1, c^m = 1, bc = cb, (b^r c^s)^{mn} = 1, b = (b^r c^s)^m, c = (b^r c^s)^n \rangle \,.$$

- Use T_2: Observe that we can derive

$$\begin{gathered}(b^r c^s)^{mn} = (b^n)^{rm}(c^m)^{sn} = 1^{rm}1^{sn} = 1 \\ (b^r c^s)^m b^{-1} = b^{rm}c^{sm}b^{-1} = b^{rm-1}c^{sm} = b^{sn}c^{sm} = (b^n)^s(c^m)^s = 1^s 1^s = 1 \\ (b^r c^s)^n c^{-1} = b^m c^{sn-1} = b^{rn}c^{rm} = (b^n)^r(c^m)^r = 1^r 1^r = 1 \,,\end{gathered}$$

so we obtain

$$G = \langle b, c \mid b^n = 1, c^m = 1, bc = cb \rangle \,.$$

□

Example 2.5. Let $G = \langle a, b \mid aba = bab \rangle$. Then

$$G = \langle c, d \mid c^3 = d^2 \rangle \,.$$

Proof.

- Use T_3 to obtain

$$G = \langle a, b, c, d \mid aba = bab, c = ab, d = aba \rangle \,.$$

- Use T_1 and T_2: Derive

$$\begin{aligned} a = cb^{-1} \quad &\Rightarrow \quad d = cb^{-1}bcb^{-1} = c^2b^{-1} \\ &\Rightarrow \quad b = d^{-1}c^2 \\ &\Rightarrow \quad a = c(d^{-1}c^2)^{-1} = c^{-1}d \end{aligned}$$

to obtain

$$G = \langle a, b, c, d \mid aba = bab, a = c^{-1}d, b = d^{-1}c^2 \rangle \,.$$

- Use T_4: Replace a by $c^{-1}d$ and b by $d^{-1}c^2$ and derive

$$c^{-1}dd^{-1}c^2c^{-1}d = d^{-1}c^2c^{-1}dd^{-1}c^2 \quad \Rightarrow \quad d = d^{-1}c^3 \quad \Rightarrow \quad d^2 = c^3$$

to obtain

$$G = \langle c, d \mid d^2 = c^3 \rangle \,.$$

□

2.2 Theorem of Malcev

Theorem 2.3 (Malcev)**.** *Let G be a finitely generated, linear group, i.e. $G \subseteq GL_n(\mathbb{C})$ for some $n \in \mathbb{N}$. Then G is residually finite.*

Proof. [48] We will distinguish two cases:

- First let $G \leq GL_n(\mathbb{Z})$ be finitely generated. Since $GL_n(\mathbb{Z})$ itself is finitely generated, it suffices to show that $GL_n(\mathbb{Z})$ is residually finite, since the then by Proposition 1.17 any subgroup will be residually finite as well.
  So choose $A \in GL_n(\mathbb{Z}) \setminus \{\mathbb{I}\}$. That means that $A - \mathbb{I}$ has at least one entry $x \in \mathbb{Z} \setminus \{0\}$. By the Fundamental Theorem of Arithmetic we know there exists a prime $p \in \mathbb{P}$ such that $p \nmid x$.
  Now consider the homomorphism

$$\begin{aligned}\phi_p : GL_n(\mathbb{Z}) &\to GL_n(\mathbb{Z}/p\mathbb{Z})\\ M &\mapsto M \quad (mod\, p)\end{aligned}$$

  Since $|\mathbb{Z}/p\mathbb{Z}| < \infty$ implies $|GL_n(\mathbb{Z}/p\mathbb{Z})| < \infty$, we know by the Isomorphism Theorem that $|GL_n(\mathbb{Z})/ker(\phi_p)| < \infty$. But $\phi_p(A) = A\,(mod\, p) \neq \mathbb{I}$ since $x \neq 0\,(mod\, p)$, so $A \notin ker(\phi_p)$. Therefore $GL_n(\mathbb{Z})$ is residually finite.

- Now let $G \leq GL_n(\mathbb{C})$ be finitely generated. The finite collection of entries of generators of G, (finitely) generate a ring $R \subseteq \mathbb{C}$. In particular, R cannot have any zero-divisors, since $\mathbb{C}$ is a field, so R is an integral domain. Again, since with R also $GL_n(R)$ is finitely generated, it suffices to show that $GL_n(R)$ is residually finite by Proposition 1.17.
  So choose $A \in GL_n(R) \setminus \{\mathbb{I}\}$. Then $A - \mathbb{I}$ has an entry $x \in R \setminus \{0\}$. So there exists a prime ideal $P \subseteq R$ such that $x \notin P$. But since R is finitely generated, P has to be finitely generated and therefore R/P is finitely generated. Also, since $\mathbb{C}$ is abelian, so are R and P and therefore R/P is abelian. So R/P is a finite integral domain and therefore a finite field.
  Now consider, analogously to the first case, the homomorphism

$$\begin{aligned}\phi_P : GL_n(R) &\to GL_n(R/P)\\ M &\mapsto M \quad (mod\, P)\,.\end{aligned}$$

  Since $|R/P| < \infty$ implies $|GL_n(R/P)| < \infty$, we know by the Isomorphism Theorem that $|GL_n(R)/ker(\phi_P)| < \infty$. But $\phi_P(A) = A\,(mod\, P) \neq \mathbb{I}$ since $x \notin P$, so $A \notin ker(\phi_P)$. Therefore $GL_n(R)$ and with it the subgroup $G \leq GL_n(R)$ is residually finite.

$\square$

Remark 2.1. Malcev's Theorem 2.3 gives us a second proof for Example 1.11 that $SL_n(\mathbb{Z})$ is residually finite.

The converse of Malcev's Theorem 2.3 is not true:

Example 2.6 (Drutu-Sapir). The group $\langle a, t \mid a^{t^2} = a^2 \rangle$ is residually finite but not linear.

2.3 Bass-Serre Theory

The following results are adopted from Jean-Pierre's Serre standard reference *Trees* [43]. To fully understand his work, it is helpful to use details, reformulations and examples introduced in [5] and [13].

Remark 2.2. An introduction on the necessary graph theoretical concepts and definitions appearing in this chapter can be found Chapter 1 of [5] and in Chapter 2 of [43].

Nonetheless let us recall the most important definitions:

Definition 2.3. Let $X = (V(X), E(X))$ be a directed graph.

- The *inverse* of an edge $e = (x, y) \in E(X)$ denoted by $\bar{e}$ is the edge $(y, x) \in E(X)$ connecting the same vertices in the opposite direction.
- For an edge $e = (x, y) \in E(X)$, x is called the *origin* of e, denoted by $o(e)$ and y is called the *terminal vertex* of e, denoted by $t(e)$.
- If $A \subseteq E(X)$ is a subset of the edge set, then A is called an *orientation of* X if $|A \cap \{e, \bar{e}\}| = 1$ for every $e \in E(X)$.

Definition 2.4. Let X and Y be two graphs. A map $f : X \to Y$ is called *graph morphism* if

$$f(V(X)) \subseteq V(Y) \qquad \text{and} \qquad f(E(X)) \subseteq E(Y)$$

and the following three conditions hold for any $e \in E(X)$:

(i). $f(o(e)) = o(f(e))$

(ii). $f(t(e)) = t(f(e))$

(iii). $f(\bar{e}) = \overline{f(e)}$.

Definition 2.5. Let G be a group and X be a directed graph.

(i). We say G *acts on* X if there are maps

$$\begin{aligned} G \times V(X) &\to V(X) \qquad & G \times E(X) &\to E(X) \\ (g, v) &\mapsto gv \qquad & (g, e) &\mapsto ge \end{aligned}$$

such that $go(e) = o(ge)$ and $g\bar{e} = \overline{ge}$ for any $e \in E(X)$.

(ii). The *orbit* of the vertex $v \in V(X)$ under the action of G is the set $G \cdot v = \{gv \mid g \in G\}$ and the *orbit* of the edge $e \in E(X)$ under the action of G is the set $G \cdot e = \{ge \mid g \in G\}$.

(iii). The *factor graph* or *quotient graph* ${}_G\backslash^X$ is the graph given by the vertex set $V({}_G\backslash^X) = \{G \cdot v \mid v \in V(X)\}$ and the edge set $E({}_G\backslash^X)$ given by all the orbits $G \cdot e$ with $e \in E(X)$ such that

- $G \cdot v = o(G \cdot e)$ if there is a $g \in G$ such that $gv = o(e)$,
- $\overline{G \cdot e} = G \cdot \bar{e}$.

Definition 2.6. Let G be a group acting on a directed graph X. A pair $(g, e) \in G \times E(X)$ is called an *inversion* if $ge = \bar{e}$, i.e. g reverses the orientation of the edge e. If there is no such pair, we say G acts on X *without inversion.*

Proposition 2.4. *Let G be a group acting on the connected, directed graph X without inversion. Then every subtree T' of ${}_G\backslash^X$ lifts to a subtree of X.*

Proof. Define the set $\Omega := \{T \subseteq X \text{ subtree} \mid T \text{ projects injectively into } T'\}$, that is partially ordered by inclusion and any ascending chain of elements in Ω have an upper bound given by their union. Hence, by Zorn's Lemma there is a maximal element T_0 of Ω. Define the image of T_0 in T' under the projection as T_0'. If we suppose that $T_0' \neq T'$, then there is an edge $e' \in E(T') \setminus E(T_0')$. Since T as a tree is connected, we can assume that $o(e') \in V(T_0')$ and $t(e') \notin V(T_0')$. Now let e' be lifted to e. Since we could replace the lift by ge for any $g \in G$, without loss of generality we can assume that $o(e) \in V(T_0)$. Now define T_1 as the graph obtained by adjoining vertex $t(e)$ and edges $e, \bar{e}$ to T_0. But since $t(e)$ is a terminal vertex, T_1 is still a tree, that projects injectively into T', in contradiction to T_0 being the maximal tree with that property. So $T_0' = T'$ and T' lifts to a subtree of a X. $\square$

Definition 2.7. Let G be a group and let it act on a graph X. A lift $T \subseteq X$ of a maximal subtree in ${}_G\backslash^X$ is called *tree of representatives of X mod G.*

Definition 2.8. Let φ_1 and φ_2 be two actions of G on the set Ω_1 and Ω_2 respectively. We call φ_1 and φ_2 *equivalent* if there is a bijective map $t : \Omega_1 \to \Omega_2$ such that

$$t(\varphi_1(\alpha, g)) = \varphi_2(t(\alpha), g)$$

for all $\alpha \in \Omega_1$ and for all $g \in G$.

2.3.1 Groups acting freely on a Tree

In case the graph is a tree, we need to refine the definition of a group action on it:

Definition 2.9. Let G be a group and let $T = T(V(T), E(T))$ be a tree. We say G *acts on* T *(by isometry)* if there are maps

$$\begin{aligned} G \times V(T) &\to V(T) \qquad\qquad & G \times E(T) &\to E(T) \\ (g, v) &\mapsto gv & (g, e) &\mapsto ge \end{aligned}$$

such that for any $e = (v_1, v_2) \in E(T)$ we get $ge = (gv_1, gv_2)$.

Definition 2.10. Let G be a group acting on some graph $X = X(V(X), E(X))$. We say that the action is *minimal* if there is no proper subset of $Y \subset X$, left invariant under the action of G, i.e.

$$G \cdot Y \neq Y \qquad \text{for any } Y \neq \emptyset\,,\, Y \subset X$$

Definition 2.11. Let G be a group acting on some (directed) graph $X = X(V(X), E(X))$. We say that the action is *free* if

- the action is without inversions, i.e.

$$ge = g(v_1, v_2) \neq (v_2, v_1) = \bar{e} \quad \text{ for any } e = (v_1, v_2) \in E(X)\,;$$

- the action does not fix any vertex, i.e.

$$gv \neq v \quad \text{ for any } v \in V(T)\,.$$

Remark 2.3. Unless stated otherwise, if a group G acts on a graph X, we assume from now on that the graph is directed and the action is always without inversion, independent of the property of being a free action. This is not a loss of generality, because we can use the following definition:

Definition 2.12. Let $X = X(V(X), E(X))$ be a graph. Its *barycentric subdivision* $B(X)$ is the graph that is obtained by dividing every edge $e \in E(X)$ exactly in the middle by a new vertex $v_e \in V(B(X))$. So for every edge $e = (x, y) \in E(X)$ there are two edges $e_1 = (x, v_e) \in E(B(X))$ and $e_2 = (v_e, y) \in E(B(X))$. Furthermore, we impose that

$$(\bar{e})_2 = \bar{e_1}, \quad (\bar{e})_1 = \bar{e_2} \qquad \text{and} \qquad v_e = v_{\bar{e}}\,.$$

If there is a group G acting on X we can extend the action of G on $B(X)$ by defining

$$ge_1 = (ge)_1\,, \quad ge_2 = (ge)_2 \qquad \text{and} \qquad gv_e = v_{ge}$$

for every $g \in G$, $e_1, e_2 \in E(B(X))$ and for every $v_e \in V(B(X))$ with $e \in E(X)$.

It is apparent that G always acts on $B(X)$ without inversion, so if necessary we can consider $B(X)$ instead of X.

Remark 2.4. Recall Example 1.2: If G is generated by S, its Cayley graph $Cay(G, S)$ is an infinite tree if and only if G is the free group F_S.

Proposition 2.5. *Let G be a group. G is a free group if and only if G acts freely on a tree.*

Proof. Let G be a free group freely generated by the set S. Since any group acts freely on its Cayley graph $Cay(G, S)$, by Remark 2.4, G acts freely on a tree. Conversely let G act freely on a tree T. We need to consider two cases, as proposed in [31]:

(i). First assume the action to be transitive. Let $v \in V(T)$ and define $\mathcal{N}(v) \coloneqq \{x \in V(T) \mid \exists\, e \in E(T) :\ e = (x, v)\}$ to be the set of neighbors of v. Since the action is transitive and free, for every $n \in \mathcal{N}(v)$ there is a unique $g_n \in G$ such that $g_n v = n$. Define the set of all those g_n as $S \coloneqq \{g_n \mid n \in \mathcal{N}(v)\}$. Choose $S' \subseteq S$ such that $S' \cap S'^{-1} = \emptyset$, so it contains exclusively either g_n or g_n^{-1} for every $n \in \mathcal{N}(v)$. Now we want to show that $G = F_{S'}$: Let $g \in G$ and let $x_0 = v, x_1, \ldots, x_m, x_{m+1} = gv \in V(T)$ be the path from v to gv in T. Since the action is transitive, for every x_i with $i \in \{1, 2, \ldots, m\}$ there is a g_i such that $g_i v = x_i$. But since x_i and x_{i+1} are connected for every $i \in \{0, 1, \ldots, m\}$, we conclude that $g_{i+1}^{-1} g_i \in S$. Hence $g \in \langle S \rangle = \langle S' \rangle$, so S is a generating set for G. It is left to show that G is in fact a free group. So take a reduced word in $S' \cup S'^{-1}$. We can associate a path in T to it. If the path would also represent the trivial element 1_G, then T would have a circle, in contradiction to T being a tree. So the group G cannot have any relations and is therefore free.

(ii). Now assume that the action of G on T is not transitive. Let again be $v \in V(T)$. Define its orbit $V'(T') \coloneqq G \cdot v = \{gv \mid g \in G\}$. Two elements $gv, hv \in V'(T')$ are connected in $T' \coloneqq T'(V'(T'), E'(T')) \subseteq T(V(T), E(T))$ if and only if they are connected in $T(V(T), E(T))$ or if the unique geodesic between gv and hv in $T(V(T), E(T))$ does not contain any element of $V'(T')$. By construction, T' cannot have a circle and is connected, so T contains a subtree that is a subdivision of T'. But as G

acts on T, it also acts on T'. This action of G on T' is now transitive. So by the first case, G is a free group.

$\square$

Corollary 2.6 (Nielsen-Schreier)**.** *Every subgroup of a free group is free.*

Proof. Let G be a free group with subgroup $H \leq G$. Then by Proposition 2.5, G acts freely on some tree T. But as a subgroup of G, H acts freely on T as well. So again by Proposition 2.5, $H$is a free group. $\square$

2.3.2 Amalgamated Product

Definition 2.13. Let $G \coloneqq \langle S_G \mid R_G \rangle$ and $H \coloneqq \langle S_H \mid R_H \rangle$ be two groups. Let K be some group and let $\varphi : K \to G$ and $\psi : K \to H$ be two monomorphisms. Let $G * H = \langle S_G \cup S_H \mid R_G \cup R_H \rangle$ be the free product. The *free product with amalgamation* $G *_K H$ of G and H, is the free product adjoined with the relations $\varphi(k)\psi(k)^{-1} = 1_{G*H}$ for any $k \in K$, i.e.

$$G *_K H = \langle S_G \cup S_H \mid R_G \cup R_H \cup \{\varphi(k)\psi(k)^{-1} = 1_{G*H} \mid k \in K\} \rangle .$$

Remark 2.5.

(i). Analogously we can think of the Definition 2.13 of the amalgamated product in the following way: Let N be the smallest normal subgroup of $G * H$ generated by the conjugates of the set $\{\varphi(k)\psi(k)^{-1} \mid k \in K\}$, i.e. the normal closure, then the amalgamated product is given by the quotient $G *_K H = (G * H)/N$.

(ii). Analogous to Definition 2.13 we can define the amalgamated product of more than two groups: Let $(G_i)_{i\in I}$ a family of groups with $I \subseteq \mathbb{N}$ and let $(\varphi_i : K \to G_i)_{i\in I}$ be a family of monomorphism for some group K. Then the *free product with amalgamation* $*_K^{i\in I} G_i$ of the family $(G_i)_{i\in I}$ is defined as the quotient of the free product $*_{i\in I} G_i$ by the normal closure of the set $\{\varphi_i(k)\varphi_{i'}(k)^{-1} \mid k \in K, i, i' \in I\}$.

(iii). In fact, we can simplify Definition 2.13: Since the images $A \coloneqq \varphi(K) \leq G$, and $B \coloneqq \psi(K) \leq H$ of the monomorphisms $\varphi : K \to G$ and $\psi : K \to H$ are subgroups, we can assume that there is an isomorphism $\tilde{\varphi} : A \to B$. To form the amalgamated product, we then only need to adjoin the relations $\tilde{\varphi}(a)a^{-1} = 1_{G*H}$ for any $a \in A$. If this interpretation is used, we denote the amalgamated product as $G *_{A\cong B} H$.

Example 2.7.

(i). If $K = \{1\}$, then the amalgamated product $G *_K H = G *_{\{1\}} H$ is simply the free product $G * H$ of G and H.

(ii). Let $G = \mathbb{Z}/4\mathbb{Z} = \langle a \mid a^4 = 1 \rangle$ and $H = \mathbb{Z}/6\mathbb{Z} = \langle b \mid b^6 = 1 \rangle$ be two cyclic groups. Let $K = \mathbb{Z}/2\mathbb{Z} = \langle c \mid c^2 = 1 \rangle$ be a group with monomorphisms

$$\begin{aligned} \varphi : K &\to G & \qquad\qquad \psi : K &\to H \\ c &\mapsto a^2 & c &\mapsto b^3 \,. \end{aligned}$$

Then the amalgamated product of G and H along K is given by

$$G *_K H = \mathbb{Z}/4\mathbb{Z} *_{\mathbb{Z}/2\mathbb{Z}} \mathbb{Z}/6\mathbb{Z} = \langle a, b \mid a^4 = b^6 = 1, a^2 b^{-3} = 1 \rangle \,.$$

(iii). Let $V_4 = \langle a, b \mid a^2 = b^2 = (ab)^2 = 1 \rangle$ be the Klein four-group with its subgroup $B = \langle b \mid b^2 = 1 \rangle$. We want to derive the amalgamated product of V_4 with itself along the subgroup B. Hence, we can first consider the free product of V_4 with itself, which is

$$\begin{aligned} V_4 * V_4 &= \langle a, b \mid a^2 = b^2 = (ab)^2 = 1 \rangle * \langle c, d \mid c^2 = d^2 = (cd)^2 = 1 \rangle \\ &= \langle a, b, c, d \mid a^2 = b^2 = (ab)^2 = c^2 = d^2 = (cd)^2 \rangle \,. \end{aligned}$$

Since the subgroups B and $D = \langle d \mid d^2 = 1 \rangle$ are isomorphic, we obtain for the amalgamated product

$$V_4 *_{B \cong D} V_4 = \langle a, b, c \mid a^2 = b^2 = c^2 = (ab)^2 = (bc)^2 = 1 \rangle \,.$$

Definition 2.14.

(i). Let G and H be groups with isomorphic subgroups $A \leq G$ and $B \leq H$. Define $F = G *_{A \cong B} H$ to be their amalgamated product along $A \cong B$ and let $i : G * H \to F$ be the canonical homomorphism from their free product onto F. Then every element $f \in F$ can be written as $f = i(x_0)i(x_1) \cdots i(x_n)$ where $x_i \in G \cup H$. To make matters easier we will denote such an element $f \in F$ simply as $f = x_0 x_1 \cdots x_n$.

(ii). Let $A \leq G$ be a subgroup of the group G and $B \leq H$ be a subgroup of the group H. Then T_A denotes the system of representatives of right cosets of A in G and T_B denotes the system of representatives of right cosets of B in H. Assume $1_G \in T_A$ and $1_H \in T_B$. Moreover, denote the unique presentation of an element $g \in G$ as the product of an element of A and an element of T_A by

$$g = \underline{x}\overline{x} \qquad \text{for } \underline{x} \in A \text{ and } \overline{x} \in T_A$$

and analogously for an element $h \in H$

$$h = \underline{y}\overline{y} \qquad \text{for } \underline{y} \in B \text{ and } \overline{y} \in T_B\,.$$

(iii). For an element $f = x_0 x_1 \cdots x_n \in F$ a sequence $(x_0, x_1, \ldots, x_n)$ is called *A-normal form* if

- $x_0 \in A$
- $x_i \in T_A \setminus \{1_G\}$ or $x_i \in T_B \setminus \{1_H\}$ and consecutive elements x_i and x_{i+1} lie in distinct systems of representatives.

Analogously the sequence is called *B-normal form* if

- $x_0 \in B$
- $x_i \in T_A \setminus \{1_G\}$ or $x_i \in T_B \setminus \{1_H\}$ and consecutive elements x_i and x_{i+1} lie in distinct systems of representatives.

Example 2.8. Let $G = \langle a \mid a^{12} = 1\rangle$ and $H = \langle b \mid b^{15} = 1\rangle$ with subgroups $A = \langle a^4 \mid a^{12} = 1\rangle \leq G$ and $B = \langle b^5 \mid b^{15} = 1\rangle \leq H$. Let

$$\begin{aligned} \varphi : A &\xrightarrow{\sim} B \\ a^4 &\mapsto b^5 \end{aligned}$$

be the isomorphism between A and B. Then the amalgamated product of G and H along φ is given by

$$G *_{A\cong B} H = \langle a, b \mid a^{12} = 1, b^{15} = 1, a^4 = b^5\rangle\,.$$

Consider the systems of coset representatives given by

$$T_A = \{1_G, a, a^2, a^3\} \qquad \text{and} \qquad T_B = \{1_H, b, b^2, b^3, b^4\}$$

and let $f \in G *_{A\cong B} H$ be the element $f = a^3ba^5$. Then we can deduce

$$f = a^3ba^5 = a^3ba^4a = a^3bb^5a = a^3b^5ba = a^3a^4ba = a^4a^3ba\,,$$

hence the A-normal form of $f = a^3ba^5$ is given by (a^4, a^3, b, a) since $a^4 \in A$, $a^3, a \in T_A \setminus \{1_G\}$ and $b \in T_B \setminus \{1_H\}$. Furthermore, since $\varphi(a^4) = b^5$, the B-normal form of f is given by (b^5, a^3, b, a).

Theorem 2.7 (Normal Form Theorem for Amalgamated Products)**.** *[5, p. 73, Theorem 11.3] Let G and H be two groups with isomorphic subgroups $A \leq G$ and $B \leq H$. Then any element $f \in G *_{A\cong B} H$ in the amalgamated product can be uniquely written in the form $f = x_0 x_1 \cdots x_n$, where $(x_0, x_1, \ldots, x_n)$ is an A-normal form.*

The Normal Form Theorem 2.7 lets us deduce:

Corollary 2.8. *Let G and H be two groups with isomorphic subgroups $A \leq G$ and $B \leq H$ respectively. If an element $f \in F := G *_{A\cong B} H$ in the amalgamated subgroup is of the form $f = x_0 x_1 \cdots x_n$ with $n \geq 1$ and $x_i \in G \backslash A$ for $i \equiv 1\,(mod\,2)$ and $x_i \in H \setminus B$ for $i \neq 0$ and $i \equiv 0\,(mod\,2)$, then $f \neq 1_F$.*

Definition 2.15. Let G be a group acting on a graph X. A *fundamental domain* of X *mod* G is a subgraph $F \subseteq X$ of X such that $F \to {}_G\backslash^X$ is an isomorphism.

Remark 2.6. Equivalently we can say that F is a fundamental domain of the graph X under the action of G if $|F \cap [x]| = 1$ for every $[x] \in {}_G\backslash^X$, i.e. F contains precisely one element per orbit.

Proposition 2.9. *Let G be a group acting on a tree T. Then T has a fundamental domain mod G if and only if ${}_G\backslash^T$ is a tree.*

Proof. If ${}_G\backslash^T$ is a tree, then by Proposition 2.4, ${}_G\backslash^T$ lifts to a subtree of T, so T has a fundamental domain. If conversely T has a fundamental domain F *mod* G, we need to show that ${}_G\backslash^T$ is a tree. But since as a tree, T is connected and non-empty, so is ${}_G\backslash^T$. Furthermore F is a non-empty connected subgraph of the tree T, hence a tree itself. But since ${}_G\backslash^T$ is by Definition 2.15 isomorphic to F, it has to be a tree as well. □

Definition 2.16. A graph isomorphic to the graph consisting of two vertices and one connecting edge is called *segment* or *interval.*

Theorem 2.10. *Let G be a group acting on a graph X with fundamental domain Y, that is a segment consisting of two vertices p and q and one connecting edge e. Let G_p, G_q and G_e denote the stabilizers of p, q and e respectively. Then X is a tree if and only if the map $\varphi : G_p *_{G_e} G_q \to G$ is an isomorphism with restrictions $\varphi|_{G_p}$ and $\varphi|_{G_q}$ equal to the identity map.*

Proof. Since G_p are G_q are subgroups of G, it is clear that φ is a homomorphism and that its restrictions to G_p and G_q are the identity map. To prove the bijectivity we need two lemmata:

Clearly, the first lemma implies surjectivity, the second injectivity:

Lemma 2.11. *Let G be a group acting on a graph X with fundamental domain Y as above. Then X is connected if and only if $G = \langle G_p \cup G_q \rangle$.*

Lemma 2.12. *Let G be a group acting on a graph X with fundamental domain Y as above. Then X has no circuit if and only if $\varphi : G_p *_{G_e} G_q \to G$ is injective.*

Proof. (of Lemma 2.11)
Define X' to be the connected component of X containing Y, $G' := \{g \in G \mid gX' = X'\}$ to be the subset of G that leaves X' invariant and let $G'' := \langle G_p \cup G_q \rangle \leq G$. First observe that $G'' \subseteq G'$: Let $g \in G''$ and assume $g \in G_p$. That means that $gp = p$ and therefore $gY \cap Y = p$. So Y and gY have the same connected component X'. Therefore $gX' = X'$, which implies $g \in G'$.
Next, observe that $X = G''Y \,\dot\cup\, (G \setminus G'')Y$: Clearly, X is given by the union, but suppose the union is not disjoint. Then there is an $x \in G''$ and a $y \in G \setminus G''$ such that either

$$y^{-1}xp = p \text{ or } y^{-1}xq = q$$

or

$$y^{-1}xp = q \text{ or } y^{-1}xq = p\,.$$

But the first case implies that $y^{-1}x \in G_p \cup G_q = G''$ and therefore $y \in G''$, in contradiction to $y \in G \setminus G''$ and the second case contradicts the assumption of Y being a fundamental domain. So the union is disjoint. But since X' is connected, it cannot have elements in two disjoint sets, so we can conclude $X' \subseteq G''Y$ and therefore $G' \subseteq G''$. So $G' = G''$. Clearly, the graph X is connected if and only $X = X'$, which is the case if and only if $G = G' = G'' = \langle G_p \cup G_q \rangle$. □

Proof. (of Lemma 2.12)
First assume X to have a circuit $c = (w_0, w_1, \ldots, w_n) \subseteq E(X)$ without backtracking. There is an $h_i \in G$ such that $w_i = h_i y_i$ for $y_i \in \{e, \bar{e}\}$ and for $i \in \{0, 1, \ldots n\}$.
Now project c to ${}_G\backslash^{X} = Y$. Since the w_i build a circuit, n is odd and $o(y_i) = t(y_{i-1}) =: r_i \in \{p, q\}$ and $y_{i-1} = \overline{y_i}$ for any $i \in \{1, 2, \ldots, n\}$. Next note

$$h_i r_i = h_i\, o(y_i) = o(h_i y_i) = o(w_i) = t(w_{i-1}) = t(h_{i-1} y_{i-1}) = h_{i-1} r_i\,.$$

So there is a $g_i \in G_{r_i} \subseteq G$ such that $h_i = h_{i-1} g_i$. Suppose $g_i \in G_e$ as well. Then

$$\overline{w_i} = \overline{h_i y_i} = \overline{h_{i-1} g_i y_i} = \overline{h_{i-1} y_i} = h_{i-1}\overline{y_i} = h_{i-1} y_{i-1} = w_{i-1}\,,$$

which contradicts the assumption of c being non-backtracking. Thus, for every

$i \in \{0, 1, \ldots, m\}$ we have $h_i = h_{i-1} g_i$ with $g_i \in G_{r_i} \setminus G_e$. Also, since c is a circuit, $o(w_0) = t(w_n)$, so $r_0 = o(y_0) = t(y_n)$, hence

$$h_0 r_0 = o(w_0) = t(w_n) = h_n\, t(y_n) = h_n r_0\,,$$

which gives

$$h_0 r_0 = h_n r_0 = h_{n-1} g_n r_0 = \cdots = h_0 g_1 \cdots g_n r_0\,.$$

So there is a $g_0 \in G_{r_0} \subseteq G$ such that $g_0 g_1 \cdots g_n = 1_G$, but by Corollary 2.8 of the Normal Form Theorem 2.7, we know $g_0 g_1 \cdots g_n \neq 1_{G_p *_{G_e} G_q}$, so $g_0 g_1 \cdots g_n \in ker(\varphi)$ and non-trivial, thus φ is not injective. □

This completes the proof of Theorem 2.10. □

Theorem 2.13. *Let $G = G_1 *_A G_2$ be an amalgamated product. Then there is a tree X unique up to isomorphism such that G acts on X with fundamental domain a segment Y consisting of two vertices p and q connected by an edge e with stabilizers $G_p = G_1$, $G_q = G_2$ and $G_e = A$.*

Proof. Construct a graph X, which is clearly unique up to isomorphism and on which G acts: Define the vertex set $V(X) \coloneqq G/G_1 \,\dot{\cup}\, G/G_2$, the edge set $E(X) \coloneqq G/A$ and edge maps

$$\begin{aligned} o : G/A &\to G/G_1 & \qquad t : G/A &\to G/G_2 \\ gA &\mapsto gG_1 & gA &\mapsto gG_2\,. \end{aligned}$$

Let Y be the segment of X with vertices G_1 and G_2 connected by the edge A. Clearly G acts on X by left multiplication with stabilizers $G_p = G_1$, $G_q = G_2$ and $G_e = A$, so we are left to show that X is a tree. But this follows directly from Lemma 2.11 and Lemma 2.12. □

Example 2.9. Let $D_\infty = \langle r, s \mid s^2 = 1, srs = r^{-1} \rangle$ be the infinite dihedral group. Consider the barycentric subdivision $B(\mathcal{C}_\infty)$ of the infinite cycle graph $\mathcal{C}_\infty$, i.e. the number line. Clearly, the map

$$\begin{aligned} D_\infty \times V(B(\mathcal{C}_\infty)) &\to V(B(\mathcal{C}_\infty)) \\ (r, i) &\mapsto i + 1 \\ (s, i) &\mapsto -i \end{aligned}$$

extends to an action of D_∞ on $B(\mathcal{C}_\infty)$. Note that it is necessary to consider the barycentric subdivision, because if we would let D_∞ act on $\mathcal{C}_\infty$, the action would not be without inversions, since the element $sr \in D_\infty$ acting on the edge $(-1, 0) \in E(\mathcal{C}_\infty)$ would give $(0, -1) \in E(\mathcal{C}_\infty)$.

We can now consider the factor graph ${}_{D_\infty}\backslash^{B(\mathcal{C}_\infty)}$ and observe that it is a segment with the vertices $0, \frac{1}{2} \in V(B(\mathcal{C}_\infty))$ and their connecting edge $(0, \frac{1}{2}) \in E(B(\mathcal{C}_\infty))$ being a lift in $B(\mathcal{C}_\infty)$. The vertex stabilizer of $0 \in V(B(\mathcal{C}_\infty))$ is $\langle s \rangle \subseteq D_\infty$, since $-0 = 0$ and the vertex stabilizer of $\frac{1}{2} \in V(B(\mathcal{C}_\infty))$ is $\langle sr^{-1} \rangle \subseteq D_\infty$ since $\frac{1}{2} - 1 = -\frac{1}{2}$ and $-(-\frac{1}{2}) = \frac{1}{2}$. The edge stabilizer of $(0, \frac{1}{2}) \in E(B(\mathcal{C}_\infty))$ is by construction $\{1\}$. Hence, since $\mathcal{C}_\infty$ is a tree, Theorem 2.10 implies that D_∞ is isomorphic to $\langle s \rangle *_{\{1\}} \langle sr^{-1} \rangle$. But with the relations of D_∞ we can deduce

$$s^2 = 1 \quad \Leftrightarrow \quad s = s^{-1}$$

and thus

$$\begin{aligned} srs = r^{-1} \quad &\Leftrightarrow \quad rs = s^{-1}r^{-1} \\ &\Leftrightarrow \quad rs^{-1} = sr^{-1} \\ &\Leftrightarrow \quad (sr^{-1})^{-1} = sr^{-1} \,. \end{aligned}$$

We can define $sr^{-1} =: t$, thus $\langle s \rangle \cong \mathbb{Z}_2 \cong \langle t \rangle$, which implies that

$$D_\infty \cong \mathbb{Z}_2 * \mathbb{Z}_2 \,.$$

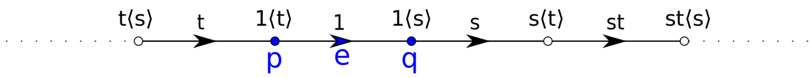


Figure 3: *As the infinite dihedral group is isomorphic to the amalgamated product $\langle s \rangle * \langle t \rangle$, where $\langle s \rangle \cong \mathbb{Z}_2 \cong \langle t \rangle$, we obtain this Bass-Serre tree [c.f. Remark 2.16].*

Example 2.10. Consider the amalgamated product

$$G := \mathbb{Z}/3\mathbb{Z} *_{\{1\}} \mathbb{Z}/4\mathbb{Z} = \langle a \mid a^3 = 1 \rangle * \langle b \mid b^4 = 1 \rangle \,.$$

By Theorem 2.13, there exists a tree on which G acts. By using the proof of the tree, we know how to construct the tree. We obtain the $(3, 4)$-biregular tree given in Figure 4 below. The vertices are cosets of the form $g\mathbb{Z}/3\mathbb{Z}$ and $g\mathbb{Z}/4\mathbb{Z}$ with $g \in G$ and the edge maps are the following:

$$\begin{aligned} o : G &\to G/\mathbb{Z}/3\mathbb{Z} \\ g &\mapsto g\mathbb{Z}/3\mathbb{Z} \end{aligned} \qquad\qquad \begin{aligned} t : G &\to G/\mathbb{Z}/4\mathbb{Z} \\ g &\mapsto g\mathbb{Z}/4\mathbb{Z} \,. \end{aligned}$$

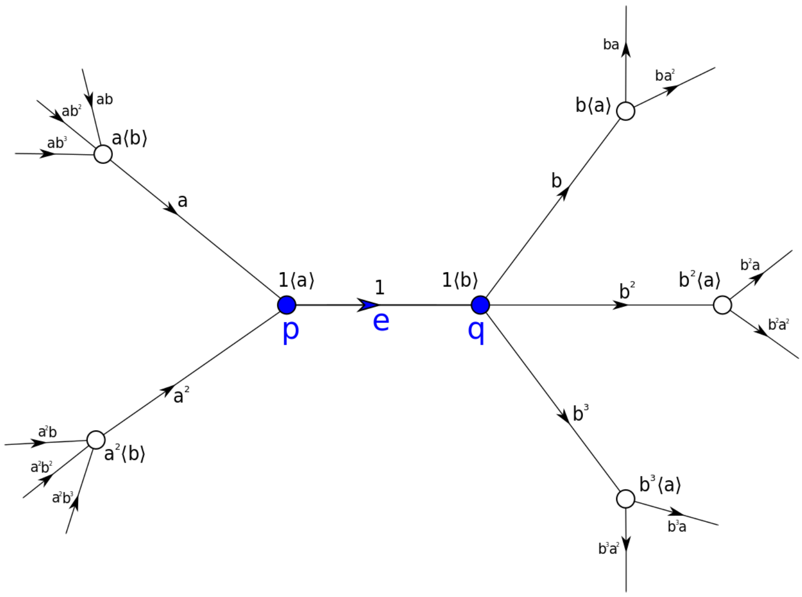


Figure 4: *The amalgamated product $G = \mathbb{Z}/3\mathbb{Z} * \mathbb{Z}/4\mathbb{Z} = \langle a \mid a^3 = 1\rangle * \langle b \mid b^4 = 1\rangle$ acts on this tree. The cosets of $\mathbb{Z}/3\mathbb{Z}$ and $\mathbb{Z}/4\mathbb{Z}$ are abbreviated by elements of the form $g\langle a\rangle$ and $g\langle b\rangle$ respectively.*

It will be useful later to state the following generalization of the necessary condition of Lemma 2.11: If a group is acting on a connected graph with fundamental domain a segment, then it is generated by the union of the vertex stabilizers.

Lemma 2.14. *[43, p.34, Lemma 4] Let G be a group acting on a connected graph X. Let T be a tree of representatives of $Y = {}_G\backslash^X$ and let $W \subseteq X$ be a subgraph of X containing T such that every edge in $E(W)$ has one vertex in $V(T)$ and $G \cdot W = X$. Furthermore define for every edge $e \in E(W \setminus T)$ with $o(e) \in V(T)$ an element $\gamma_e \in G$ such that $\gamma_e t(e) \in V(T)$. Then the group H generated by the elements γ_e and the vertex stabilizers G_p for $p \in V(T)$ is equal to G.*

Remark 2.7. If in Lemma 2.14 T is a fundamental domain, then we can take $W = T$ and all the $\gamma_e = 1_G$ for $e \in E(T)$.

We want to give a first application of when an amalgamated product is used to determine whether a group is residually finite, introduced by Graham Higman in [17]:

Example 2.11. We aim to construct a group with three generators and two defining relations that is isomorphic to a proper quotient of itself and therefore not Hopfian and thus by Malcev's Theorem 1.20 not residually finite.
We start with the following definitions: Let H_1 be a group with two subgroups $V_1, U_1 \leq H_1$ such that $V_1 \subseteq U_1$ and let H_2 be a group with subgroup U_2. Let there be an isomorphism $\alpha : U_1 \to U_2$ and define the image of V_1 under α as $V_2 \coloneqq \alpha(V_1)$. Denote the amalgamated products of H_1 and H_2 with the identifications $U_1 \cong U_2$ and $V_1 \cong V_2$ as G and K respectively, i.e.

$$G = H_1 *_{U_1 \cong U_2} H_2 \quad \text{and} \quad K = H_1 *_{V_1 \cong V_2} H_2 \,.$$

Then, by construction, there is a homomorphism $\varphi : G \to K$, that cannot be an isomorphism, as long as $V_1 \subsetneq U_1$ is a proper subset. Hence, by the Isomorphism Theorem

$$G \cong K/ker(\varphi)\,. \tag{10}$$

Now we want to show that G is in fact isomorphic to K, to ensure G has the desired property. In order to do so we suppose there is an isomorphism $\mu : U_1 \to V_1$ such that it can be extended to an automorphism $\bar{\mu} : H_1 \to H_1$ and that the isomorphism $\alpha \circ \mu \circ \alpha^{-1} : U_2 \to V_2$ can be extended to an automorphism $\alpha \circ \bar{\mu} \circ \alpha^{-1} : H_2 \to H_2$.
As $H_1, H_2 \leq G$ and $H_1, H_2 \leq K$ are subgroups of the amalgamated products, we can think of the automorphisms as restrictions of maps from G to K and observe the following: The maps do not only agree on the intersection of H_1 and H_2 in G, which is $U_1 \cong U_2$, but they also map it to the intersection of H_1 and H_2 in K, which is $V_1 \cong V_2$. But since G and K are amalgamated products where the intersections of H_1 and H_2 are identified, G and K are isomorphic, hence by condition (10), G is isomorphic to a proper quotient of itself.
With the existence of such a group G insured, we want to find the desired presentation for it. Take

$$H_1 = \langle x, z \mid x^{-1}zx = z^2 \rangle \quad \text{and} \quad H_2 = \langle y, w \mid y^{-1}wy = w^2 \rangle$$

with

$$U_1 = \langle z \rangle \quad \text{and} \quad V_1 = \langle z^2 \rangle\,.$$

Then the obvious choice of

$$U_2 = \langle w \rangle \quad \text{and} \quad V_2 = \langle w^2 \rangle$$

gives us the group

$$G = \langle x, y, z \mid x^{-1}zx = y^{-1}zy = z^2 \rangle \, .$$

Now consider the map induced by

$$\begin{aligned} \phi : G &\to G \\ x &\mapsto x \\ y &\mapsto y \\ z &\mapsto z^2 \, . \end{aligned}$$

Since

$$\begin{aligned} \phi([x^{-1}y, z]) &= \phi((x^{-1}y)^{-1}z^{-1}x^{-1}yz) \\ &= \phi(y^{-1}xz^{-1}x^{-1}yz) \\ &= y^{-1}xz^{-2}x^{-1}yz^2 \\ &= y^{-1}xx^{-1}z^{-1}xx^{-1}yz^2 \\ &= y^{-1}z^{-1}yy^{-1}zy \\ &= 1_G \, , \end{aligned}$$

we conclude that $[x^{-1}y, z] \in ker(\phi)$. But the Normal Form Theorem 2.7 tells us that $[x^{-1}y, z] \neq 1_G$, which implies that ϕ is not injective. So we can consider $G/ker(\phi)$ and deduce, since $[x^{-1}y, z] \in ker(\phi)$ from the identity $[x^{-1}y, z]\, ker(\phi) = 1_{G/ker(\phi)}$, the relation

$$\begin{aligned} [x^{-1}y, z]\, ker(\phi) = 1_{G/ker(\phi)} &\Leftrightarrow [x^{-1}y, z] = 1 \\ &\Leftrightarrow y^{-1}xz^{-1}x^{-1}yz = 1 \\ &\Leftrightarrow xz^{-1}x^{-1}yz = z \\ &\Leftrightarrow xz^{-1}x^{-1} = yz^{-1}y^{-1} \\ &\Leftrightarrow (xz^{-1}x^{-1})^{-1} = (yz^{-1}y^{-1})^{-1} \\ &\Leftrightarrow xzx^{-1} = yzy^{-1} \, . \end{aligned}$$

Now we define

$$G' := \langle x, y, z \mid x^{-1}zx = y^{-1}zy = z^2, xzx^{-1} = yzy^{-1} \rangle \tag{11}$$

and want to check that G' is isomorphic to G. Therefore we first observe that the relation

$$xzx^{-1} = yzy^{-1} \tag{12}$$

in G' is not deducible from the relation

$$x^{-1}zx = y^{-1}zy = z^2 \tag{13}$$

given in the presentation of G, since (12) equates an element in $\langle x, z\rangle \setminus \langle z\rangle$ with an element in $\langle y, z\rangle$, whereas (13) implies that $\langle x, z\rangle$ and $\langle y, z\rangle$ intersect in $\langle z\rangle$. But on the other hand, we can use Tietze transformations to get another presentation of G'. In particular we use T_3 and T_4: Add the new generator v with the relation $v = xzx^{-1}$ and derive from it with (13)

$$\begin{gathered}
xzx^{-1} = v \Leftrightarrow x^{-1}x^{-1}xzx^{-1}xx = x^{-1}x^{-1}vxx \Leftrightarrow x^{-1}zx = x^{-2}vx^2 \\
xzx^{-1} = v \Leftrightarrow z = x^{-1}vx \Leftrightarrow y^{-1}zy = y^{-1}x^{-1}vx^{-1}y \\
xzx^{-1} = v \Leftrightarrow z = x^{-1}vx \Leftrightarrow z^2 = x^{-1}v^2x
\end{gathered}$$

to obtain the new relation

$$x^{-2}vx^2 = y^{-1}x^{-1}vx^{-1}y = x^{-1}v^2x\,.$$

That immediately implies (via the first and the last term)

$$x^{-1}vx = v^2\,. \tag{14}$$

Furthermore with inserting that new relation in (12) we can derive

$$\begin{gathered}
xzx^{-1} = v \Leftrightarrow z = x^{-1}vx \Leftrightarrow x^{-1}xvxx^{-1} = yx^{-1}vxy^{-1} \\
\Leftrightarrow v = yx^{-1}vxy^{-1} \Leftrightarrow y^{-1}vy = x^{-1}vx
\end{gathered}$$

to obtain the new relation

$$y^{-1}vy = x^{-1}vx\,. \tag{15}$$

So putting (14) and (15) together we obtain

$$y^{-1}vy = x^{-1}vx = v^2\,.$$

It is an easy check that this relation is not only a conclusion of, but also implies relations (12) and (13), which means that the presentation (11) of G' is in fact via Tietze transformations equivalent to the presentation

$$G' = \langle x, y, v \mid y^{-1}vy = x^{-1}vx = v^2 \rangle \,,$$

which is isomorphic to the original presentation of G. So we have found a group G isomorphic to a proper quotient of itself, i.e. G is not Hopfian and the proof is done.

Remark 2.8. It is apparent that in the above Example 2.11 the choice of H_1 and H_2 and with it of V_1 can be generalized to

$$H_1 = \langle x, z \mid x^{-1}zx = z^h \rangle \,, \quad H_2 = \langle y, w \mid y^{-1}wy = w^2 \rangle \quad \text{and} \quad V_1 = \langle z^h \rangle$$

for any $h \in \mathbb{Z} \setminus \{-1, 0, 1\}$.

2.3.3 HNN-extension

Definition 2.17. Let $G = \langle S \mid R \rangle$ a group with subgroups $H, K \leq G$ and let $\alpha : H \to K$ be a homomorphism. Let $t \notin S$ be a new symbol. The group

$$G*_\alpha \coloneqq \langle S \cup \{t\} \mid R \cup \{tht^{-1} = \alpha(h) \mid h \in H\} \rangle$$

is called *HNN-extension* of G relative to α.

Example 2.12.

(i). Let $G = \mathbb{Z}/2\mathbb{Z} = \langle a \mid a^2 = 1 \rangle$ be the cyclic group of order two with the automorphism

$$\begin{aligned} \alpha : G &\to G \\ a &\mapsto a \,. \end{aligned}$$

Then the HNN-extension of G relative to α is given by

$$G*_\alpha = \langle G \cup \{t\} \mid a^2 = 1, t^{-1}at = a \rangle \,.$$

(ii). Let $V_4 = \langle a, b \mid a^2 = b^2 = (ab)^2 = 1 \rangle$ be again the Klein four-group. Let $A = \langle a \mid a^2 = 1 \rangle$ and $B = \langle b \mid b^2 = 1 \rangle$ be two subgroups of V_4 that are isomorphic via the map

$$\begin{aligned} \alpha : A &\to B \\ a &\mapsto b \,. \end{aligned}$$

Then the HNN-extension of V_4 relative to α is given by

$$V_4 *_\alpha = \langle a, b, t \mid a^2 = b^2 = (ab)^2 = 1, t^{-1}at = b \rangle \,.$$

The contruction of the HNN-extension was first proven by Graham Higman, Bernhard H. Neumann and Hanna Neumann in [18], hence the name was chosen as initials of their names in their honour.

Proposition 2.15 (HNN-construction)**.** *Let G be a group with a subgroup $A \leq G$ and an injective homomorphism $\theta : A \to G$. Then there is a group G', that contains G and an element $s \in G' \setminus G$ such that $\theta(a) = sas^{-1}$ for every $a \in A$.*

Proof. Define for every $n \in \mathbb{Z}$ two groups $A_n := A$ and $G_n := G$. Hence for every $n \in \mathbb{Z}$ we can also find two injective homomorphisms: $\theta : A_n = A \to G = G_n$ and the canonical injection $id : A_n = A \to G = G_{n+1}$. Let $H = *^n_{i(A) \cong \theta(A)} G_n$ be the amalgamated product of all G_n, identifying the images of the homomorphisms with each other. Next, let u_n be the canonical homomorphism $G_n \to G_{n+1}$. The u_n define an automorphism $u : H \to H$. If we consider an element $a \in A \leq G = G_0$, then its image $u(a) \in G_1 = G$ is equal to the element $\theta(a) \in G_0$. Hence, u extends θ.
Now let $S = \langle s \rangle$ be an infinite, cyclic group generated by an element s $\notin G$ and consider the semi-direct product $G' := H \rtimes_u S$, given by the homomorphism

$$\begin{aligned} \phi : (S, H) &\to H \\ (s, h) &\mapsto u(h) \,. \end{aligned}$$

By [24, p. 140, Corollary 2], we know if H has the presentation $H = \langle T \mid R \rangle$ that G' is presented as

$$\langle T \cup \{s\} \mid R, sts^{-1} = \phi(s)(t); \forall\, t \in T \rangle \,,$$

hence

$$G' = \langle T \cup \{s\} \mid R, sts^{-1} = u(h); \forall\, t \in T \rangle \,.$$

But since u extends θ, we know that $u(a) = \theta(a)$ for every $a \in A$, implying that

$$\theta(a) = sas^{-1}$$

for every $a \in A$. $\square$

Remark 2.9. One often says that the group G' in Proposition 2.15 is derived from (A, G, θ) by the *HNN-construction*.

As for amalgamated products in Definition 2.14, we can introduce a normal form for elements of an HNN-extension.

Definition 2.18.

(i). Let $G = H*_\alpha = \langle H, t \mid t^{-1}at = \alpha(a),\, a \in A\rangle$ be an HNN-extension given by the group H and its subgroups A and B isomorphic via the map α. Let $i : H * \langle t\rangle \to G$ be the canonical isomorphism from the free product into the HNN-extension. Then every element $g \in G$ can be written as $g = i(g_0)i(t)^{\varepsilon_1}i(g_1)\cdots i(t)^{\varepsilon_n}i(g_n)$ where $g_i \in H$ and $\varepsilon_i \in \{\pm 1\}$ for all $i \in \{1, 2, \ldots, n\}$. To make matters easier we will denote such an elements $g \in G$ simply as $g = g_0t^{\varepsilon_1}g_1\cdots t^{\varepsilon_n}g_n$.

(ii). Let $A, B \leq H$ be two subgroups of the group H. Then, analogously to Definition 2.14 (ii), T_A denotes the system of representatives of right cosets of A in H and T_B denotes the system of representatives of right cosets of B in H. Assume that $1_H \in T_A$ and $1_H \in T_B$. Moreover, denote the unique presentation of an element $g \in G$ as the product of an element of A and an element of T_A by

$$g = \underline{x}\overline{x} \qquad \text{for } \underline{x} \in A \text{ and } \overline{x} \in T_A$$

and analogously for an element $h \in H$

$$h = \underline{y}\overline{y} \qquad \text{for } \underline{y} \in B \text{ and } \overline{y} \in T_B\,.$$

(iii). For an element $g = g_0t^{\varepsilon_1}g_1\cdots t^{\varepsilon_n}g_n \in G$ a sequence $(g_0, t^{\varepsilon_1}, g_1, \ldots, t^{\varepsilon_n}, g_n)$ is called *normal form* if

- $g_0 \in G$
- $\varepsilon_i = -1$ implies that $g_i \in T_A$
- $\varepsilon_i = 1$ implies that $g_i \in T_B$
- there is no consecutive subsequence $t^{\varepsilon_i}, 1, t^{-\varepsilon_i}$ for any $i \in \{1, 2, \ldots n\}$.

We call n be called the G-length of g.

Example 2.13. Let $H = \langle a, b\rangle = F_2$ be the free group of rank 2 and let $A = \langle a^2\rangle$ and $B = \langle b^3\rangle$ be two subgroups isomorphic via the map

$$\begin{aligned} \alpha : A &\to B \\ a^2 &\mapsto b^3\,. \end{aligned}$$

Let T_A be the set of all reduced words in H not starting with any power of a united with all the reduced words in H starting with a and let T_B be the set of all reduced words in H not starting with any power of b united with all the reduced words in H starting with either b or b^2. Consider the element

$$g = b^2 t^{-1} a^{-4} t b^5 a b t^{-1} a^4 b^3 a \in H *_\alpha =: G\,.$$

We can start rewriting g from the right side using the relations $t^{-1}a^2 = b^3 t^{-1}$ and $tb^3 = a^2 t$ to obtain its normal form:

$$\begin{aligned} g = b^2 t^{-1} a^{-4} t b^5 a b t^{-1} a^4 b^3 a &= b^2 t^{-1} a^{-4} t b^5 a b b^3 b^3 t^{-1} b^3 a \\ &= b^2 t^{-1} a^{-4} t b^5 a b^7 t^{-1} b^3 a \\ &= b^2 ((b^3)^{-1})^2 b^5 a b^7 t^{-1} b^3 a \\ &= b a b^7 t^{-1} b^3 a\,. \end{aligned}$$

Hence, the normal form of $g = b^2 t^{-1} a^{-4} t b^5 a b t^{-1} a^4 b^3 a$ is $(bab^7, t^{-1}, b^3 a)$ since $bab^7 \in H$ and $b^3 a \in T_A$.

Theorem 2.16 (Normal Form Theorem for HNN-extensions)**.** *[5, p. 82, Theorem 14.3] Let H be a group with subgroups A and B isomorphic via the map $\alpha : A \to B$ and let $G = \langle H, t \mid t^{-1}at = \varphi(a),\, a \in A\rangle$ be its HNN-extension. Then for every $g \in G$ there is a unique presentation $g = g_0 t^{\varepsilon_1} g_1 \cdots t^{\varepsilon_n} g_n \in G$, where the sequence $(g_0, t^{\varepsilon_1}, g_1, \ldots, t^{\varepsilon_n}, g_n)$ is a normal form.*

Again, we can deduce:

Corollary 2.17 (Britton's Lemma)**.** *Let H be a group with two isomorphic subgroups A and B via the map α and let $G = H*_\alpha$ be the associated HNN-extension. If an element $g \in G$ is of the form $g = g_0 t^{\varepsilon_1} g_1 \cdots t^{\varepsilon_n} g_n$ with $n \geq 1$ and $g_i \in H$ for all $i \in \{1, 2, \ldots, n\}$, then $g \neq 1_G$.*

Theorem 2.18. *Let G be a group acting on a graph X with the quotient ${}_G\backslash^X$ being a loop. Let $F \subseteq X$ be a segment in X consisting of two vertices p and q connected by the edge f with stabilizers G_p, G_q and G_f. Let $x \in G$ such that $xp = q$ and define $G'_f := x^{-1} G_f x$ with $\phi : G_f \to G'_f$ the homomorphism induced by conjugation. Then X is a tree if and only if $G'_f \leq G_p$ and the map*

$$\begin{aligned} \psi : G_p *_\phi = \langle G_p \cup \{t\} \mid R_p \cup \{t^{-1}at = \phi(a) \mid a \in G_f\}\rangle &\to G \\ g &\mapsto g \qquad \forall\, g \in G_p \\ t &\mapsto x \end{aligned}$$

is an isomorphism.

Proof. The proof is similar to the proof of Theorem 2.10 and will be the Corollary 2.33 of the Structure Theorem 2.32 in the next Chapter. □

Theorem 2.19. *Let $G = H*_\alpha$ be an HNN-extension of H with subgroups A and $\alpha(A)$ for a monomorphism $\alpha : A \to G$. Then there is a tree X unique up to isomorphism such that G acts on X with the quotient ${}_G\backslash^X$ being a loop. Moreover there is a segment $Y \subseteq X$ such that the stabilizers of vertices and edges in the group G are given by H, sHs^{-1} and A respectively.*

Proof. The proof is very similar to the proof of Theorem 2.13 and can therefore be omitted. □

2.3.4 Graph of groups

So far we have only dealt with two special cases of actions on trees, namely when the action is free and when the quotient of the graph by the action is a segment or a loop. So next we want to broaden our concept to a more general case:

Definition 2.19. Let X be a directed, connected graph and let $\mathcal{G}$ be a set of groups. A *graph of groups* is the pair $(\mathcal{G}, X)$, where every vertex $v \in V(X)$ is identified with a group $G_v \in \mathcal{G}$ and every edge $e \in E(X)$ is identified with a group $G_e \in \mathcal{G}$ such that $G_e = G_{\bar{e}}$ and for every $e \in E(X)$ there is a monomorphism

$$\begin{aligned} \alpha_e : G_e &\hookrightarrow G_{t(e)} \\ a &\mapsto a^e \,. \end{aligned}$$

We denote the image $\alpha_e(G_e)$ of G_e in $G_{t(e)}$ under the monomorphism α_e with G_e^e for every $e \in E(X)$.

Remark 2.10. Instead of requiring $G_e = G_{\bar{e}}$ in Definition 2.19 of the graph of groups, we can also demand to have a second monomorphism $G_e \hookrightarrow G_{o(e)}$ for every $e \in E(X)$.

Definition 2.20. Let $(\mathcal{G}, X)$ be a graph of groups. Define the group $\mathcal{F}(\mathcal{G}, X)$ as follows:
$\mathcal{F}(\mathcal{G}, X)$ is generated by all groups $G_v = \langle G_v \mid R_v \rangle$ for $v \in V(X)$ and by the elements $e \in E(X)$ with the additional relations $\bar{e} = e^{-1}$ and $ea^e e^{-1} = a^{\bar{e}}$ for any $e \in E(X)$, i.e.

$$\mathcal{F}(\mathcal{G}, X) = \langle \{G_v \mid v \in V(X)\} \cup E(X) \mid \{R_v \mid v \in V(X)\} \cup \tilde{R}\} \rangle$$

$$\text{with } \tilde{R} = \{\bar{e} = e^{-1}, ea^e e^{-1} = a^{\bar{e}} \mid e \in E(X),\, a \in G_e\} \,.$$

Definition 2.21. Let $(\mathcal{G}, X)$ be a graph of groups and let $c = c_1 c_2 \cdots c_n$ be a path in X with $c_i \in E(X)$ for all $i \in \{1, , 2 \ldots n\}$. A word $w \in \mathcal{F}(\mathcal{G}, X)$ is said to be of *type* c given by $(v_0, v_1, \ldots, v_n)$ with $v_i \in G_{o(c_{i+1})}$ for any $i \in \{0, 1, \ldots, n-1\}$ and with $v_n = G_{t(c_n)}$ if

$$w = v_0 c_1 v_1 c_2 \cdots c_n v_n \,.$$

Definition 2.22. Let $(\mathcal{G}, X)$ be a graph of groups. Its *fundamental group* $\pi_1(\mathcal{G}, X, p_0)$ *with base point* $p_0 \in V(X)$ is the set of all words in $\mathcal{F}(\mathcal{G}, X)$ of type c, where c is a circuit starting and ending at point p_0, i.e.

$$\pi_1(\mathcal{G}, X, p_0) = \{w \in \mathcal{F}(\mathcal{G}, X) \mid w \text{ of type } c, p_0 = o(c) = t(c)\} \,.$$

Example 2.14. If $G_v = \{1\}$ for every $v \in V(X)$, then $\pi_1(\mathcal{G}, X, p_0) = \pi_1(X, p_0)$, i.e. the fundamental group with base point p_0 in the sense of algebraic topology.

Definition 2.23. Let $(\mathcal{G}, X)$ be a graph of groups and let $T \subseteq X$ be a maximal subtree of X. Its *fundamental group* $\pi_1(\mathcal{G}, X, T)$ *with respect to* T is the quotient of $\mathcal{F}(\mathcal{G}, X)$ by the smallest normal subgroup containing all edges of T, i.e.

$$\mathcal{F}(\mathcal{G}, X) / \langle\langle \{e \in E(T)\} \rangle\rangle \,.$$

Thus, if we denote the image of the edge $e \in \mathcal{F}(\mathcal{G}, X)$ under the projection $p : \mathcal{F}(\mathcal{G}, X) \to \pi_1(\mathcal{G}, X, T)$ by g_e, the group $\pi_1(\mathcal{G}, X, T)$ is generated by the groups G_p for $p \in V(X)$ and the elements g_e for $e \in E(X)$ subject to the relations

- if $e \in E(X) \setminus E(T)$ and $a \in G_e$:

$$\begin{aligned} g_e a^e g_e^{-1} &= a^{\bar{e}} \\ g_{\bar{e}} &= g_e^{-1} \,; \end{aligned}$$

- if $e \in E(T)$ and $a \in G_e$:

$$\begin{aligned} g_e &= 1 \\ a^e &= a^{\bar{e}} \,. \end{aligned}$$

In fact, these two definitions for the fundamental group are equivalent:

Theorem 2.20. *[43, p. 44, Proposition 20] Let $(\mathcal{G}, X)$ be a graph of groups, let $p_0 \in V(X)$ and let T be a maximal subtree of X. Then the canonical projection*

$$p : \mathcal{F}(\mathcal{G}, X) \to \pi_1(\mathcal{G}, X, T)$$

induces an isomorphism of $\pi_1(\mathcal{G}, X, p_0)$ *onto* $\pi_1(\mathcal{G}, X, T)$.

As a direct consequence we get independence of the choice of the base point and the the maximal subtree:

Corollary 2.21. *Let* $(\mathcal{G}, X)$ *be a graph of groups. Then* $\pi_1(\mathcal{G}, X, p_0)$ *is isomorphic to* $\pi_1(\mathcal{G}, X, T)$ *for every* $p_0 \in V(X)$ *and for every maximal subtree* $T \subseteq X$.

Remark 2.11. We will denote this isomorphism class in Corollary 2.21 by $\pi_1(\mathcal{G}, X)$ and use from now on whichever of the two Definitions 2.22 and 2.23 is better suited for the problem we are dealing with.

Example 2.15.

(i). Let $(\mathcal{G}, X)$ be a graph of groups with X a segment given by to vertices $p, q \in V(X)$ with one connecting edge $e \in E(X)$ and associated groups G_p, G_q and G_e in $\mathcal{G}$ respectively. Then X itself is its maximal spanning tree. Hence in the construction of the fundamental group $\pi_1(\mathcal{G}, X, X)$, we get $g_e = 1$ for every $e \in E(X)$, which leads to the relation

$$a^e = a^{\bar{e}} \qquad \text{for every } e \in E(X) \text{ and every } a \in G_e\ .$$

But this implies that the subgroups of G_p and G_q isomorphic to G_e are identified with each other, hence

$$\pi_1(\mathcal{G}, X, X) = G_p *_{G_e} G_q\,.$$

(ii). More generally, let $(\mathcal{G}, X)$ be a graph with X a tree. Then, again X itself is its maximal spanning tree. Hence $\pi_1(G, X, X) = *_{G_e}^{p \in V(X)} G_p$ along the edge groups G_e for $e \in E(X)$.

(iii). Let $(\mathcal{G}, X)$ be a graph of groups with X a single vertex $p \in V(X)$ with one loop $e \in E(X)$ attached. Define $A \coloneqq G_e = G_{\bar{e}}$ and observe that we have two monomorphisms

$$A \overset{e}{\hookrightarrow} G_p \quad \text{and} \quad A \overset{\bar{e}}{\hookrightarrow} G_p\,.$$

Since every path c in X can only go along e and $\bar{e}$, we get that $\mathcal{F}(\mathcal{G}, X) = \pi_1(\mathcal{G}, X, p)$. Furthermore $\pi_1(\mathcal{G}, X, p)$ is generated by G_p and an element $g \coloneqq g_e$ with the defining relation

$$g a^e g^{-1} = a^{\bar{e}} \quad \text{for every } a \in A\,.$$

We can now identify A with a subgroup of $G \coloneqq G_p$ via the homomorphism

$$\begin{aligned} A &\to G \\ a &\mapsto a^e \end{aligned}$$

and define a second homomorphism

$$\begin{aligned} \theta : A &\to G \\ a &\mapsto a^{\bar{e}} \end{aligned}$$

to conclude that $\pi_1(\mathcal{G}, X, p)$ is the group derived from (A, G, θ) by the HNN-construction in Proposition 2.15. Hence $\pi_1(\mathcal{G}, X, p)$ is the semi-direct product of the infinite cyclic group generated by g and the normal subgroup $R = \langle\langle G \rangle\rangle$ generated by all the conjugates $G_n \coloneqq g^n G g^{-n}$ of G for $n \in \mathbb{Z}$, i.e. $\pi_1(\mathcal{G}, X, p) = R \rtimes \langle g \rangle$. Furthermore, the HNN-construction gives us that R is the amalgamated product of the G_n according to the homomorphisms

$$\begin{aligned} A &\to G_{n-1} \\ a &\mapsto g^{n-1} a^{\bar{e}} g^{1-n} \end{aligned} \qquad \text{and} \qquad \begin{aligned} A &\to G_n \\ a &\mapsto g^n a^e g^{-n} . \end{aligned}$$

(iv). For any arbitrary graph of groups $(\mathcal{G}, X)$, we can obtain its fundamental group by successively performing an amalgamated product for every edge in the maximal subtree and an HNN-extension for every edge not in the maximal subtree.

2.3.4.1 Reduced Words

Use the definition of a reduced word in our setting:

Definition 2.24. Let $(\mathcal{G}, X)$ be a graph of groups and let w be a word of type $c = c_1 \cdots c_n$, i.e.

$$w = v_0 c_1 v_1 c_2 \cdots c_n v_n$$

with $v_i \in G_{o(c_{i+1})}$ for any $i \in \{0, 1, \dots, n-1\}$ and with $v_n = G_{t(c_n)}$. If $n = 0$, the word w is called *reduced* if $v_0 \neq 1$. If $n \geq 1$, the word w called *reduced* if $v_i \notin G_{c_i}^{c_i}$, whenever $c_{i+1} = \bar{c}_i$ holds for an $i \in \{1, 2, \dots n\}$.

Remark 2.12. By definition, every word of type c, where c is a path without backtracking, is reduced.

The following example shows that a reduced expression of an element in the fundamental group of a graph of groups is in general not unique:

Example 2.16. Let $(\mathcal{G}, X)$ be a graph of groups with X a graph with vertex set $V(X) = \{u, v\}$ and two connecting edges, i.e. $E(X) = \{e_1, \bar{e_1}, e_2, \bar{e_2}\}$ with $o(e_1) = o(e_2) = u$ and $t(e_1) = t(e_2) = v$. Let the associated vertex and edge groups be given by

$$\begin{aligned} G_u = \langle a \mid a^{12} = 1\rangle \qquad &\text{and} \qquad G_v = \langle b \mid b^{18} = 1\rangle \\ G_{e_1} = G_{\bar{e}_1} = \langle c \mid c^2 = 1\rangle \qquad &\text{and} \qquad G_{e_2} = G_{\bar{e}_2} = \langle d \mid d^3 = 1\rangle \end{aligned}$$

with monomorphisms

$$\begin{aligned} \alpha_{e_1} : G_{e_1} &\hookrightarrow G_u & \alpha_{\bar{e}_1} : G_{\bar{e}_1} &\hookrightarrow G_v \\ c &\mapsto a^6 & c &\mapsto b^9 \\ \\ \alpha_{e_2} : G_{e_2} &\hookrightarrow G_u & \alpha_{\bar{e}_2} : G_{\bar{e}_2} &\hookrightarrow G_v \\ d &\mapsto a^4 & d &\mapsto b^6 \,. \end{aligned}$$

Define T to be the maximal subtree of Y given by the vertices $u, v \in V(X)$ and the edges $e_1, \bar{e}_1 \in E(X)$. Then the group $\mathcal{F}(\mathcal{G}, X)$ is generated by the elements $\{a, b, e_1, e_2, \bar{e}_1, \bar{e}_2\}$ and obeys the relations

$$a^{12} = b^{18} = 1,\, e_1^{-1} = \bar{e}_1,\, e_2^{-1} = \bar{e}_2,\, e_1 a^6 e_1^{-1} = b^9 \quad \text{and} \quad e_2 a^4 e_2^{-1} = b^6 \,.$$

To obtain the fundamental group $\pi_1(\mathcal{G}, X, T)$ we need to take the quotient of $\mathcal{F}(\mathcal{G}, X)$ by the normal closure of the edges in T, i.e. by the normal closure of $\{e_1, \bar{e}_1\}$. Hence, we obtain

$$\pi_1(\mathcal{G}, X, T) = \langle a, b, e_2, \bar{e}_2 \mid a^{12} = b^{18} = 1, e_2^{-1} = \bar{e}_2, a^6 = b^9, e_2 a^4 e_2^{-1} = b^6\rangle \,.$$

To make the notation a bit more accessible, note that via Tietze transformations $\pi_1(\mathcal{G}, X, T)$ is isomorphic to the group

$$G = \langle a, b, t \mid a^{12} = b^{18} = 1, a^6 = b^9, t a^4 t^{-1} = b^6\rangle \,,$$

by adding the generator t and the relation $e_2 = t$. Consider now the element $g = bta^3t^{-1}a^6b^3t \in G$ and compute two reduced expressions for it by using the relations. First observe

$$
\begin{aligned}
g = bta^3t^{-1}a^6b^3t \quad & \overset{a^6=b^9}{=} \quad bta^3t^{-1}b^{12}t \\
& \overset{b^{12}=ta^8t^{-1}}{=} \quad bta^3t^{-1}ta^8t^{-1}t \\
& = \quad bta^{11} \\
& \overset{a^{12}=1}{=} \quad bta^{-1}\,,
\end{aligned}
$$

which is reduced. Now, since we are allowed to insert the trivial element $1 = b^{-6}b^6$ at any position, we can also derive

$$
\begin{aligned}
g = bta^3t^{-1}a^6b^3t \quad & = \quad bb^{-6}b^6ta^3t^{-1}a^6b^3t \\
& \overset{b^6=ta^4t^{-1}}{=} \quad b^{-5}ta^4t^{-1}ta^3t^{-1}a^6b^3t \\
& = \quad b^{-5}ta^7t^{-1}a^6b^3t^{-1} \\
& \overset{a^6=b^9}{=} \quad b^{-5}ta^7t^{-1}b^{12}t \\
& \overset{t^{-1}b^{12}t=a^8}{=} \quad b^{-5}ta^{15} \\
& \overset{a^{12}=1}{=} \quad b^{-5}ta^3\,,
\end{aligned}
$$

which gives a second reduced expression for the element g.

Theorem 2.22. *Let w be a reduced word of type c in the graph of groups $(\mathcal{G}, X)$. Then its associated element in $\mathcal{F}(\mathcal{G}, X)$ is not trivial.*

Before we can tackle the proof, we need some observations and preliminaries. Let us first draw the following conclusions supposing that Theorem 2.22 is true:

Corollary 2.23. *The homomorphism $G_p \to \mathcal{F}(\mathcal{G}, X)$ is injective for every $G_p \in \mathcal{G}$ with $p \in V(X)$.*

Proof. Since Theorem 2.22 holds for every path c in X, we can choose it to be consisting of only one vertex $p \in V(X)$ and no edge, hence it is of length zero. Then all words of type c need to be in G_p and since their image cannot be trivial in $\mathcal{F}(\mathcal{G}, X)$, the homomorphism has to be injective. $\square$

Corollary 2.24. *Let $w = v_0c_1v_1c_2 \cdots c_nv_n$ be a reduced word of type $c = c_1 \cdots c_n$ in the graph of groups $(\mathcal{G}, X)$ with length $l(c) \geq 1$ and $o(c) = p_0$. Then $w \notin G_{p_0}$.*

Proof. Suppose that $w \in G_{p_0}$. Then $w^{-1}v_0c_1v_1c_2 \cdots c_{n-1}v_{n-1}c_nv_n$ is a reduced word of type c with trivial image in $\mathcal{F}(\mathcal{G}, X)$ in contradiction to Theorem 2.22. $\square$

Corollary 2.25. *Let w be a reduced word of type c in the graph of groups $(\mathcal{G}, X)$ with c a closed path and let $T \subseteq X$ be a maximal tree of X. Then the image of w in the fundamental group $\pi_1(\mathcal{G}, X, T)$ is non-trivial.*

Proof. Define $o(c) = t(c) =: p_0$ to be the origin of the path c. The image of w is in the fundamental group $\pi_1(\mathcal{G}, X, p_0)$ and by Theorem 2.22 it is non-trivial. But since $\pi_1(\mathcal{G}, X, p_0)$ is isomorphic to the fundamental group $\pi_1(\mathcal{G}, X, T)$ for any maximal tree T by Corollary 2.21, the image of w has to be also non-trivial in $\pi_1(\mathcal{G}, X, T)$. □

Next, we need the following construction:

Construction 2.1. Let $(\mathcal{G}, X)$ be a graph of groups and let $X' \subseteq X$ be a connected, non-empty subgraph of X with its graph of groups $(\mathcal{G}|_{X'}, X')$ the restriction of $(\mathcal{G}, X)$. Assume that Theorem 2.22 is true for $(\mathcal{G}|_{X'}, X')$. Then Corollary 2.23 implies that the map

$$G_p \to \mathcal{F}(\mathcal{G}|_{X'}, X') \tag{16}$$

is a monomorphism for every $p \in V(X')$.
Next define $W = X/X'$ to be the graph derived by contraction of the subgraph X' in X to a vertex $x' \in V(W)$. Then we get for W

$$V(W) = (V(X) \setminus V(X')) \cup \{x'\} \quad \text{and} \quad E(W) = E(X) \setminus E(X')$$

and for $e \in E(W)$

$$o_w(e) = \begin{cases} o(e) & \text{if } o(e) \notin V(X') \\ x' & \text{otherwise} \end{cases} \quad \text{and} \quad t_w(e) = \begin{cases} t(e) & \text{if } t(e) \notin V(X') \\ x' & \text{otherwise} \end{cases}.$$

We want to define the graph of groups $(\mathcal{H}, W)$. The groups associated to the the vertices and edges are given by

$$\begin{aligned} H_p &= G_p && \text{if } p \in V(X) \setminus V(X') \\ H_p &= \mathcal{F}(\mathcal{G}|_{X'}, X') && \text{if } p = x' \\ H_e &= G_e && \text{for } e \in E(W) \end{aligned}$$

and the monomophisms $H_e \to H_{t(e)}$ are given by the ones in $(\mathcal{G}, X)$ whenever $t(e) \in V(W) \setminus \{x'\}$ and for $t(e) = x'$ we can define the monomorphism

$$H_e \to \mathcal{F}(\mathcal{G}|_{X'}, X')$$

as the composition of

$$H_e = G_e \to G_p \quad \text{and} \quad G_p \to \mathcal{F}(\mathcal{G}|_{X'}, X'),$$

where $p = t(e) \in V(X')$, the first map is the monomorphism given in $(\mathcal{G}, X)$ and the second is the monomorphism (16). Clearly, the projection $(\mathcal{G}, X) \to (\mathcal{H}, W)$ induces a homomorphism

$$\varphi : \mathcal{F}(\mathcal{G}, X) \to \mathcal{F}(\mathcal{H}, W). \tag{17}$$

Lemma 2.26. *The homomorphism $\varphi : \mathcal{F}(\mathcal{G}, X) \to \mathcal{F}(\mathcal{H}, W)$ in (17) of Construction 2.1 is in fact an isomorphism.*

Proof. This becomes apparent by simply considering the definition of the groups. □

Next, let us expand Construction 2.1 by associating to each word in $(\mathcal{G}, X)$ a word in $(\mathcal{H}, W)$ via the isomorphism from Lemma 2.26:

Construction 2.2. Let $c = c_1 \cdots c_n$ be a path in $E(X)$ with vertices $p_0, p_1, \dots, p_n$ in $V(X)$ and let $v = (v_0, v_1, v_2, \dots, v_n)$ be an $n+1$-tuple of elements $v_i \in G_{o(c_{i+1})}$ for every $i \in \{0, 1, \dots, n-1\}$ and $v_n \in G_{t(c_n)}$. Our aim is to define subwords of the word $w = v_0 c_1 v_1 c_2 \cdots c_{n-1} v_{n-1} c_n v_n$ of type c in X' and replace them by their images in $H_{x'}$.
So for any pair (i, j) with $0 \leq i \leq j \leq n$ define $w_{i,j}$ to be the subword of w that consists of the subpath $c_{i,j} = (p_i, c_i, p_{i+1}, c_{i+1}, \dots, c_{j-1}, p_j)$ of c and $v_{i,j} := (v_i, \dots, v_j)$. In case $c_{i,j}$ is contained in X', then let the image of $w_{i,j}$ in $H_{x'} = \mathcal{F}(\mathcal{G}|_{X'}, X')$ be denoted by $r_{i,j}$.
Now we define an increasing sequence of indices

$$0 \leq i_0 \leq j_0 < i_1 \leq j_1 < \cdots i_m \leq j_m \leq n,$$

satisfying the two conditions:

(i). The subpath c_{i_k, j_k} has to be fully contained in X' for every $0 \leq k \leq m$.

(ii). For every vertex $p_l \in V(X')$ of the path c that is contained in the graph X' there has to be a $k_l \in \{0, \dots, m\}$ such that p_l is an element of $c_{i_{k_l}, j_{k_l}}$.

Thus, the intermediate paths $c_{j_k, i_{k+1}}$ are of length $l(c_{j_k, i_{k+1}}) \geq 1$ for every $k \in \{0, \dots, m\}$ and apart from their origin and terminal vertex, no vertex is contained in $V(X')$. Therefore they are paths in W.
Lastly, we can define a word w' in $(\mathcal{H}, W)$ of type c', where c' is the path

$$c' = \dots, c_{j_{k-1}, i_k}, c_{j_k, i_k + 1}, \dots$$

given by

$$v' = (\dots, v_{j_{k-1}+1,i_k-1}, r_{i_k,j_k}, v_{j_k+1,i_{k+1}-1}, \dots),$$

with the convention that $v_{h,k}$ is empty whenever $h > k$.
If $i_0 = 0$, then c' starts with c_{j_0,i_1} and if $i_0 \neq 0$, it starts with c_{0,i_0}. Similarly we get that v' either starts with r_{i_0,j_0} if $i_0 = 0$ and with v_{0,i_0-1} if $i_0 \neq 0$. Analogously we can determine the "ends" depending on whether $j_m = n$.

To illustrate this rather technical Construction 2.2, we want to give an accessible example.

Example 2.17. Let (G, X) be a graph of groups. Suppose that $n = 4$ and we have a path $c = c_1c_2c_3c_4$ in $E(X)$ with vertices $\{p_0, p_1, p_2, p_3, p_4\} =: V_c \in V(X)$. Now let $X' \subseteq$ be a subgraph of X with $c \cap E(X') = \{c_2\}$ and $V_c \cap V(X') = \{p_0, p_1, p_2\}$. Then, following Construction 2.2, we obtain

$$c' = c_1, c_3, c_4 \qquad \text{and} \qquad v' = (v_0, v_1c_2v_2, v_3, v_4).$$

Clearly, w of type c, given by v in $\mathcal{F}(\mathcal{G}, X)$ corresponds to w' of type c' given by v' in $\mathcal{F}(\mathcal{H}, W)$ under the isomorphism (17) of Lemma 2.26.

Lemma 2.27. *[43, p. 48, Lemma 7] If w is a reduced word in $(\mathcal{G}, X)$ of type c and given by v, then w' defined in the Construction 2.2 is a word of type c' and given by v' is reduced in $(\mathcal{H}, W)$.*

This leads to the following lemma:

Lemma 2.28. *If Theorem 2.22 is true for the graph of groups $(\mathcal{H}, W)$ given by Construction 2.1, then it is also true for $(\mathcal{G}, X)$.*

Proof. This is a direct consequence of Lemma 2.26 and Lemma 2.27. □

Proof. (of Theorem 2.22) Let $(\mathcal{G}, X)$ be a graph of groups and let w be a reduced word of type $c = c_1c_2\cdots c_n$ given by $v = (v_0, v_1, \dots, v_n)$. We want to show that w is non-trivial in $\mathcal{F}(\mathcal{G}, X)$, as long as c and v are not both trivial. We have to distinguish four cases for X.

(i). Let X be a segment consisting of two vertices $p_{-1}, p_1 \in V(X)$ and one connecting edge $e \in E(X)$. Hence, w is of the form

$$w = v_0e^{k_1}v_1e^{k_2}\cdots e^{k_n}v_n,$$

with $k_i = \pm 1$ and $k_{i+1} = -k_i$, $v_0 \in G_{p_{-k_1}}$ and $v_i \in G_{p_{k_i}} \setminus G_e^{e^{k_i}}$. If $n = 0$, $w = v_0 \neq 1$, so w is non-trivial. If $n \geq 1$, we can consider the canonical projection p into the fundamental group, which is precisely $G_{p_{-1}} *_{G_e} G_{p_1}$

by Example 2.15 (i), i.e.

$$p : \mathcal{F}(\mathcal{G}, X) \to \pi_1(\mathcal{G}, X, X) = G_{p_{-1}} *_{G_e} G_{p_1} .$$

Since $p(w) = v_0 v_1 \cdots v_n$ and by Corollary 2.8 of the Normal Form Theorem 2.7, we know $v_0 v_1 \cdots v_n \neq 1$ and conclude that $w \neq 1$ as well.

(ii). Let X be a tree. With a standrad direct limit argument [43, c.f., Chapter 1], we assume without loss of generality that X is finite. Now, we want to use induction on $n = \frac{1}{2}|E(X)|$, half of the cardinality of the edge set (i.e. the cardinality of an orientation of X): If $n = 0$, the path c has to be trivial, so $w = v_0 \neq 1$. So assume $n \geq 1$. Define $X' \subseteq X$ to be a segment contained in X. By (i) we can apply Theorem 2.22 on X'. Next, we can consider the quotient $W = X/X'$, which is again a tree, with $|E(W)| = n - 1$. Hence, by the induction hypothesis, Theorem 2.22 applies to $(\mathcal{H}, W)$ and therefore by Lemma 2.28 also to $(\mathcal{G}, X)$.

(iii). Let X be a single vertex $p \in V(X)$ with one attached loop $e \in E(X)$. Then we know by Example 2.15 (iii) that the fundamental group is a semi-direct product, namely

$$\mathcal{F}(\mathcal{G}, X) \cong \pi_1(\mathcal{G}, X, p) = R \rtimes \langle g \rangle ,$$

where R is the normal subgroup $R = \langle\langle G_p \rangle\rangle$ and $\langle g \rangle$ is the infinite cyclic group generated by $g := g_e \notin R$. Moreover, we know R to be the amalgamated product of the G_n according to the homomorphisms

$$\begin{aligned} A &\to G_{n-1} & \quad\text{and}\quad & & A &\to G_n \\ a &\mapsto g^{n-1} a^{\bar{e}} g^{1-n} & & & a &\mapsto g^n a^e g^{-n} . \end{aligned} \tag{18}$$

Let us now consider a reduced word $w = v_0 e^{k_1} v_1 e^{k_2} \cdots e^{k_n} v_n$ in $\mathcal{F}(\mathcal{G}, X)$ of type $c = e^{k_1} e^{k_2} \cdots e^{k_n}$ with $k_i = \pm 1$ and given by $v = (v_0, v_1, \ldots, v_n)$ with $v_0 \in G_0$ and $v_i \notin A^{e^{k_i}}$ if $k_{i+1} = -k_i$. Since in this case the projection of $\mathcal{F}(\mathcal{G}, X)$ into $\pi_1(\mathcal{G}, X, p)$ is an isomorphism, the image of w is simply $v_0 g^{k_1} v_1 g^{k_2} \ldots g^{k_n} v_n$. Thus, we can distinguish two cases: If $\sum_{i=1}^{n} k_i \neq 0$, then $w \notin R$, hence $w \neq 1$. So we can assume $\sum_{i=1}^{n} k_i = 0$. Define

$$d_0 = d_n = 0; \quad d_j := \sum_{i=1}^{j} k_i = k_1 + \cdots + k_j \qquad \text{and} \qquad s_j := g^{d_j} v_j g^{-d_j} ,$$

to obtain

$$w = s_0 s_1 \cdots s_n \qquad \text{with} \qquad s_j \in G_{d_j}\,,$$

$$d_{j+1} - d_j = k_{j+1} = \pm 1 \qquad \text{and} \qquad s_j \notin g^{d_j} A^{e^{k_j}} g^{-d_j} \;\; \text{if } d_{j+1} = d_{j-1}\,.$$

Next, let $T = Cay(\mathbb{Z}, \{1\})$ be the Cayley graph of $\mathbb{Z}$ and let $(\mathcal{K}, T)$ be the tree of groups associated to it, where to every vertex $n \in \mathbb{Z}$ we assign the group G_n and to every edge we assign the group A with the homomorphisms given in (18). Then we have that $R = \pi_1(\mathcal{K}, T, T)$ and $s_0 s_1 \cdots s_n$ is associated with a reduced word of $(\mathcal{K}, T)$ whose type is a closed path since $d_0 = d_n = 0$. But since T is a tree, by (ii) we can apply Theorem 2.22 and thus also Corollary 2.25 to it. That implies that $s_0 s_1 \cdots s_n \neq 1$, but since $w = s_0 s_1 \cdots s_n$ also $w \neq 1$.

(iv). Let X be an arbitrary graph. As in part (i) we may assume that X is finite and use indcution on $n = \frac{1}{2}|E(X)|$, half of the cardinality of the edges in X. If $n = 0$, the path c has to be trivial and $w = v_0 \neq 1$. So assume $n \geq 1$. Define $X' \subseteq X$ either as a segment or a loop in X. In both cases, Theorem 2.22 is applicable to $(\mathcal{G}|_{X'}, X')$ by (i) and (iii). Also, by the induction hypothesis, it is applicable to $(\mathcal{H}, W)$ for $W = X/X'$. Thus, by Lemma 2.28, Theorem 2.22 holds also for $(\mathcal{G}, X)$.

□

Remark 2.13. Case (iii) of the Theorem 2.22 dealing with HNN-extensions is equivalent to the statement of Corollary 2.17 and often referred to as *Britton's Lemma.*

2.3.4.2 Universal Covering

Remark 2.14. In the following sections, let $(\mathcal{G}, X)$ be always a non-empty, connected graph of groups, with T a maximal spanning tree and A an orientation of X.

Definition 2.25. Let X be a non-empty connected graph with orientation A and let $e \in E(X)$. Then we define the *sign* of an edge $e \in E(X)$ by the following map

$$s : E(X) \to \mathbb{Z}/2\mathbb{Z}$$
$$e \mapsto \begin{cases} 0 & \text{if } e \in A \\ 1 & \text{if } e \notin A \end{cases}$$

and the *absolute value* of an edge $e \in E(X)$ by

$$|e| := \begin{cases} e & \text{if } e \in A \\ \bar{e} & \text{if } e \notin A \Leftrightarrow \bar{e} \in A\,. \end{cases}$$

It is our aim to construct a graph $\tilde{X} = \tilde{X}(\mathcal{G}, X, T)$ on which $\pi := \pi_1(\mathcal{G}, X, T)$ acts and a morphism

$$p : \tilde{X} \to X$$

inducing an isomorphism

$${}_{\pi}\backslash^{\tilde{X}} \to X$$

with right inverses

$$\begin{aligned} V(X) &\to V(\tilde{X}) & \qquad E(X) &\to E(\tilde{X}) \\ p &\mapsto \tilde{p} & e &\mapsto \tilde{e}\,. \end{aligned} \tag{19}$$

Furthermore we have the following two requirements:

- If $p \in V(X)$ we want that the stabilizer $\pi_{\tilde{p}} \subseteq \pi$ of $\tilde{p}$ in π equals G_p, the stabilizer of p in G.
- If $e \in E(X)$, we define $w := |\bar{e}|$ and want that the stabilizer $\pi_{\tilde{e}} \subseteq \pi$ of $\tilde{e}$ in π equals the subgroup $G_w^w \leq G_{t(w)}$.

To achieve these conditions, we simply define $\tilde{X}$ in the following way:

Definition 2.26. Let $\tilde{X}$ be graph with vertex set the disjoint union of the vertex orbits $\pi \cdot \tilde{p} \cong \pi/\pi_{\tilde{p}}$ for $p \in V(X)$, i.e.

$$V(\tilde{X}) = \bigsqcup_{p \in V(X)} \pi/\pi_{\tilde{p}}$$

and edge set the disjoint union of the edge stabilizers $\pi \cdot \tilde{e} \cong \pi/\pi_{\tilde{e}}$ for $e \in E(X)$, i.e.

$$E(\tilde{X}) = \bigsqcup_{e \in E(X)} \pi/\pi_{\tilde{e}}$$

where

$$\pi_{\tilde{p}} = G_p \qquad \text{and} \qquad \pi_{\tilde{e}} = G_w^w \text{ for } w = |\bar{e}|$$

and $\tilde{p}$ is the image of 1 in $\pi/\pi_{\tilde{p}}$ and $\tilde{e}$ is the image of 1 in $\pi/\pi_{\tilde{e}}$. We call $\tilde{X}(\mathcal{G}, X, T) = \tilde{X} = \tilde{X}(V(\tilde{X}), E(\tilde{X}))$ the *universal covering* of X.

Next, we need to figure out how π acts on $\tilde{X}$ such that ${}_{\pi}\backslash^{\tilde{X}} = X$:

Lemma 2.29. *[43, p. 51] The map*

$$\begin{aligned}
\pi \times \tilde{X} &\to \tilde{X} \\
(g, \tilde{e}) &\mapsto g\tilde{e} \\
(g, \tilde{\bar{e}}) &\mapsto g\tilde{\bar{e}} = \overline{g\tilde{e}} \\
(g, o(\tilde{e})) &\mapsto o(g\tilde{e}) = gg_e^{-s(e)}\tilde{o}(e) \\
(g, t(\tilde{e})) &\mapsto t(g\tilde{e}) = gg_e^{1-s(e)}\tilde{t}(e)
\end{aligned}$$

is a well-defined action of π on the universal covering $\tilde{X}$.

Remark 2.15. So now we have found the graph $\tilde{X}$ with an action of π on it such that by construction ${}_{\pi}\backslash^{\tilde{X}} = X$. Moreover if $e \in E(T) \subseteq E(X)$, then $g_e = 1$ and so

$$\begin{aligned}
o(\tilde{e}) = o(1 \cdot \tilde{e}) = 1 \cdot g_e^{-s(e)}\tilde{o}(e) = \tilde{o}(e) \\
t(\tilde{e}) = t(1 \cdot \tilde{e}) = 1 \cdot g_e^{1-s(e)}\tilde{t}(e) = \tilde{t}(e)\,,
\end{aligned}$$

which implies that the maps (19) define a lifting $T \xrightarrow{\sim} \tilde{T} \subseteq \tilde{X}$ of T into $\tilde{X}$.

Theorem 2.30. *Let $(\mathcal{G}, X)$ be a connected, non-empty graph of groups with maximal tree T and an orientation A. Then the universal covering $\tilde{X} = \tilde{X}(\mathcal{G}, X, T)$ is a tree.*

Proof. First show that $\tilde{X}$ is connected. Let W be the smallest subgraph of $\tilde{X}$ that contains all $\tilde{e}, e \in E(Y)$. Let now $e \in E(Y)$ be an edge. By Remark 2.15 we know since T is the maximal subtree of Y that either $o(\tilde{e}) = \tilde{o}(e)$ in case $s(e) = 0$ or $t(\tilde{e}) = \tilde{t}(e)$ in case $s(e) = 1$. Either way, at least one of the extremities of $\tilde{e}$ belongs to the lift $\tilde{T}$ of T. Thus W is connected and $\pi \cdot W = \tilde{X}$. That means that it suffices to show that there is a generating set S of π such that $W \cup sW$ is connected for all $s \in S$. But by Definition 2.23 we know that $\pi = \pi_1(\mathcal{G}, X, T)$ is generated by the set

$$S = \{G_{\tilde{p}} \mid p \in V(X)\} \cup \{g_e \mid e \in E(X)\}\,.$$

So let first $a \in \{G_{\tilde{p}} \mid p \in V(X)\} \subseteq S$: Then $a \in G_{\tilde{q}}$ for some $\tilde{q} \in V(\tilde{X})$, but $\tilde{q} \in W$ as well, so aW and W share the vertex $\tilde{q}$, so $aW \cup W$ has to be connected. Let now $a \in \{g_e \mid e \in E(X)\} \subseteq S$: Then $a = g_f$ for some $f \in E(X)$. If $s(f) = 0$, then $o(\tilde{f}) = o(g_f\tilde{f}) = o(a\tilde{f}) =: o$ and W and aW share this vertex o. If $s(f) = 1$, then $t(\tilde{f}) = t(g_f\tilde{f}) = t(a\tilde{f}) =: t$ and W and aW share this vertex t. Either way, $W \cup aW$ is connected.

Secondly, we need to make sure that $\tilde{X}$ does not contain any circuit. Suppose that on the contrary $\tilde{c} = s_1\tilde{e}_1 \cdots s_n\tilde{e}_n$ is a closed path of length $n > 0$ in $\tilde{X}$ without backtracking and with edges $s_i\tilde{e}_i \in E(\tilde{X})$ where $s_i \in S$ and $e_i \in E(X)$ for any $i \in \{1, 2, \dots, n\}$. Let c be the projection of $\tilde{c}$ in X with vertices $(p_0, \dots, p_n)$ with $p_0 = p_n$. To make the notation more accessible, define $s(e_i) =: k_i$ and $g_{e_i} =: g_i$ and observe

$$\begin{aligned} t(s_n\tilde{e}_n) &= s_n g_n^{1-k_n}\tilde{p}_0 = s_1 g_1^{-k_1}\tilde{p}_0 = o(s_1\tilde{e}_1) \\ t(s_1\tilde{e}_1) &= s_1 g_1^{1-k_1}\tilde{p}_1 = s_2 g_2^{-k_2}\tilde{p}_1 = o(s_2\tilde{e}_2) \\ &\vdots \\ t(s_{n-1}\tilde{e}_{n-1}) &= s_{n-1} g_{n-1}^{1-k_{n-1}}\tilde{p}_{n-1} = s_n g_n^{-k_n}\tilde{p}_{n-1} = o(s_n\tilde{e}_n)\,. \end{aligned}$$

Now defining $q_i \coloneqq s_i g_i^{-k_i}$ gives for $r_i \in G_{p_i} = \pi_{\tilde{p}_i}$

$$\begin{aligned} q_n g_n r_n &= q_1 \\ q_1 g_1 r_1 &= q_2 \\ &\vdots \\ q_{n-1} g_{n-1} r_{n-1} &= q_n\,. \end{aligned}$$

Thus we obtain

$$\begin{aligned} g_1 r_1 &= q_1^{-1} q_2 \\ g_2 r_2 &= q_2^{-1} q_3 \\ &\vdots \\ g_n r_n &= q_n^{-1} q_1 \end{aligned}$$

and multiplying gives us

$$g_1 r_1 g_2 r_2 \cdots g_n r_n = 1\,. \tag{20}$$

Let now $w = s_1\tilde{e}_1 r_1 s_2\tilde{e}_2 r_2 \dots s_n\tilde{e}_n r_n$ be a word of type c for $(1, r_1, \dots, r_n)$. If we can show that w is reduced, we are done, since then its image $g_1 r_1 g_2 r_2 \cdots g_n r_n$ cannot be equal to 1 by Corollary 2.25 of Theorem 2.22, since c is a closed path. So suppose that w is not reduced, i.e. $e_{i+1} = \bar{e}_i$ for some $i \in \{1, 2, \dots, n-1\}$. That implies $g_{i+1} = g_i^{-1}$ and $k_{i+1} = 1 - k_i$. Using the formula

$$s_i g_i^{-k_i} g_i r_i = q_i g_i r_i = q_{i+1} = s_{i+1} g_{i+1}^{-k_{i+1}}$$

gives

$$r_i = g_i^{k_i - 1} s_i^{-1} s_{i+1} g_i^{1-k_i}\,.$$

By assumption $r_i \in G_{e_i}^{e_i}$, which is equivalent to $s_i^{-1} s_{i+1} \in g_i^{1-k_i}\, G_{e_i}^{e_i} g_i^{k_i - 1} = \pi_{\tilde{e_i}}$. But $\overline{s_{i+1}\tilde{e}_{i+1}} = s_{i+1}\tilde{e}_i \neq s_i \tilde{e}_i$ since c is not backtracking, thus $s_i \pi_{\tilde{e}} \neq s_{i+1}\pi_{\tilde{e_i}}$, in contradiction to $s_i^{-1} s_{i+1} \in \pi_{\tilde{e_i}}$. Therefore w cannot be a reduced word, contradicting our assumption. Hence, w is a reduced word of type c and the proof is done. □

Remark 2.16. If the universal covering is a tree, it is often called *Bass-Serre tree*.

Example 2.18.

(i). If all the groups $G_p = \{1\}$ associated to the vertices $p \in V(X)$ in X are trivial, then $\pi = \pi_1(Y, T)$ and $\tilde{X}$ is the universal covering relative to T in the usual sense known from algebraic topology.

(ii). Let the quotient Y be a segment consisting of two vertices p and q connected by one edge e. Then we know by Theorem 2.13 that the fundamental group $\pi = \pi_1(\mathcal{G}, Y, T) = G_p *_{G_e} G_q$ where G_p, G_q and G_e are the vertex respectively edge stabilizers. Hence we obtain for the universal covering $\tilde{X}$ the tree associated to it.

2.3.4.3 Structure Theorem

Our aim is now to construct a graph of groups $(\mathcal{G}, Y)$ where Y is the quotient of the connected, non-empty graph X by a group action without inversions of G and where its fundamental group is isomorphic to G.

Construction 2.3. Let T be a maximal tree of Y and let $j' : T \to X$ be a lifting, that exists by Lemma 2.4. Let A be an orientation of Y with associated sign map s. We want to extend this lifting j' to a map

$$\begin{aligned} j : E(Y) &\to E(X) \\ e &\mapsto j(e) \end{aligned}$$

with the properties $\forall\, e \in E(Y)$

$$j(\overline{e}) = \overline{j(e)} \tag{21}$$

$$o(j(e)) = j(o(e))\,. \tag{22}$$

It suffices to define the element $j(e)$ for every edge $e \in A \setminus E(T)$, since A is an orientation and $j|_{E(T)} = j'$. To satisfy property (22), we need to choose $j(e)$ such that the origin

$$o(j(e)) \in V(j(T)) = V(j'(T))\,.$$

For the terminal vertex we observe that $t(j(e))$ and $j(t(e))$ need to have the same projection $t(e) \in V(Y)$ for any $e \in E(Y)$. Hence they have to lie in the same orbit and we can choose an element

$$\gamma_e \begin{cases} \in G & \text{if } e \in A \setminus E(T) \\ =: 1 \in G & \text{if } e \in E(T) \end{cases}$$

with

$$\gamma_{\bar{e}} = \gamma_e^{-1}$$

such that for any $e \in E(Y)$

$$t(j(e)) = \gamma_e \, j(t(e))\,.$$

So we obtain

$$\begin{aligned} o(j(e)) &= \gamma_e^{-s(e)} \, j(o(e)) \\ t(j(e)) &= \gamma_e^{1-s(e)} \, j(t(e))\,. \end{aligned}$$

Next, we attach to every vertex $p \in V(Y)$ the vertex stabilizer $G_{j(p)}$ of $j(p) \in V(X)$ in G and to every edge $e \in E(Y)$ the edge stabilizer $G_{j(e)}$ of $j(e) \in V(X)$ in G. The graph of groups is completed by the monomorphisms

$$\begin{aligned} G_e &\hookrightarrow G_{t(e)} \\ a &\mapsto a^e =: \gamma_e^{s(e)-1} \, a \, \gamma_e^{1-s(e)}\,, \end{aligned}$$

which are well defined since $\gamma_e^{s(e)-1} \, G_{j(e)} \, \gamma_e^{1-s(e)} \subseteq G_{j(t(e))}$.

Now we are in the position to define two (homo-)morphims, which will then give us the isomorphism of G and the graph of groups $(\mathcal{G}, Y)$ we are looking for:

Definition 2.27. Define ϕ to be the homomorphism given by

$$\begin{aligned} \phi : \pi_1(\mathcal{G}, Y, T) &\to G && \\ G_p &\hookrightarrow G && \forall\, p \in V(Y) \\ g_e &\mapsto \gamma_e && \forall\, e \in E(Y) \end{aligned} \tag{23}$$

and ψ to be the graph morphism given by

$$\begin{aligned} \psi : \tilde{X}(\mathcal{G}, Y, T) &\to X && \\ g\tilde{p} &\mapsto \phi(g) \, j(p) && \forall\, \tilde{p} \in V(\tilde{X}(\mathcal{G}, Y, T)) \\ g\tilde{e} &\mapsto \phi(g) \, j(e) && \forall\, \tilde{e} \in E(\tilde{X}(\mathcal{G}, Y, T))\,. \end{aligned} \tag{24}$$

We want to check some properties of ϕ and ψ:

Definition 2.28. A homomorphism $f : \tilde{X} \to X$ is said to be locally injective if it is injective on every set of edges with a given common origin.

Lemma 2.31. *ϕ is surjective and ψ is surjective and locally injective.*

Proof. Let W be the smallest subgraph of X containing $j(e)$ for every $e \in E(Y)$. Then by our construction of j, we know $o(j(e)) \in V(j(T))$ for any $e \in A$, so at least one vertex of the egde $e \in E(W)$ is in $j(T)$ and we have $G \cdot W = X$. Furthermore, we defined $\gamma_e \in G$ such that $\gamma_e t(e) \in V(T)$. Therefore we can apply Lemma 2.14 and conclude that G is generated by the set

$$\{G_p \mid p \in V(Y)\} \cup \{\gamma_e \mid e \in E(Y)\},$$

which means that ϕ is surjective.
Also, by the definition of ψ, we know that $\psi(1 \cdot \tilde{y}) = \phi(1)j(y) = j(y)$, hence $W \subseteq \psi(\tilde{X})$, which implies with the surjectivity of ϕ that ψ has to be surjective as well. Moreover, ϕ induces isomorphisms between the stabilizers of corresponding vertices and edges in $\tilde{X}$ and X, therefore ψ is locally injective. $\square$

The following result is often referred to as the *Fundamental Theorem of Bass-Serre Threory*:

Theorem 2.32 (Structure Theorem)**.** *Let G be a group acting without inversions on a connected, non-empty graph X. Let $Y := {}_G\backslash^X$ denote the quotient of X be the action on G. Let $(\mathcal{G}, Y)$ be the graph of groups given by the Construction 2.3. Then the following three statements are equivalent:*

(i). X is a tree.

(ii). The map $\psi : \tilde{X} \to X$ is an isomorphism.

(iii). The map $\phi : \pi_1(\mathcal{G}, Y, T) \to G$ is an isomorphism.

Proof.

- $(i) \Rightarrow (ii)$: We already know by Lemma 2.31 that ψ is a locally injective, surjective homomorphism. We need to show that it is also injective if X is a tree. So let c be an injective path in $\tilde{X}$. Since X is a tree, geodesics are unique, thus it suffices to check that $\psi \circ c$ has no backtracking. But that follows directly from the fact that c is injecitve and ψ is locally injective.

- $(ii) \Rightarrow (i)$: Let ψ be an isomorphism. Then X has to be a tree, since $\tilde{X}$ is a tree by Theorem 2.30.

- $(ii) \Rightarrow (iii)$: Assume that ψ is bijective. Since we know by Lemma 2.31 that ϕ is surjective anyways, we are left to show injectivity. Let $N := ker(\phi)$ be the kernel of the homomorphism ϕ and let $p \in V(Y)$. Since ϕ defines an isomorphism of G_p onto $G_{j(p)}$, we have trivial intersection $N \cap G_p = \{1_{\pi_1(\mathcal{G},Y,T)}\}$. Hence no non-trivial element in N can fix the vertex $\tilde{p}$ in $\tilde{X}$, since $G_p = \pi_{\tilde{p}}$. That means that for $n \in N \setminus \{1_{\pi_1(\mathcal{G},Y,T)}\}$, the two vertices $\tilde{p} \in V(\tilde{X})$ and $n\tilde{p} \in V(\tilde{X})$ are distinct in $\tilde{X}$. But since

  $$\psi(n\tilde{p}) = \phi(n)j(p) = 1 \cdot j(p) = j(p) = \psi(\tilde{p}),$$

  they have the same image $j(p)$ in X under ψ. So if ψ is an isomorphism, in particular if ψ is injective, we conclude $n\tilde{p} = \tilde{p}$. But that means $n = 1_{\pi_1(\mathcal{G},Y,T)}$, so ϕ is injective as well.

- $(iii) \Rightarrow (ii)$: If ϕ is an isomorphism, then so is ψ per Definition.

□

Remark 2.17. The most important implication of the Structure Theorem 2.32 is $(i) \Rightarrow (iii)$: It tells us that if X is a tree, the group G acting on it is generated by the vertex stabilizers G_p for $p \in V(Y)$ and the elements γ_e for $e \in E(Y)$ subject to the relations of the G_p and

$$\gamma_e a^e \gamma_e^{-1} = a^{\bar{e}}, \quad \gamma_{\bar{e}} = \gamma_e^{-1} \qquad \text{and} \qquad \gamma_e = 1 \text{ if } e \in E(T).$$

Let us revisit Theorem 2.18 as a reformulation into a Corollary of the Structure Theorem 2.32:

Corollary 2.33. *Let G be a group acting on a graph X with the quotient $Y := {}_G\backslash^X$ being a loop. Let $F \subseteq X$ be a segment in X consisting of two vertices p and q connected by the edge f with stabilizers G_p, G_q and G_f. Since Y is a loop, there is an element $\gamma_f \in G$ such that $\gamma_f p = q$. Define $G_f' := \gamma_f^{-1} G_e \gamma_f$ with $\phi : G_f \to G_f'$ the homomorphism induced by conjugation. Then X is a tree if and only if $G_f' \leq G_p$ and the map*

$$\begin{aligned} \psi : G_p *_\phi = \langle G_p \cup \{t\} \mid R_p \cup \{t^{-1}at = \phi(a) \mid a \in G_f\}\rangle &\to G \\ g &\mapsto g \qquad \forall\, g \in G_p \\ t &\mapsto x \end{aligned}$$

is an isomorphism. In particular, if X is a tree, then G is an HNN-extension.

Proof. First, let X be a tree. Then we know by the Structure Theorem 2.32 that G has to be isomorphic to $\pi_1(\mathcal{G}, Y, T)$. But Example 2.15 (iii) implies

that $\pi_1(\mathcal{G}, Y, T)$ is precisely the HNN-extension $G_p *_\phi$. Furthermore we know by Construction 2.3 that the only (positively oriented) edge $e \in E(Y)$ has to be the preimage of $f \in V(X)$ under the map j, and the only vertex $v \in V(Y)$ has to be the preimage of $p \in V(X)$. With the definition of the graph of groups $(\mathcal{G}, Y)$ we conclude that $\gamma_f^{s(e)-1} G_{j(e)} \gamma_f^{1-s(e)} \leq G_{j(t(e)}$ has to hold, hence $G'_f = \gamma_f^{s(e)-1} G_f \gamma_f^{1-s(e)} \leq G_p$, which gives the second claim of the necessary condition.
For the converse we can use the same argument: Since $G'_f \leq G_p$, the graph of groups $(\mathcal{G}, Y)$ of Construction 2.3 is well defined. Hence, we can apply the Structure Theorem 2.32, which implies since the map $\phi : \pi_1(\mathcal{G}, Y, T) \to G$ is an isomorphism that X is a tree. $\square$

Example 2.19. Let $G = \langle a, b \mid a^2 = b^3 \rangle$ be a group and ϕ a homomorphism into the symmetric group $S_3 = \langle d, s \mid s^2 = d^3 = 1, dsds = 1 \rangle$ of a triangle, given by

$$\begin{aligned} \phi : G &\to S_3 \\ a &\mapsto s \\ b &\mapsto d \,. \end{aligned}$$

Define $H = ker(\phi) \leq G$ to be the kernel of ϕ. We aim to find a presentation of H in form of a fundamental group of a graph of groups. Therefore observe first that $G = \langle a \rangle *_{\langle a^2 \rangle \cong \langle b^3 \rangle} \langle b \rangle$. By Theorem 2.13 there is a tree X with fundamental domain a segment consisting of two vertices a and b connected with an edge e with stabilizers $G_a = \langle a \rangle$, $G_b = \langle b \rangle$ and $G_e = \langle a^2 \rangle \cong \langle b^3 \rangle$. As H is a subgroup of G, it acts on X by left multiplication as well. Now we want to determine the factor graph $Y := {}_H\backslash^X$, since then the Structure Theorem 2.32 implies that H is isomorphic to the fundamental group $\pi_1(\mathcal{H}, Y, T)$. Therefore consider the system of left coset representatives of H in G given by $\{1, b, b^2, a, ba, b^2a\}$. Then any vertex $g\langle a \rangle \in V(X)$ can be either in the same H-orbit as $\langle a \rangle$, $b\langle a \rangle$ or $b^2\langle a \rangle$. If we consider the system of left coset representatives of H in G given by $\{1, b, b^2, ab, ab^2\}$, which we obtain from the previous one by the relations that have to hold in H

$$\begin{aligned} baba = 1 \quad &\Leftrightarrow \quad ab = b^{-1}a^{-1} \quad \Leftrightarrow \quad ab = b^2a \\ baba = 1 \quad &\Leftrightarrow \quad ba = a^{-1}b^{-1} \quad \Leftrightarrow \quad ba = ab^2 \,, \end{aligned}$$

then we conclude that any vertex $g\langle b \rangle$ is in the same H-orbit as $\langle b \rangle$ or $a\langle b \rangle$. And since $\langle a \rangle$, $b\langle a \rangle$, $b^2\langle a \rangle$, $\langle b \rangle$ and $a\langle b \rangle$ are not in the same H-orbit, we can conclude that there are precisely 5 H-orbits, hence Y consists of 5 vertices, let us call them A, D, E, B and C respectively. As the system of left coset representatives of H

in G contains 6 elements, we have 6 H-orbits for the edges, whose representatives we denote as the tree $\tilde{Y}$. Hence Y has 6 edges. Since the relations that have to hold in H imply that

$$b\langle a\rangle = 1 \cdot b\langle a\rangle = (baba)^{-1} b\langle a\rangle = a^{-1}b^{-1}a^{-1}b^{-1}b\langle a\rangle = ab^2a\langle a\rangle = ab^2\langle a\rangle$$
$$\text{and} \quad b^2\langle a\rangle = 1b^2\langle a\rangle = ababb^2\langle a\rangle = abab^3\langle a\rangle = aba\langle a\rangle = ab\langle a\rangle\,,$$

we can deduce that the vertex E has to be connected to the vertices B and C. Furthermore we have the obvious edges (B, A), (A, C), (B, D) and (B, E) in $E(Y)$. Now let T be a maximal tree of Y, given by all vertices in $V(Y)$ and all the edges except for (D, C) and (E, C) in $E(Y)$. Let $\tilde{T} \subseteq X$ be the lift of T given by the map $B \mapsto \langle b\rangle$, $D \mapsto b\langle a\rangle$, $A \mapsto \langle a\rangle$ and $E \mapsto b^2\langle a\rangle$.
Next, we want to compute the vertex and edge stabilizers in H. Therefore observe that $g\langle a\rangle g^{-1} \leq G$ is the vertex stabilizer in G for the vertex $g\langle a\rangle$, since $g\langle a\rangle g^{-1}g\langle a\rangle = g\langle a\rangle$. In a similar manner we obtain for the edge stabilizer $g\langle a^2\rangle g^{-1} \leq G$. Intersecting those with H gives us for the vertex as well as for the edge stabilizers $g\langle a^2\rangle g^{-1}$ for any $g \in G$. But by the relation $b^3 = a^2$, we obtain that any power of a^2 commutes with any $g \in G$. Hence $g\langle a^2\rangle g^{-1} = gg^{-1}\langle a^2\rangle = \langle a^2\rangle$, which means that $\langle a^2\rangle$ is the vertex as well as the edge stabilizer for any edge $g\langle a^2\rangle$ and any vertex $g\langle a\rangle$ or $g\langle b\rangle$. Hence for the graph of groups $(\mathcal{H}, Y)$ we attach to every vertex and every edge the group $\langle a^2\rangle$ and the edge maps are simply given by the identity map. Thus, we have constructed a graph of group $(\mathcal{H}, Y)$, whose fundamental group $\pi_1(\mathcal{H}, Y, T)$ with respect to the maximal tree T is isomorphic to H. So to obtain a presentation for H, we simply need to find the presentation for $\pi_1(\mathcal{H}, Y, T)$: As we only have one stabilizer $\langle a^2\rangle$, we have only one generator from them, namely $a^2 := x$ and no relations. Furthermore, there are two (positively oriented) edges, that are not in the normal closure, that is quotient out, namely (D, C) and (E, C) in $E(Y)$. Let their corresponding generators be denoted by t_1 and t_2. They have to be subject to the relations $t_1^{-1}xt_1 = x$ and $t_2^{-1}xt_2 = x$, since all the edge maps in $(\mathcal{H}, Y)$ are identity maps. Hence we obtain for H the presentation

$$H = \langle x, t_1, t_2 \mid t_1^{-1}xt_1 = x, t_2^{-1}xt_2 = x\rangle\,.$$

2.3.4.4 Property FA

The fact that Bass-Serre theory can come in very handy with algebraic properties of groups should be first illustrated by the following application:

Definition 2.29. Let G be a group acting on a tree T. Let

$$T^G := \{v \in V(T) \mid gv = v \quad \forall\, g \in G\} \cup \{e \in E(T) \mid ge = e \quad \forall\, g \in G\}$$

be the set of all elements in T that are fixed by G. We say that G has *Property FA* if $T^G \neq \emptyset$ for every tree T.

Proposition 2.34. *Let G be a countable group. Then G has Property FA if and only if the following three statements hold*

(i). G is finitely generated.

(ii). G is not an amalgamated product.

(iii). G has no quotient that is isomorphic to $\mathbb{Z}$.

Proof. First suppose G has Property FA and show that this implies all three statements:

(i). Suppose that G is not finitely generated. Let $G_1 \subseteq G_2 \subseteq \dots$ be a nested family of finitely generated groups and let G be their union, i.e.

$$G = \bigcup_{i \in \mathbb{N}} G_i \,.$$

Construct a tree T by defining $V(T) = \bigsqcup_{i \in \mathbb{N}} G/G_i$ and let there be an edge between gG_i and gG_{i+1} for all $g \in G$ and $i \in \mathbb{N}$. Then G acts on T, so by Property FA the set T^G is not empty, so there is a $gG_p \in V(T)$ for some $g \in G$ and for some $p \in \mathbb{N}$ that is fixed by all elements in G. But that implies that $G = G_p$, so G has to be finitely generated after all.

(ii). Suppose that G is an amalgamated product $G = G_1 *_A G_2$. Then by Theorem 2.13 G acts on a tree T with fundamental domain a segment consisting of two vertices P and Q connected by one edge e with vertex stabilizers $G_P \lneq G$ and $G_Q \lneq G$ proper subgroups of G. But that means that each vertex of T can only be stabilized by a proper subgroup of G and not G itself. Therefore $X^G = \emptyset$ in contradiction to our assumption of Property FA.

(iii). Suppose that there is an epimorphism $G \twoheadrightarrow \mathbb{Z}$. Then G acts on the Cayley graph of $\mathbb{Z}$, which is just the integer line, by translation. Thus, there cannot be any fixed point, i.e. $X^G = \emptyset$ in contradiction to the assumption of Property FA.

Now assume conversely that all three statements hold and show that G has to have Property FA. Let G act on a tree T and define the quotient $Y \coloneqq {}_G\backslash^X$. Then by Theorem 2.32 G is isomorphic to the fundamental group of the associated graph of groups $\pi_1(\mathcal{G}, Y, T)$. There is an epimorphism $\pi_1(\mathcal{G}, Y, T) \twoheadrightarrow \pi_1(Y) \cong F_S$ for some generating set S. But since any free group has $\mathbb{Z}$ as a subgroup, (iii).

implies that $F_S = \{1_{F_S}\}$. So Y has to be just a tree, in particular Y has no loops. Now Example 2.15 (ii) implies that $\pi_1(\mathcal{G}, Y, T)$ can be written as an amalgamated product

$$\pi_1(\mathcal{G}, Y, T) = *_{G_e}^{i \in I} G_i$$

along the edge stabilizers $G_e \in E(T)$, where G_i is a vertex stabilizer for some $p_i \in V(T)$ and $I \subseteq \mathbb{N}$ is a finite set by property (i). . But since (ii). states that G cannot be an amalgamated product and $G \cong \pi_1(\mathcal{G}, Y, T) = *_{G_e}^{i \in I} G_i$, we can conclude that $G = G_p$ for one $p \in V(T)$, which means that $X^G \supseteq G_p \neq \emptyset$. So G has Property FA. □

Example 2.20. Let G be a group with Property FA. If G is contained in an amalgamated product $G_1 *_A G_2$, then by Theorem 2.34 G has to be contained in a conjugate of G_1 or G_2.

3 Well-Studied Examples

3.1 Free Groups

The proofs and approaches in this chapter are guided by [42] and [48].

3.1.1 Hyperbolicity

Remark 3.1. As already pointed out in Example 1.5, any free group is hyperbolic, since its Cayley graph with respect to a free generating set is always a locally finite tree.

3.1.2 Residual Finiteness

Proposition 3.1. *Free groups are residually finite.*

We will explore three different ways to prove that fact:

Proof. (A) [of 3.1] Let F be a free group of rank $n \in \mathbb{N}$ generated by S. Let $x \coloneqq x_1 x_2 \cdots x_k \in F$. We want to find a homomorphism $\phi : F \to S_{n+1}$ such that $\phi(x) \neq 1_{S_{n+1}}$ since then F is residually finite.

So define

$$f : S \to S_{n+1}$$
$$a \mapsto \begin{cases} id & \text{if } a \notin \{x_1^{\pm 1}, x_2^{\pm 1}, \dots x_k^{\pm 1}\} \\ \begin{array}{l}\text{some choice of}\\ \text{permutation}\\ \text{sending}\end{array} \begin{cases} i \mapsto i+1 & \text{if } a = x_i \\ i+1 \mapsto i & \text{if } a = x_i^{-1} \end{cases} & \text{if } a \in \{x_1^{\pm 1}, x_2^{\pm 1}, \dots x_k^{\pm 1}\} \end{cases}$$

By Definition 1.1, f extends uniquely to a homomorphism $\phi : F \to S_{n+1}$ with $\phi(x) = f(x_1)\, f(x_2) \cdots f(x_n)$. But since $f(x_1)$ maps $1 \mapsto 2$, $f(x_2)$ maps 2 to 3 and so on up to $f(x_n)$, which maps n to $n+1$, we get that $\phi(x)$ maps 1 to $n+1$ and is therefore not the identity map. $\square$

Next, we want to use the topological interpretation of residual finiteness. Therefore we want to establish some topological properties of free groups.

Proposition 3.2. *Let X be a rose with $|S| < \infty$ petals, that is the wedge of $|S|$ copies of the unit circle S^1 indexed by S. Then $\pi_1(X)$, the fundamental group of X, is isomorphic to F_S.*

Proof. We will use induction on $|S| =: n$:

- Let $n = 0$, i.e. $S = \emptyset$: Then the rose consists of no circles, but only a point x. So $\pi_1(X) = \{1\} \cong F_S$.
- Let $|S| = n$ and suppose the claim is true for all the cases with cardinality less or equal $n-1$: Take $s_0 \in S$ fixed and let U be a small open neighborhood of the circle in the rose that corresponds to s_0. Let $T := S \setminus \{s_0\}$ contain all the other elements and let V be a small open neighborhood of all the circles corresponding to the elements in T. Now, since $|T| = n-1$ and $|\{s_0\}| = 1$, we get by the induction hypothesis that $F_T \cong \pi_1(V)$ and $\mathbb{Z} = F_{\{s_0\}} \cong \pi_1(U)$.
  Next, choose an orientation on X, i.e. choose the direction for each circle and define the map

$$i : S \to \pi_1(X)$$
$$s \mapsto \text{path around circle corresponding to } s$$

  Now let G be a group and consider any set map $f : S \to G$. Since $\pi_1(U) \cong \langle i(s_0) \rangle$ is free there is a unique group homomorphism $f_1 : \pi_1(U) \to G$ such that $(f_1 \circ i)(s_0) = f(s_0)$. And since $\pi_1(V) \cong F_T$ is free there is a unique group homomorphism $f_2 : \pi_1(V) \to G$ such that $(f_2 \circ i)(t) = f(t)$ for every

$t \in T$. Hence, by the Seifert-van Kampen Theorem we obtain that there is a unique group homomorphism $\tilde{f} : \pi_1(X) \to G$, that extends f_1 and f_2. So in conclusion we have found for the set S with map $i : S \to \pi_1(X)$ and any arbitrary group G with map $f : S \to G$ an unique homomorphism $\tilde{f} : \pi_1(X) \to G$ such that $\tilde{f} \circ i = f$. Thus $\pi_1(X)$ is free over S by Definition 1.1.

□

Proposition 3.3. *Let F be a group. F is free over some finite generating set S if and only if there is a finite graph Δ such that $F \cong \pi_1(\Delta)$.*

Proof. First let F be a free group generated by S. By Proposition 3.2 we know, that F is isomorphic to the fundamental group of the rose with $|S|$ petals, which is a graph.
Conversely assume that there exists some finite graph Δ, such that its fundamental group is isomorphic to F. If we can show that any finite graph is homotopy equivalent to a finite rose, we are done again by Proposition 3.2.
So let T be the maximal tree in Δ. Then $\tilde{\Delta} := \Delta/T$ is a rose, since we contract T when building the quotient. Now consider the map $q : \Delta \to \tilde{\Delta}$. Since T is a tree, there is a map $r : \tilde{\Delta} \to \Delta$ such that r is unique up to homotopy. But $q \circ r$ is homotopic to $id_{\tilde{\Delta}}$ and $r \circ q$ is homotopic to id_{Δ}. So the original graph Δ and the rose $\tilde{\Delta}$ are homotopy equivalent. □

Now we are ready for the second version of the proof of the residual finiteness of free groups:

Proof. (B) [of 3.1] Let F_n be the fundamental group of the n-circle wedge M_n and let $w \in F_n \setminus \{1_{F_n}\}$. We want to find a finite sheeted cover M_W of M_n such that w is not a loop in M_W. So take the universal cover $\tilde{M_n}$ of the n-circle wedge, which is the $2n$-regular tree and Cayley graph of F_n and define as W the path in $\tilde{M_n}$ corresponding to w. Then we can define M_W as follows: The vertices of M_W are precisely all vertices appearing in W and the edges of M_W are those appearing in W plus all necessary edges to make M_W a $2n$-regular tree. This is no problem, since $W \subseteq \tilde{M_n}$. So M_W is finite and $2n$-regular, so it is a finite sheeted cover of the n-circle wedge M_n and W is not a loop in M_W. So $w \notin \pi_1(M_c)$ but $w \in \pi_1(M_n)$ and $[F_n : \pi_1(M_W)] < \infty$, which implies that F_n is residually finite. □

For the third version of the proof we need some facts about subgroups of free groups at hand:
First recall Corollary 2.6, which we can deduce from Proposition 3.3:

Corollary 3.4 (Nielsen-Schreier)**.** *Every subgroup of a free group is free.*

Proof. Let F be a free group. Then by Proposition 3.3 there is a graph X with $F \cong \pi_1(X)$. But any covering X' of X is a graph as well and its fundamental group $\pi_1(X') =: H$ is a subgroup of $\pi_1(X) = F$. Now Proposition 3.3 implies that H is a free group as well. □

Lemma 3.5 (Schreier Index-Formula)**.** *Let F_r be a free group of rank $r \geq 2$ and let $H \leq F_r$ be a subgroup with $[F_r : H] = k < \infty$. Then the rank of H, call it s, is given by $s = 1 + k \cdot (r-1)$.*

Proof. Let $F_r = \pi_1(X)$ with X the rose with r petals and let $H = \pi_1(X')$ for some covering X' of X. By the properties of the Euler characteristic we have $\chi(X') = k \cdot \chi(X)$. But since X is a rose with r petals we know on the other hand that $\chi(X) = 1-r$ and also $\chi(X') = 1-s$. Hence $s = 1-\chi(X') = 1-k\cdot(1-r)$. □

Lemma 3.6. *The free group F_2 of rank 2 has a finite index subgroup isomorphic to F_k for any $k \geq 2$.*

Proof. Let $C_n = \langle c \mid c^n = 1\rangle$ be the cyclic group of size $n \in \mathbb{N}$ generated by the element c. Define the map

$$\begin{aligned} f_n : F_2 &\to C_n \\ x &\mapsto c \\ y &\mapsto 1_{C_n} \end{aligned}$$

Then we get $[F_2 : ker(f_n)] = n$ and since by Lemma 3.5 $[F_2 : ker(f_n)] = \frac{rank(ker(f_n))-1}{rank(F_2)-1}$ we obtain $rank(ker(f_n)) = n+1$. □

To use Malcev's Theorem 2.3 we need to check whether free groups are linear. We do that with the following Proposition:

Proposition 3.7 (Ping-Pong-Lemma)**.** *Let G be a group acting on some set X. Let $a, b \in G$ such that*

- *The elements a and b are of infinite order.*
- *There are non-empty subsets $X_1, X_2 \subseteq X$ such that $X_1 \cap X_2 \neq X_1$ and $X_1 \cap X_2 \neq X_2$ and for any $m \in \mathbb{Z} \setminus \{0\}$*

  (i). $a^m X_2 \subseteq X_1$

  (ii). $b^m X_1 \subseteq X_2$.

Then $\langle a, b\rangle$ is isomorphic to F_2, the free group of rank 2.

Proof. Consider the map

$$\begin{aligned} \varphi : F_2 = \langle x, y\rangle &\to \langle a, b\rangle \leq G \\ x &\mapsto a \\ y &\mapsto b \quad . \end{aligned}$$

Clearly, this is a homomorphism. We want to show that φ is bijective.
Since F_2 does not have any relations, φ has to be surjective. For injectivity we need to show, that $ker(\varphi) = \{1_{F_2}\}$. So take a reduced word $w \in F_2 \setminus \{1_{F_2}\}$ and show that it is not in $ker(\varphi)$, i.e. that $\varphi(w) \neq 1_G$. Since $w \in F_2 \setminus \{1_{F_2}\}$ is a reduced word, $w = x^{m_1}y^{n_1}x^{m_2}y^{n_2}\cdots x^{m_k}y^{n_k}$ with $n_1, m_k \in \mathbb{Z}\setminus\{0\}$, $m_j, n_j \in \mathbb{Z}\setminus\{0\}$ for $j \in \{2, 3, \ldots, k-1\}$ and $m_1, n_k \in \mathbb{Z}$.
Now we need a case distinction:

- *Case 1*: $m_1 \neq 0$ and $n_k = 0$, so $w = x^{m_1}y^{n_1}x^{m_2}y^{n_2}\cdots x^{m_k}$.
Then

$$\begin{aligned} \varphi(w)X_2 &= \varphi(x^{m_1})\varphi(y^{n_1})\varphi(x^{m_2})\varphi(y^{m_2})\cdots\varphi(y^{n_{k-1}})\varphi(x^{m_k})X_2 \\ &= a^{m_1}b^{n_1}a^{m_2}b^{m_2}\cdots b^{n_{k-1}}a^{m_k}X_2 \end{aligned}$$

But now we can use the two properties (i) and (ii) from the second assumption

$$\begin{aligned} & \varphi(w)X_2 = a^{m_1}b^{n_1}a^{m_2}b^{n_2}\cdots b^{n_{k-1}}a^{m_k}X_2 \\ \overset{(i)}{\Rightarrow}\quad & \varphi(w)X_2 \subseteq a^{m_1}b^{n_1}a^{m_2}b^{n_2}\cdots b^{n_{k-1}}X_1 \\ \overset{(ii)}{\Rightarrow}\quad & \varphi(w)X_2 \subseteq a^{m_1}b^{n_1}a^{m_2}b^{n_2}\cdots a^{m_{k-1}}X_2 \\ & \qquad\vdots \\ \Rightarrow\quad & \varphi(w)X_2 \subseteq a^{m_1}X_2 \overset{(i)}{\subseteq} X_1 \\ \overset{X_2 \not\subseteq X_1}{\Rightarrow}\quad & \varphi(w)X_2 \neq X_2 \\ & \quad \varphi(w) \neq 1_G \end{aligned}$$

- *Case 2*: $m_1 = 0$ and $n_k \neq 0$, so $w = y^{n_1}x^{m_2}y^{n_2}\cdots x^{m_k}y^{n_k}$.
Consider $xwx^{-1} \in F_2$ since then by Case 1 we know $\varphi(xwx^{-1}) \neq 1_G$. Suppose now $\varphi(w) = 1_{F_2}$. Then

$$1_{F_2} \neq \varphi(xwx^{-1}) = \varphi(x)\varphi(w)\varphi(x^{-1}) = a\,1_{F_2}\,a^{-1} = aa^{-1} = 1_{F_2},$$

which is a contradiction, i.e. $\varphi(w) \neq 1_G$.

- *Case 3*: $m_1 \neq 0$ and $n_k \neq 0$, so $w = x^{m_1} y^{n_1} x^{m_2} y^{n_2} \cdots x^{m_k} y^{n_k}$.
  Consider $x^l w x^{-l} \in F_2$ for $l \neq -m_1$, since then we get by Case 1 that $\varphi(x^l w x^{-l}) \neq 1_G$ and we can proceed as in Case 2 to get $\varphi(w) \neq 1_G$.

- *Case 4*: $m_1 = 0$ and $n_k = 0$, so $w = y^{n_1} x^{m_2} y^{n_2} \cdots x^{m_k}$.
  Consider $x^l w x^{-l} \in F_2$ for $l \neq m_k$, since then we get by Case 1 that $\varphi(x^l w x^{-l}) \neq 1_G$ and we can proceed as in Case 2 to get $\varphi(w) \neq 1_G$.

□

We apply the Ping-Pong Lemma 3.7 to get:

Proposition 3.8. *The free group F_2 of rank 2 is a linear group.*

Proof. Let $G := SL_2(\mathbb{R})$, let $X := \mathbb{R}^2$ and let G act on X by linear transformation, i.e.

$$\begin{gathered} SL_2(\mathbb{R}) \times \mathbb{R}^2 \to \mathbb{R}^2 \\ \left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}, \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \right) \mapsto \begin{bmatrix} ax_1 + bx_2 \\ cx_1 + dx_2 \end{bmatrix} \end{gathered}$$

Next define $X_1 := \left\{ \begin{bmatrix} x \\ y \end{bmatrix} \middle| |x| > |y| \right\}$ and $X_2 := \left\{ \begin{bmatrix} x \\ y \end{bmatrix} \middle| |x| < |y| \right\}$, since then $X_1 \cap X_2 = \emptyset$.
Lastly define $a = \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix}$ and $b = \begin{bmatrix} 1 & 0 \\ 2 & 1 \end{bmatrix}$ since $\langle a, b \rangle \leq SL_2(\mathbb{Z})$.
Note that $a^m = \begin{bmatrix} 1 & 2m \\ 0 & 1 \end{bmatrix}$ and $b^m = \begin{bmatrix} 1 & 0 \\ 2m & 1 \end{bmatrix}$.
Consider $a^m X_2$

$$a^m \begin{bmatrix} x_2 \\ y_2 \end{bmatrix} = \begin{bmatrix} 1 & 2m \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_2 \\ y_2 \end{bmatrix} = \begin{bmatrix} x_2 + 2my_2 \\ y_2 \end{bmatrix} \in X_1$$

and $b^m X_1$

$$b^m \begin{bmatrix} x_1 \\ y_1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 2m & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ y_1 \end{bmatrix} = \begin{bmatrix} x_2 \\ 2mx_2 + y_2 \end{bmatrix} \in X_2 \,.$$

So all the conditions for the Ping-Pong-Lemma 3.7 are satisfied, hence we can conclude

$$F_2 \cong \langle a, b \rangle \leq SL_2(\mathbb{Z}) \,,$$

i.e. F_2 is linear. □

Now we are ready for the third proof of the residual finiteness of linear groups:

Proof. (C) [of 3.1] Since by Proposition 3.8, the free group of rank 2, F_2 is linear, by Malcev's Theorem 2.3 it is residually finite. But since by Proposition 3.6 any free group F_r of rank $r > 2$ is a subgroup of F_2, and F_2 is finitely generated, by Proposition 1.17 the residual finiteness is passed to F_r as well for any $r > 2$. □

Corollary 3.9. *Finitely generated free groups are Hopfian.*

Proof. Since we know by Proposition 3.1 that free groups are residually finite, Propostion 1.20 implies that any finitely generated free group is Hopfian. □

Corollary 3.10. *Let $k, l \geq 2$. Then the free groups F_k and F_l are isomorphic if and only if $k = l$.*

Proof. Trivially, F_k and F_l are isomorphic if $k = l$. So assume conversely that $F_k \cong F_l$. Let $X = \{x_1, x_2, \dots, x_k\}$ be the generating set of F_k. Without loss of generality we can assume that $k > l$ and define F_l to be the free subgroup of F_k generated by $\{x_1, x_2, \dots, x_l\}$, the first l generators of F_k. Then there is a surjective homomorphism

$$\begin{aligned} \phi : F_k &\to F_l \\ x_i &\mapsto x_i && \forall\, i \in \{1, 2, \dots l\} \\ x_i &\mapsto 1_{F_l} && \forall\, i \in \{l+1, \dots, k\}\,, \end{aligned}$$

which has a non-trivial kernel by definition. But since $F_l \cong F_k$, we get by composing ϕ with this isomorphism a surjective homomorphism $F_k \to F_k$, that has a non-trivial kernel. So F_k is not Hopfian, in contradiction to Corollary 3.9. Thus, we conclude $l = k$. □

3.2 Baumslag-Solitar Groups

Definition 3.1. Let $m, n \in \mathbb{Z} \setminus \{0\}$. The finitely generated group

$$BS(m, n) = \langle a, b \mid b^{-1} a^m b = a^n \rangle$$

is called the *Baumslag-Solitar group* (associated to the pair of integers (m, n)).

Remark 3.2.

(i). When considering $BS(m, n)$ we may assume $0 < m \leq |n|$ as we will see in Theorem 3.12 that $BS(m, n)$ is isomorphic to $BS(-m, -n)$.

(ii). Some authors define the Baumslag-Solitar group $BS(m,n)$ also as

$$BS(m,n) = \langle a, b \mid ba^m b^{-1} = a^n \rangle ,$$

but we will see in Theorem 3.12 that the two groups are isomorphic, thus this difference does not really matter. We will stick with Definition 3.1.

(iii). The Baumslag-Solitar group $BS(m,n)$ is the HNN-extension $A*_\alpha$ of the infinite cyclic group $A = \langle a \rangle$ with associated subgroups $A_1 = \langle a^m \rangle$ and $A_2 = \langle a^n \rangle$ along the isomorphism

$$\begin{aligned} \alpha : A_1 &\to A_2 \\ a^m &\mapsto a^n . \end{aligned}$$

3.2.1 Hyperbolicity

Proposition 3.11. *The Baumslag-Solitar group $BS(m,n)$ is not hyperbolic for any integers $m, n \in \mathbb{Z} \setminus \{0\}$.*

Proof. We need to convince ourselves that bigons in the Cayley graph of $BS(m,n)$ cannot be ε-thin, since then the group is not hyperbolic by Theorem 1.14. Observe that with the relation $a^n = b^{-1}a^m b$ we can deduce for $x \in \mathbb{Z}$

$$a^{nx} = (a^n)^x = (b^{-1}a^m b)^x = b^{-1}a^m b b^{-1} a^m b \cdots b^{-1} a^m b = b^{-1} a^{mx} b . \tag{25}$$

With relation (25) at hand we want to show $b^{-k}a^{m^k}b^k = a^{n^k}$ holds for any $k \in \mathbb{N}$. We use induction on $k \in \mathbb{N}$: The case $k = 0$ holds trivially. So assume $b^{-k}a^{m^k}b^k = a^{n^k}$ is true. Then for any $y \in \mathbb{Z}$ we have

$$(a^{n^k})^y = b^{-k}a^{m^k}b^k b^{-k}a^{m^k}b^k \cdots b^{-k}a^{m^k}b^k = b^{-k}a^{m^k y}b^k .$$

Now we obtain for the case $k+1$:

$$\begin{aligned} b^{-k-1}a^{m^{k+1}}b^{k+1} &= b^{-k}b^{-1}a^{mm^k}bb^k \overset{x:=m^k}{=} b^{-k}b^{-1}a^{mx}bb^k \\ &= b^{-k}a^{nx}b^k = b^{-k}a^{nm^k}b^k \overset{y:=n}{=} b^{-k}a^{m^k y}b^k \\ &= a^{n^k y} = a^{n^k n} \\ &= a^{n^{k+1}} \end{aligned}$$

Thus, $b^{-k}a^{m^k}b^k = a^{n^k}$ and with it $b^{-k}a^{-m^k}b^k = a^{-n^k}$ holds for all $k \in \mathbb{N}$.
Let first $m \neq |n|$. Define $x = 1_{BS(m,n)}$ and $y = a^{|n|^k+1}$ to be two vertices in the Cayley graph C of $BS(m,n)$ with $k \in 2\mathbb{N} \setminus \{0\}$. We choose k to be even, because neighboring a-levels in C have edges oriented in opposite directions in case n is

negative. Since we want to use the above derived relation $b^{-k}a^{m^k}b^k = a^{n^k}$, this means we will go down k steps along b^{-1} and we want to make sure that the a-level at which we arrive has edges oriented in the same direction as the edges of the level we started at in order to always move towards y. Thus we only allow 2-steps. As we aim to let k be arbitrarily large anyways, this is not a restriction. We want to show that there have to be two geodesics γ_1 and γ_2 connecting x and y, one, say γ_1, starting with ab^{-j} and γ_2 starting with b^{-j} where $j \leq k$ increases with k.
In order to do that we need the following observations:

- Clearly, a geodesic connecting x and y cannot begin with a^{-1} as this would mean we move away from y even further.

- A geodesic cannot start with b, i.e. the geodesic does not start upwards from x: Let γ be the path $ba^nb^{-1}a^{|n|^k+1-m}$. It connects x and y, goes up one b and then takes the first opportunity back to the a-axis on which x and y lie. But γ has length

$$1 + |n| + 1 + |n^k| + 1 - m$$

which is longer than the length of the a-level connection $a^{|n|^k+1}$ between x and y, which has length

$$|n^k| + 1\,,$$

as $|n| > m$. Thus γ cannot be a geodesic. But any alterations of γ that keep starting with b are only increasing in length: We can either continue going sideways longer and not take the first b^{-1} down, but this clearly makes the path again longer than $a^{|n|^k+1}$ by the same argument as before. The other option is to go upwards further. Since the path has to go down again eventually as it heads towards y, we can assume it to have some height $l+1$ for $l \in \mathbb{N}$, by which we mean the maximal distance of the occurring a-levels to the basis level containing x and y. Let z be the last vertex before the path makes the last upwards movement along b to the highest occurring a-level of height $l+1$ and let z' be the first vertex the path hits when going back to the a-level containing z. Then by the same two arguments as we used before (with changed signs for n and m necessary) this path is longer than the path that goes the same way to z and then instead of going up just goes along the a-level to z' before continuing as before. This new path has now height l. Thus we have shown that any path of height $l+1$ can be shortened to a path of height l for any $l \in \mathbb{N}$, concluding inductively that no geodesic can start with b.

- The straight a-level path $\gamma := a^{|n|^k+1}$ between x and y is not a geodesic, i.e. there is a shorter path containing some number of b^{-1}: As k is even we know that

$$a^{|n|^k+1} = b^{-k} a^{m^k} b^k a \,.$$

The original path has length $|n|^k + 1$ whereas the one we obtained by the substitution with the relation has length $2k + m^k + 1$. We claim that the second one is always shorter, whenever $k \geq 3$, i.e.

$$2k + m^k < |n|^k \,.$$

The least m and $|n|$ can differ is 1, so if the claim is true for $|n| = m + 1$ it will clearly hold for any greater difference as well. Furthermore if it holds for $k = 3$ it will hold for any $k > 3$ as the exponential growth of m^k and $|n|^k$ outpowers the linear growth of $2k$. But to see that the claim is true for $|n| = m + 1$ and $k = 3$ is an easy check:

$$\begin{aligned} & 2 \cdot 3 + m^3 < (m+1)^3 \\ \Rightarrow \quad & 6 + m^3 < m^3 + 3m^2 + 3m + 1 \\ \Rightarrow \quad & \frac{5}{3} < m^2 + m \,, \end{aligned}$$

which holds for any $m \in \mathbb{N} \setminus \{0\}$. Thus we can always shorten γ by this substitution as long as we choose $k \geq 4$.

- We can arrange the geodesic to start with all the occurring b^{-1}: We know that a acts by isometry on the Cayley graph C. Thus, translating a path γ by a keeps its length unchanged. This means that the path $\gamma * a^p$ has the same length as the path $a^p * \gamma$ for all $p \in \mathbb{Z}$ (where $*$ denotes the concatenation). In particular we can choose γ to have endpoints on the same a-level. Then the paths $\gamma * a^p$ and $a^p * \gamma$ have the same endpoints. Taking γ' now to be a geodesic connecting x and y starting with a^p for some $p \in \mathbb{N} \setminus \{0\}$, i.e. it is of the form $\gamma' = a^p * \gamma$, we know that $\gamma * a^p$ has to be a geodesic as well. Again, this argument can be applied inductively: Suppose γ' is a path that goes down $n+1$ levels. By the hypothesis we may assume γ' starts with b^{-n}. So $\gamma' = b^{-n} a^p * \gamma * \gamma_r$ for some $p \in \mathbb{N} \setminus \{0\}$, where γ is a subpath starting with b^{-1} and going back to the same a-level where it started. γ_r is simply the rest of the geodesic. Then by the same argument as before we know that $a^p * \gamma$ has the same length as $\gamma * a^p$, thus we see that γ' has the same length as $b^{-n} * \gamma * a^p * \gamma_r$ and we have found a geodesic starting with $b^{-(n+1)}$ as desired.

- We can find a condition on how many b^{-1} the geodesic must contain at least: We need to figure out in which cases going further down shortens the path. By the relation (25) we know that instead of going along $a^{p|n|}$ we can take $b^{-1}a^{\pm pm}b$ (where the sign of the exponent depends on the sign of n). This will make the path shorter or keep its length in case

$$2 + pm \leq p|n| \quad \Leftrightarrow \quad \frac{2}{|n| - m} \leq p\,.$$

So let us start with the path $a^{|n|^k}$ and shorten it by going down. The new path is $b^{-1}a^{\pm \frac{|n|^k}{n} m}b = b^{-1}a^{\pm |n|^{k-1}m}b$, which is shorter or equally long if

$$2 + |n|^{k-1}m \leq |n|^k\,.$$

Shortening (or maintaining the length) again is possible in case

$$2 + |n|^{k-2}m^2 \leq |n|^{k-1}m\,.$$

For step $j < k$ we obtain

$$2 + |n|^{k-j}m^j \leq |n|^{k-j+1}m^{j-1}\,.$$

Expressing this in terms of the general case we have

$$p = |n|^{k-j}m^{j-1}\,.$$

Thus as long as

$$|n|^{k-j}m^{j-1} = p \geq \frac{2}{|n| - m}\,, \tag{26}$$

the path is shortened or remains equally long. Thus we know the geodesic must contain at least j times the element b^{-1} where $j \in \{1, \ldots, k\}$ is maximal such that the condition (26) still holds. Clearly with increasing k, the maximal j still satisfying (26) increases as well as m and $|n|$ remain fixed.

Thus we know that γ_2 starting with b^{-j} is well-chosen.

- The geodesic γ_2 will eventually be back on the a-level of x and y where it started. As x and y are $|n|^k + 1$ steps apart on their shared a-level, we know that γ_2 does not reach this a-level exactly at y, but already before, because the b^{-1}-edges starting at x and y respectively do not lie in the same sheet. Thus γ_2 is of the form $\gamma * a$ for a subpath γ starting at x with

b^{-j} and ending at some point lying on the same a-level as x and y. Thus again by the same argument as before the path $a * \gamma$ has the same length as γ_2. But since γ_2 is a geodesic, so is $a * \gamma$ which starts with ab^{-j}.

Thus also γ_1 starting with ab^{-j} is well-chosen.
Now consider for $t \in \{1, \dots, j\}$ the distance

$$d(\gamma_1(t), \gamma_2(t)) \,.$$

We want to estimate it from below by considering the distance d_T of the projections of the points in the Bass-Serre tree T. Let $\pi : C \to T$ be the projection. Then we know that $\pi(x) = \pi(\gamma_1(1))$ since $\gamma_1(1) = a$. Thus we obtain

$$\begin{aligned} d_T(\pi(\gamma_1(t)), \pi(\gamma_2(t))) &= d_T(\pi(\gamma_1(t)), \pi(x)) + d_T(\pi(x), \pi(\gamma_2(t))) \\ &= d_T(\pi(\gamma_1(t)), \pi(\gamma_1(1))) + d_T(\pi(\gamma_2(0)), \pi(\gamma_2(t))) \\ &= t - 1 + t = 2t - 1 \,. \end{aligned}$$

Hence

$$d(\gamma_1(t), \gamma_2(t)) \geq 2t - 1 \,.$$

With $t = j$ we obtain $d(\gamma_1(j), \gamma_2(j)) \geq 2j - 1$, so with growing $k \in 2\mathbb{N} \setminus \{0\}$ we have that j increases and thus the distance becomes arbitrarily large. This means that by Theorem 1.14 the group $BS(m, n)$ cannot be hyperbolic for $m \neq |n|$.
Let now $m = |n|$. Choose $k \in m\mathbb{N}$, again such that $k = 0 \mod 2$ and let $x = 1_{BS(m,n)}$ and $y = a^k b^k$ be two vertices in the Cayley graph C of $BS(m, n)$. Consider the element

$$a^k b^k a^{-k} b^{-k} \,.$$

Since k is an (even) multiple of m there is a $t \in \mathbb{N}$ such that $k = mt$. Thus we have in case $m = n$ with relation (25) that $b^{-1} a^{mt} b = a^{mt}$ and thus $b a^{-mt} b^{-1} = b^{-mt}$. We obtain

$$\begin{aligned} a^k b^k a^{-k} b^{-k} &= a^k b^k a^{-mt} b^{-k} \\ &= a^k b^{k-1} b a^{-mt} b^{-1} b^{-k+1} \\ &= a^k b^{k-1} a^{-mt} b^{-k+1} \\ &\vdots \\ &= a^k b a^{-mt} b^{-1} = a^k a^{-mt} = a^k a^{-k} = 1_{BS(m,n)} \,. \end{aligned}$$

In case $m = -n$ we have with relation (25) that $b^{-1} a^{mt} b = a^{-mt}$ and thus

$ba^{-mt}b^{-1} = a^{mt}$ and $ba^{mt}b^{-1} = a^{-mt}$. We obtain

$$\begin{aligned}
a^k b^k a^{-k} b^{-k} &= a^k b^k a^{-mt} b^{-k} \\
&= a^k b^{k-1} b a^{-mt} b^{-1} b^{-k+1} \\
&= a^k b^{k-1} a^{mt} b^{-k+1} \\
&= a^k b^{k-2} b a^{mt} b^{-1} b^{-k+2} \\
&= a^k b^{k-2} a^{-mt} b^{-k+2} \\
&\textit{even number of steps} \\
&\quad\vdots \\
&= a^k b^2 a^{-mt} b^{-2} \\
&= a^k b a^{mt} b^{-1} \\
&= a^k a^{-mt} = a^k a^{-k} = 1_{BS(m,n)}\,.
\end{aligned}$$

So either way, the element $a^k b^k a^{-k} b^{-k}$ describes a loop in C. Furthermore the elements $a^k b^k$ and $(a^{-k} b^{-k})^{-1} = b^k a^k$ both connect x and y. This helps us to find again two geodesics: Suppose γ is a geodesic connecting x and y. We make the following observations: We can assume that γ stays in one sheet of C, namely the one that contains y, as $m = |n|$. Furthermore γ stays within the area enclosed by the paths $a^k b^k$ and $b^k a^k$ (those paths included) as leaving this area will make the path longer than $a^k b^k$ and $b^k a^k$ by an argument similar to the one used in the previous case: When leaving this area the path will eventually have to return to the area, making staying in this area the shorter option. Furthermore γ has a staircase-shape, i.e. the only possible directions in which γ can move are upwards or to the right. Because again, if it would go down, it would need to go up again eventually, which is longer than just going up. The same applies to moving to the left. Thus γ is of the form

$$\gamma = a^{i_1} b^{j_1} \cdots a^{i_l} b^{j_l}$$

where $i_2, \ldots, i_l \in m\mathbb{Z}$, $j_1, \ldots, j_{l-1} \in \mathbb{N} \setminus \{0\}$, $i_0, j_l \in \mathbb{N}$. We want to rewrite the element representing γ into normal form, starting at the end and using the relation $ba^m = a^{\pm m} b$ as illustrated in Figure 5 below. This is possible at every position as the exponents of a are multiples of m. Also it does not change the length of the element: If we have the relation $ba^m = a^m b$, this is obvious and since all the exponents of a in γ are positive, we end up with the element $a^k b^k$. In the case where the relation is $ba^m = a^{-m} b$, we observe that if γ goes along the basis a-level containing x or any other a-level that is of even distance to the basis, the exponents of the in γ occurring a's is positive. Along the other a-levels the exponent is negative. This means: An element a has a positive exponent in γ if the exponent sum of the b's occurring before is even and otherwise it is negative. Thus, applying the relation to γ to rewrite it, a positive-exponent a changes place with an even number of b's, resulting in a positive exponent in the end and a negative-exponent a changes place with an odd number of b's, thus the last change gives a a positive exponent. So the length of γ is maintained and again we end up at the element $a^k b^k$. Thus $a^k b^k$ has the same length as the geodesic γ and is therefore a geodesic connecting x and y itself. By our previous calculation we know that $b^k a^k$ connects x and y as well and it has the same length as $a^k b^k$. Therefore we can define the two geodesics $\gamma_1 := a^k b^k$ and $\gamma_2 := b^k a^k$ connecting x and y.

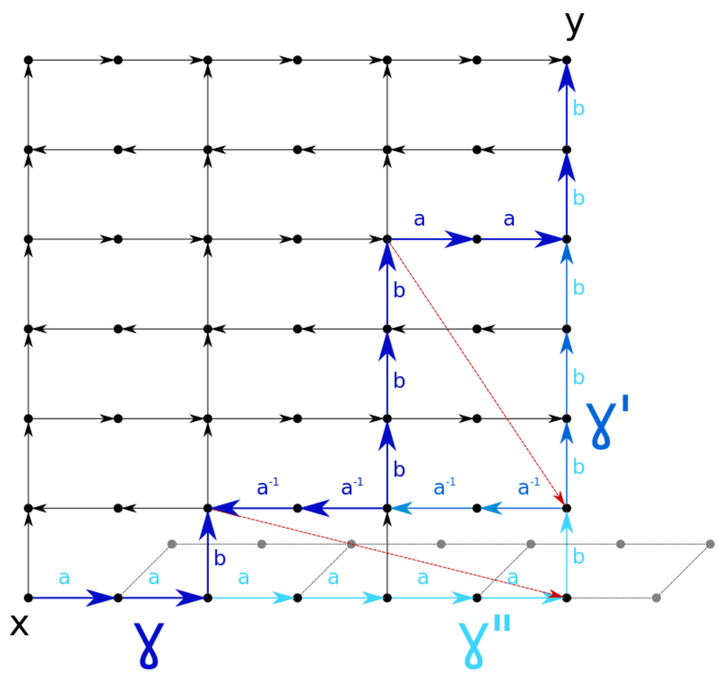


Figure 5: *Consider the Cayley graph of $BS(2,-2)$. Let $\gamma = a^2ba^{-2}b^3a^2b^2$ be a path connecting $x = 1_{BS(2,-2)}$ and $y = a^6b^6$. As illustrated we can deform γ stepwise first to $\gamma' = a^2ba^{-2}a^{-2}b^3b^2 = a^2ba^{-4}b^5$ and then to $\gamma'' = a^2a^4bb^5 = a^6b^6$. The same way any geodesic γ connecting $x = 1_{BS(m,\pm m)}$ and $y = a^kb^k$ can be deformed to a^kb^k.*

Now consider for $t = k$ the distance

$$d(\gamma_1(t), \gamma_2(t))\,.$$

Estimating it from below with the distance d_T in the image of the projection $\pi : C \to T$ of the Cayley graph to the Bass-Serre tree T gives since $\pi(\gamma_1(t)) = \pi(\gamma_1(k)) = \pi(a^k) = \pi(x)$ that

$$d(\gamma_1(t), \gamma_2(t)) \geq d_T(\pi(\gamma_1(t)), \pi(\gamma_2(t))) = d_T(\pi(x), \pi(\gamma_2(k))) = k\,.$$

Thus again for growing $k \in \mathbb{N}$ the distance becomes arbitrarily large, which means that by Theorem 1.14 the group $BS(m,n)$ cannot be hyperbolic for $m = |n|$ either. That concludes the proof. □

3.2.2 Isomorphism-Classes

Theorem 3.12. *Let $G \coloneqq BS(m,n)$ and $H \coloneqq BS(m',n')$ be two Baumslag-Solitar groups. Then G and H are isomorphic if and only if $(m',n') = (m,n)$, $(m',n') = (-m,-n)$, $(m',n') = (n,m)$ or $(m',n') = (-n,-m)$.*

A proof of this theorem can be found in [33]. We want to use a different approach and give a proof that follows similar, more general work on so-called *Generalized Baumslag-Solitar Groups*, introduced by Clay and Forester in [8] and Levitt in [28].
In preparation for our proof, we need some definitions and observations concerning a group acting on a tree, similar to what can be found in [26] and [49]:

Definition 3.2. Let G be a group acting on a tree $T = (V(T), E(T))$ by isometry and without inversions.

(i). For $g \in G$ we define $\tau(g) \coloneqq min\{d(v, gv) \,|\, v \in V(T)\}$ to be the minimal distance that $g \in G$ moves elements of $V(T)$.

(ii). For $g \in G$ we define $Min(g) \coloneqq \{v \in V(T) \,|\, d(v, gv) = \tau(g)\}$ to be the subset of $V(T)$ that contains all vertices that are moved by $g \in G$ only the minimal distance $\tau(g)$.

(iii). An element $g \in G$ is called *elliptic*, if the vertices in $Min(g)$ and edges in $E(Min(g)) \coloneqq \{(v_1, v_2) \in E(T) \,|\, v_1, v_2 \in Min(g)\}$ form a subtree of T that is fixed pointwise by g.

(iv). An element $g \in G$ is called *hyperbolic*, if $Min(g)$ is a line on which g acts by translation by the distance $\tau(g) > 0$.

Lemma 3.13. *Let G be a group acting on a tree $T = (V(T), E(T))$ by isometry. Then:*

(i). Every $g \in G$ is either elliptic or hyperbolic.

(ii). Every $g \in G$ with order $ord(g) < \infty$, i.e. of finite order, is elliptic.

(iii). *Let $g, h \in G$. If $gh = hg$, i.e. g and h commute, then $g\,Min(h) = Min(h)$.*

(iv). *Let $g \in G$ and $p \in \mathbb{Z} \setminus \{0\}$. Then $\tau(g^p) = |p|\,\tau(g)$. Moreover, if $g \in G$ is hyperbolic, then $Min(g^p) = Min(g)$.*

(v). *Let $g \in G$ be hyperbolic and $v \in V(T) \setminus Min(g)$. Then*

$$d(v, gv) = \tau(g) + 2\,d(v, Min(g))\,.$$

(vi). *Let $g \in G$ be a hyperbolic element and let $X \subseteq T$ be a subtree of T that is preserved by $g^{\pm 1}$. Then X contains $Min(g)$.*

Proof. (of Lemma 3.13)

(i). Let $g \in G$.

- ***Case 1:*** Let $\tau(g) > 0$. Show that g is hyperbolic.
  Let $x \in V(T)$ and consider the triangle with vertices $x, gx, g^2x \in V(T)$ illustrated in Figure 6 below.

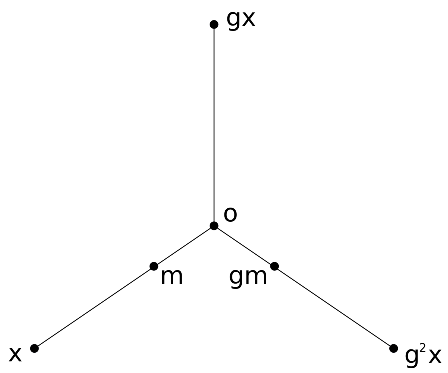


Figure 6: *This tripod illustrates the triangle with vertices x, gx and g^2x. The center of the tripod is called o, the midpoint of the edge $[x, gx]$ is called m.*

If $gx \in [x, g^2x]$, then we have already found the line $Min(g)$. So assume that this is not the case and define $o \in V(T)$ to be the center of the tripod and $m \in V(T)$ the midpoint of $[x, gx]$. If $d(m, x) \geq d(o, x)$, g fixes m, i.e. $gm = m$, which implies $d(m, gm) = 0$ and therefore $\tau(g) = 0$ in contradiction to our assumption. So we can assume $d(m, x) < d(o, x)$.
Now it suffices to show that $d(m, g^2m) = 2\,d(m, gm)$ since that implies that the g-translates of m form a g-invariant line. Since $o \in [m, gm]$ we can deduce $go \in [gm, g^2m]$. Therefore we only need to show that

$d(o, go) = 2\, d(o, gm)$:

$$\begin{aligned} d(o, go) &= d(gx, g^2x) - 2\, d(gx, o) \\ &= d(x, gx) - 2\, (\tfrac{1}{2}\, d(x, gx) - d(o, gm)) \\ &= 2\, d(o, gm)\,. \end{aligned}$$

So the desired line is given by m and all its g-translates as illustrated in Figure 7 below. The distance $d(m, gm)$ is precisely $\tau(g)$. Therefore g is hyperbolic.

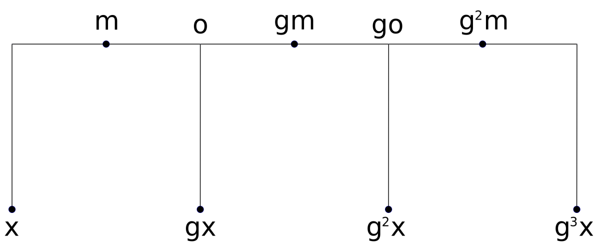


Figure 7: *The line on which g acts by translation by the distance $\tau(g)$.*

- ***Case 2:*** Let $\tau(g) = 0$. Show that g is elliptic.
  Let $x, y \in V(T)$ such that $gx = x$ and $gy = y$, i.e. $x, y \in Min(g)$. Either $|Min(g)| = 1$, i.e. $x = y$, then $(Min(g), \emptyset) \subseteq T$ trivially is a subtree invariant under g, or $|Min(g)| > 1$. Then $x \neq y$ and $[x, y] = [gx, gy]$. So g leaves the path between x and y, that exists since T is a tree, fixed. Therefore any elements in $Min(g)$ are connected, so $(Min(g), E(Min(g))$ is a g-invariant subtree of T. Therefore g is elliptic.

(ii). Let $g \in G$ be of finite order. Since translations of a line must have infinite order, hyperbolic elements cannot be of finite order. Thus, g is not hyperbolic and hence by (i), the only alternative is for g to be elliptic.

(iii). Let $g, h \in G$ such that $gh = hg$ and let $x \in Min(h)$. Then

$$\begin{aligned} x \in Min(h) &\Leftrightarrow d(x, hx) = \tau(h) \\ &\Leftrightarrow d(gx, ghx) = \tau(h) \\ &\Leftrightarrow d(gx, hgx) = \tau(h) \\ &\Leftrightarrow gx \in Min(h)\,. \end{aligned}$$

So therefore $gMin(h) = Min(h)$.

(iv). Let $g \in G$.

- ***Case 1:*** Let $\tau(g) = 0$, i.e. g is elliptic.
  $\Rightarrow$ there is an $x \in V(T)$ such that $gx = x$ and therefore $g^{-1}x = x$

  $$\Rightarrow \begin{cases} g^p x = g^{p-1}(gx) = g^{p-1}x = \cdots = gx = x & \text{if } p > 0 \\ g^p x = g^{p+1}(g^{-1}x) = g^{p+1}x = \cdots = g^{-1}x = x & \text{if } p < 0 \end{cases}$$

  $\Rightarrow \tau(g^p) = 0$
  $\Rightarrow \tau(g^p) = \tau(g) = |p|\,\tau(g)$.

- ***Case 2:*** Let $\tau(g) > 0$, i.e. g is hyperbolic.
  First let $p > 0$: Since by Definition, $Min(g)$ is a line on which g acts by translation by distance $\tau(g)$, we know that $x \in Min(g)$ implies

  $$d(x, g^p x) = d(x, gx) + d(gx, g^2 x) + \cdots + d(g^{p-1}x, g^p x)\,.$$

  But also $d(x, gx) = d(g^{n-1}x, g^n x)$ for any $n \in \mathbb{N}$, so

  $$\begin{aligned} |p|\,\tau(g) &= |p|\,d(x, gx) \\ &= \underbrace{d(x, gx) + d(x, gx) + \cdots + d(x, gx)}_{|p|\text{-times}} \\ &= d(x, gx) + d(gx, g^2 x) + \cdots + d(g^{p-1}x, g^p x) \\ &= d(x, g^p x) \\ &= \tau(g^p)\,. \end{aligned}$$

  Let now $p < 0$: Analogously we get again by Definition for $x \in Min(g)$

  $$d(x, g^p x) = d(x, g^{-1}x) + d(g^{-1}x, g^{-2}x) + \cdots + d(g^{p+1}x, g^p x)\,.$$

  and we know $d(x, g^{-1}x) = d(g^{n+1}x, g^n x)$ for any $n \in \mathbb{Z}^-$, so

  $$\begin{aligned} |p|\,\tau(g) &= |p|\,d(x, g^{-1}x) \\ &= \underbrace{d(x, g^{-1}x) + d(x, g^{-1}x) + \cdots + d(x, g^{-1}x)}_{|p|\text{-times}} \\ &= d(x, g^{-1}x) + d(g^{-1}x, g^{-2}x) + \cdots + d(g^{p+1}x, g^p x) \\ &= d(x, g^p x) \\ &= \tau(g^p)\,. \end{aligned}$$

  So the desired formula holds for any $p \in \mathbb{Z} \setminus \{0\}$.
  It is left to show that for $g \in G$ hyperbolic $Min(g^p) = Min(g)$.

Let $x \in Min(g)$ and let $p > 0$. Then

$$\begin{aligned}
x \in Min(g) &\overset{(iii)}{\Leftrightarrow} gx \in Min(g) \\
&\overset{(iii)}{\Leftrightarrow} g^{p-1}x \in Min(g) \\
&\Leftrightarrow d(g^{p-1}x, g^p x) = \tau(g) = \tfrac{1}{|p|}\,\tau(g^p) \\
&\Leftrightarrow \tau(g^p) = |p|\, d(g^{p-1}x, g^p x) = d(x, g^p x) \\
&\Leftrightarrow x \in Min(g^p)\,.
\end{aligned}$$

So $Min(g) = Min(g^p)$ as desired. Again, the argument for $p < 0$ is completely analogous.

(v). Let $v \in V(T) \setminus Min(g)$ with distance $n \geq 1$ to $Min(g)$. Then let $v' \in Min(g)$ be the vertex in $Min(g)$ that is closest to v. Clearly, by the definition of $Min(g)$ we get

$$[v, gv] = [v, v'] \cup [v', gv'] \cup [gv', gv]\,.$$

Thus the claim holds.

(vi). Let $x \in X$. Then $[x, gx] \subseteq X$, since X is preserved by g. There are two cases. Either $x \in Min(g)$ itself, or for $x \notin Min(g)$, there is an $x' \in Min(g)$ closest to x. Clearly, $x' \in [x, gx]$. Thus we deduce $Min(g) \subseteq X$.

□

Now we are ready to work through the proof of the Isomorphism Problem of Baumslag-Solitar groups, given by Theorem 3.12:

Proof. (of Theorem 3.12)
First prove via Tietze transformations:
Claim 1: $BS(k,l) \cong BS(l,k)$ and $BS(k,l) \cong BS(-k,-l)$ for all $k, l \in \mathbb{Z}$:

$$\begin{aligned}
BS(k,l) &\overset{\text{Def}}{=} \langle a, b \,|\, b^{-1}a^k b = a^l \rangle \\
&\overset{\text{T1}}{=} \langle a, b \,|\, b^{-1}a^k b = a^l,\ a^k = ba^l b^{-1} \rangle \\
&\overset{\text{T2}}{=} \langle a, b \,|\, a^k = ba^l b^{-1} \rangle \\
&\overset{\text{T3}}{=} \langle a, b, y \,|\, a^k = ba^l b^{-1},\ y = b^{-1} \rangle \\
&\overset{\text{T4}}{=} \langle a, y \,|\, a^k = y^{-1}a^l y \rangle \\
&\overset{\text{Def}}{=} BS(l,k)
\end{aligned}$$

$$\begin{aligned} BS(k,l) &\overset{\text{Def}}{=} \langle a,b \,|\, b^{-1}a^kb = a^l \rangle \\ &\overset{\text{T1}}{=} \langle a,b \,|\, b^{-1}a^kb = a^l,\, (b^{-1}a^kb)^{-1} = (a^l)^{-1} \rangle \\ &\overset{\text{T2}}{=} \langle a,b \,|\, b^{-1}a^{-k}b = a^{-l} \rangle \\ &\overset{\text{Def}}{=} BS(-k,-l) \end{aligned}$$

This gives the sufficient condition of the theorem immediately.
Also, now we can restrict ourselves to the case where $0 < m \leq |n|$ in the proof of the necessary condition, since by Claim 1 any group where the roles of m and n are interchanged or both their signs are different will be isomorphic anyways. So let $0 < m \leq |n|$.

Before entering the general case, we want to deal with $BS(1,1)$ and $BS(1,-1)$ separately. First observe that

$$BS(1,1) = \langle a,b \mid b^{-1}ab = a \rangle = \langle a,b \mid ab = ba \rangle$$

is abelian. But by Britton's Lemma 2.17 the Baumslag-Solitar group cannot be abelian for any other pair of integers (m,n), hence $BS(1,1)$ is not isomorphic to any other Baumslag-Solitar group (except $BS(-1,-1)$ of course).
Checking that $G = BS(1,-1)$ is only isomorphic to $BS(-1,1)$ requires a little more work: First observe that G has an abelian subgroup A of index 2, namely the subgroup generated by a and b^2. Now we need the following lemma:

Lemma 3.14. *If an abelian group A is acting cocompactly on a tree T, then there is a point or a line in T that is left invariant under the action.*

Proof. (of Lemma 3.14) It suffices to handle the following three cases for A being generated by a and b:

(i). *a and b elliptic:* Suppose that $Min(a)$ and $Min(b)$ are disjoint sets. Let $x \in Min(a)$ and $y \in Min(b)$ be the points where the sets are closest to each other, i.e. connected by a unique geodesic. Translating $Min(b)$ by a gives that $Min(a)$ and $aMin(b)$ have to be closest to each other in x and ay. But a and b commute, so by 3.13 (iii) we know $aMin(b) = Min(b)$, thus $ay = y$. Hence $y \in Min(a)$ as well, $x = y$ and the sets are not disjoint. That means that there is a vertex $v \in V(T)$ fixed by both a and b, hence v is a point left invariant under the group action.

(ii). *a elliptic and b hyperbolic:* Suppose that $Min(a)$ and $Min(b)$ are disjoint sets. Let $x \in Min(a)$ be the vertex that is closest to $Min(b)$. By Lemma 3.13 (iii) we have $b^n x \in b^n Min(a) = Min(a)$ for all $n \in \mathbb{N}$. Since $Min(a)$

is a subtree of T by definition, $b^n x$ and x have to be connected. But by Lemma 3.13 (v) and (iv) we know that

$$d(x, b^n x) = \tau(b^n) + 2d(x, Min(b^n)) = n\tau(b) + 2d(x, Min(b))$$

for any $n \in \mathbb{N}$. Thus the geodesic connecting $b^n x$ and x has to go through $Min(b)$ and there cannot be a second geodesic since T is a tree. Hence $Min(b) \subseteq Min(a)$ and the line $Min(b)$ is left invariant under the group action.

(iii). *a and b hyperbolic:* Since a and b commute, we know by Lemma 3.13 (iii) that b leaves $Min(a)$ invariant and a leaves $Min(b)$ invariant. Hence by Lemma 3.13 (vi) we obtain that $Min(b) \subseteq Min(a)$ and $Min(a) \subseteq Min(b)$, implying that $Min(a) = Min(b)$. Thus there is a line left invariant under the group action.

□

We know that any Baumslag-Solitar group $BS(m, n)$ with $|m|, |n| > 1$ acts cocompactly and minimally on its Bass-Serre tree, which is $|m| + |n|$-regular (compare with Figure 8). Suppose T is the Bass-Serre tree for $BS(m, n)$ with $|m| + |n| > 2$. If G is isomorphic to $BS(m, n)$, then G acts cocompacty on T as well. Thus by Lemma 1.11, so does A. But by Lemma 3.14, A fixes a point or a line in T, which implies that all of T is contained in a bounded neighborhood of a point or a line, in contradiction to T being a branching tree. Hence $G = BS(1, -1)$ cannot be isomorphic to any other Baumslag-Solitar group (except $BS(-1, 1)$ of course).

Now we are ready to deal with the general case, assuming throughout the proof that $m > 1$ and $|n| > 1$. The guideline of the proof is the following: Assume $G \cong H$ and m and n given. Show that the only choice for m' and n' is $m' = m$ and $n' = n$. In order to do so we will distinguish three cases for (m, n) with $m > 1$ and $|n| > 1$ that need to be treated differently:

(i). $n = m$

(ii). $n = -m$

(iii). $|n| \neq m$

In all three cases we try to find an intrinsic property of G in terms of m and n, i.e. a property that does not depend on the choice of the group presentation. If this property is preserved by the isomorphism, we can conclude that H and therefore m' and n' must have the same property as G and m and n. Then we can conclude that the choice of m' and n' is already given by the choice of m

and n.

In preparation for the case distinctions show the following:
Claim 2: There is a tree T' on which G acts, such that $\tau_{T'}(a) \neq 0$ if and only if $|n| = m$.
For the necessary condition observe first that for any $g \in G$, the minimal distance $\tau(g)$ is invariant under conjugation, i.e. $\tau(g) = \tau(h^{-1}gh)$ for all $h \in G$:

$$\begin{aligned}
\tau(h^{-1}gh) &= min\{d(y, h^{-1}ghy) \,|\, y \in V(T)\} \\
&= min\{d(hy, ghy) \,|\, y \in V(T)\} \\
&= min\{d(hy, ghy) \,|\, hy \in V(T)\} \\
&= min\{d(x, gx) \,|\, x \in V(T)\} \\
&= \tau(g)\,.
\end{aligned}$$

So for $G = BS(m,n) = \langle a, b \,|\, b^{-1}a^m b = a^n\rangle$ we obtain $\tau(a^m) = \tau(b^{-1}a^m b) = \tau(a^n)$. Now suppose that there is a tree T' with $\tau_{T'}(a) > 0$, then Lemma 3.13 (iv). with the fact that $m > 0$ implies

$$0 < m\,\tau_{T'}(a) = |m|\,\tau_{T'}(a) = \tau_{T'}(a^m) = \tau_{T'}(a^n) = |n|\,\tau_{T'}(a)$$

and therefore cancelling $\tau_{T'}(a) \neq 0$ gives $|n| = m$, hence the necessary condition is proven.
For the sufficient condition, assume that $|n| = m$ and show that there is a tree T' with $\tau_{T'}(a) > 0$. Observe that if $m = n$, $BS(m,n)$ quotients onto $BS(1,1) = \langle a, b \mid b^{-1}ab = a\rangle$ and if $m = -n$, then $BS(m,n)$ quotients onto $BS(1,-1) = \langle a, b \mid b^{-1}ab = a^{-1}\rangle$, both times by mapping a to a and b to b. Now take the quotient groups of $BS(1,1)$ and $BS(1,-1)$ by $\langle b^2\rangle$ by adding the relation $b^2 = 1$. In the case of $BS(1,1)$ this gives a presentation for $\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$ and in the case of $BS(1,-1)$ we obtain a presentation for the infinite dihedral group Dih_∞. In both cases we know that the group acts on a line with the element a acting by translation. Hence, so does $BS(m,n)$, which implies that the line is the tree T' with the property $\tau_{T'}(a) > 0$, that we were looking for.

Next, to obtain a condition for the first case, we want to show:
Claim 3: G has non-trivial center $\mathcal{Z}(G)$ if and only if $n = m$. In this case $\mathcal{Z}(G) = \langle a^m\rangle$.
The sufficient condition is straightforward: Let $n = m$, hence we have

$$G = \langle a, b \mid b^{-1}a^m b = a^m\rangle\,.$$

Then $\langle a^m \rangle \subseteq \mathcal{Z}(G)$ since for any $g = a^{e_1}b^{f_1}a^{e_2}b^{f_2}\cdots a^{e_i}b^{f_i} \in G$ with $e_1, f_i \in \mathbb{Z}$, $f_1, e_i \in \mathbb{Z}\setminus\{0\}$ and $e_j, f_j \in \mathbb{Z}\setminus\{0\}$ for $j \in \{2,\dots,i-1\}$ and any $l \in \mathbb{Z}$ we have

$$\begin{aligned} g^{-1}a^{lm}g &= b^{-f_i}a^{-e_i}\cdots b^{-f_1}a^{-e_1}a^{lm}a^{e_1}b^{f_1}\cdots a^{e_i}b^{f_i} \\ &= b^{-f_i}a^{-e_i}\cdots b^{-f_1}a^{lm}b^{f_1}\cdots a^{e_i}b^{f_i} \\ &= b^{-f_i}a^{-e_i}\cdots b^{-f_2}a^{-e_2}a^{lm}a^{e_2}b^{f_2}\cdots a^{e_i}b^{f_i} \\ &= \cdots = b^{-f_i}a^{-e_i}a^{lm}a^{e_i}b^{f_i} = b^{-f_i}a^{lm}b^{f_i} \\ &= a^{lm} \end{aligned}$$

So $\mathcal{Z}(G)$ is not trivial, since $a^{lm} \in \mathcal{Z}(G)$ for all $l \in \mathbb{Z}$.

For the necessary condition, we need to think of G as a graph of groups $\mathcal{G} = (V(\mathcal{G}), E(\mathcal{G}))$: Since G is an HNN-extension, the graph consists of a single vertex $\{v\} = V(\mathcal{G})$ with the edge set $E(\mathcal{G}) = \{(v,v) \mid v \in V(\mathcal{G})\}$ being just a single loop. The vertex group is $\mathcal{G}_v = \langle a \rangle$, the stable letter is $b \in G$ and the edge group is the infinite cyclic group generated by an element, say c, that includes into $\langle a \rangle$ the following way:

$$i_1 : \langle c \rangle \to \langle a \rangle \qquad\qquad i_2 : \langle c \rangle \to \langle a \rangle \tag{27}$$

$$c \mapsto a^m \qquad\qquad c \mapsto a^n \tag{28}$$

So $G = \langle \langle a \rangle, b \mid b^{-1}\, i_1(c')\, b = i_2(c') \quad \forall\, c' \in \langle c \rangle\, \rangle = BS(m,n)$.
Now let $T = (V(T), E(T))$ be the Bass-Serre tree associated to this graph of groups $\mathcal{G} = (V(\mathcal{G}), E(\mathcal{G}))$, which is illustrated in Figure 8 for the general case of m and n. The vertex set is given by $V(T) = \{g\langle a \rangle \mid g \in G\}$ and the edge set $E(T)$ is constructed as follows:

$$\begin{gathered} \exists\, e \in E(T) \text{ with } o(e) = f\langle a \rangle \in V(T) \text{ and } t(e) = h\langle a \rangle \in V(T) \\ \text{if } \exists\, k \in \mathbb{Z} \text{ such that } h\langle a \rangle = fa^k b\langle a \rangle \text{ or } h\langle a \rangle = fa^k b^{-1}\langle a \rangle. \end{gathered}$$

By construction, G acts on T by left multiplication. The vertex stabilizer for $g\langle a \rangle \in V(T)$ is given by

$$G_{g\langle a \rangle} = \{ga^p g^{-1} \mid p \in \mathbb{Z}\} \tag{29}$$

since $ga^pg^{-1}g\langle a \rangle = ga^p\langle a \rangle = g\langle a \rangle$.
Now suppose $\mathcal{Z}(G) \neq \{1\}$ and $z \in \mathcal{Z}(G)$, i.e. for all $g \in \mathcal{Z}(G)$ we have $g^{-1}zg = z$. Therefore $z\,Min(g) = Min(g)$ for all $g \in G$ by Lemma 3.13 (iii). So z fixes all minimal distance sets in T. Thus z fixes the whole tree T, in particular its vertex v, that is stabilized by $\mathcal{G}_v = \langle a \rangle$. So $z \in \langle a \rangle$, that means there is an $l \in \mathbb{Z}$

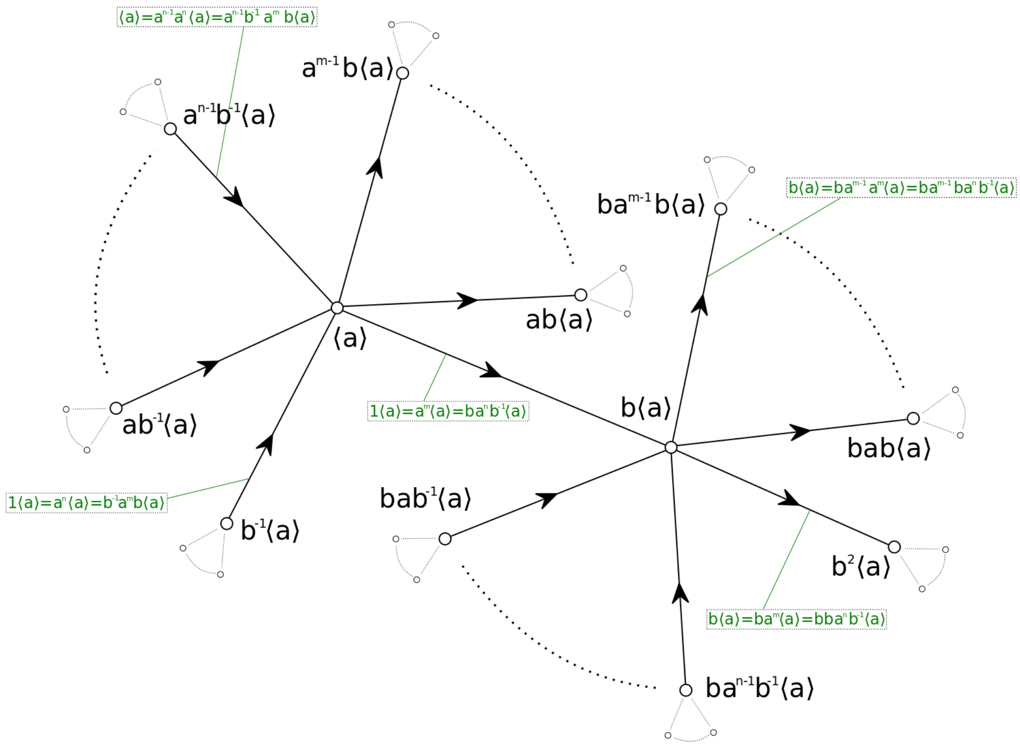


Figure 8: *The Bass-Serre tree of the Baumslag-Solitar group $BS(m,n)$ has $|m|$ outgoing and $|n|$ incoming edges. A vertex $g\langle a\rangle$ is connected to all vertices of the form $ga^pb^{-1}\langle a\rangle$ with an incoming edge for $p \in \{0, \dots n-1\}$ and with vertices of the form $ga^qb\langle a\rangle$ with an outgoing edge for $q \in \{0, \dots, m-1\}$. The green boxes illustrate the computation that has to hold for an edge to exist.*

such that $z = a^l$. Now consider

$$\begin{aligned} z^n = a^{ln} = b^{-1}a^{lm}b \quad &\Leftrightarrow \quad bz^n = a^{lm}b \\ &\overset{z^n \in \mathcal{Z}(G)}{\Leftrightarrow} \quad z^n b = a^{lm}b \\ &\Leftrightarrow \quad z^n = a^{lm}\,. \end{aligned}$$

So $a^{ln} = z^n = a^{lm}$, which implies $n = m$.
Also, in that particular setting with $n = m$ we can see with the relation $b^{-1}a^m b = a^m$, which is equivalent to $a^m b = ba^m$, that a^m commutes with every element in G, so $a^m \in \mathcal{Z}(G)$. But by the property of the edge maps (27) of the graph of groups description $\mathcal{G}$ of G, we know there is an edge whose stabilizer is $\langle a^m\rangle$. Hence the center $\mathcal{Z}(G)$ has to be a subgroup of $\langle a^m\rangle$, which implies that $\mathcal{Z}(G) = \langle a^m\rangle$.

Now we have gathered enough information to prove the first case:

(i). $n = m$: We know the center $\mathcal{Z}(G) = \langle a^m \rangle$, so we can consider $G/\mathcal{Z}(G) = G/\langle a^m \rangle$ and show that this group is isomorphic to the free product $\mathbb{Z} * \mathbb{Z}/m\mathbb{Z}$ by using the following group action on the Bass-Serre tree T:

$$\begin{aligned} \phi : G/\langle a^m \rangle \times V(T) &\to V(T) \\ (g/\langle a^m \rangle, h\langle a \rangle) &\mapsto gh\langle a \rangle \,. \end{aligned} \tag{30}$$

Now we consider the quotient graph of groups of T by this action, which is a vertex v with a single loop l. Attached to v is the group $\langle a \rangle / \langle a^m \rangle \cong \mathbb{Z}/m\mathbb{Z} = \langle x \mid x^m = 1 \rangle$ and attached to l is the group $\langle a^m \rangle / \langle a^m \rangle = \{1\}$. Hence the fundamental group of the quotient, which by the Structure Theorem 2.32 is isomorphic to $G/\langle a^m \rangle$ is given by

$$\langle x, t \mid x^m = 1 \rangle \cong \mathbb{Z}/m\mathbb{Z} * \mathbb{Z} \,.$$

Since we know by assumption $H \cong G$ and that the quotients of isomorphic groups by their center are isomorphic as well, we obtain

$$H/\mathcal{Z}(H) \cong G/\mathcal{Z}(G) \cong \mathbb{Z}/m\mathbb{Z} * \mathbb{Z}$$

which by Claim 2 leads to the conclusion $H = BS(m, m)$, i.e. $m' = m$ and $n' = n = m$.

So now we can continue assuming that G has trivial center $\mathcal{Z}(G) = \{1\}$, i.e. $n \neq m$.

We want to move on to the second case where $n = -m$. By Claim 2 we know there is a tree T', such that $\tau_{T'}(a) > 0$ and $T' = Min(a)$ is a line. Hence, by Lemma 3.13 (iv) follows $Min(a^m) = Min(a) = Min(a^{-m})$ and $\tau(a^m) = |m|\,\tau(a) = \tau(a^{-m})$, thus we can observe: $b\,Min(a^m) = Min(a^m)$, since

$$\begin{aligned} x \in Min(a^m) &\Leftrightarrow x \in Min(a^{-m}) \\ &\Leftrightarrow d(x, a^{-m}x) = \tau(a^{-m}) = \tau(a^m) \\ &\Leftrightarrow d(x, b^{-1}a^m bx) = \tau(a^m) \\ &\Leftrightarrow d(bx, a^m bx) = \tau(a^m) \\ &\Leftrightarrow bx \in Min(a^m) \end{aligned}$$

and so we conclude $bMin(a) = bMin(a^m) = Min(a^m) = Min(a)$. Therefore $Min(a)$ is fixed by b and trivially by a, so by both of the generators of G, which

implies $gMin(a) = Min(a)$ for all $g \in G$. This gives us two cases: Either every element in the group leaves the orientation of the line $Min(a)$ invariant, or there are certain elements that flip it. Since we are in the setting of $n = -m$, we can conclude from the relation $b^{-1}a^m b = a^{-m}$ that b flips the line.
Now we can define a group action ϕ of G on the orientation of the line, which is represented by $\mathbb{Z}/2\mathbb{Z}$ and its induced homomorphism ψ:

$$\begin{aligned} \phi : G \times \mathbb{Z}/2\mathbb{Z} &\to \mathbb{Z}/2\mathbb{Z} & \psi : G &\to \mathbb{Z}/2\mathbb{Z} \\ (g, x) &\mapsto \phi(g, x) = gx & g &\mapsto \psi(g) = \begin{cases} 0 & \text{if } \phi(g,x) = x \ \forall x \in \mathbb{Z}/2\mathbb{Z} \\ 1 & \text{otherwise}. \end{cases} \end{aligned}$$

Since b flips the line, we know $\psi(b) = 1$ and therefore $Im(\psi) \neq \{0\}$. That implies $2 = |Im(\psi)| = [G : ker(\psi)]$. Let us denote that index-2 subgroup by $ker(\psi) =: G'$. We can describe G' explicitly: The elements contained in G' do not flip the line. So let $g \in G'$ be such an element. If there is one occurrence of the line-flipping $b^{\pm 1}$ in g, it needs to contain a second $b^{\pm 1}$ as well, to reverse the flip. Hence, the sum of the exponents of the occurring $b^{\pm 1}$'s needs to be an even integer or 0. This leads to the observation that any $g \in G'$ commutes with a^m: Of course any power of a will commute with a^m, so we only need to figure out, how b and b^{-1} interact with a^m. By the relation $b^{-1}a^m b = a^{-m}$ we can derive

$$\begin{aligned} b^{-1}a^m &= a^{-m}b^{-1} & b^{-1}a^{-m} &= a^m b^{-1} \\ ba^m &= a^{-m}b & ba^{-m} &= a^m b \quad . \end{aligned}$$

So b and b^{-1} can switch place with $a^{\pm m}$ by the cost of a sign change in the exponent. But in $g \in G'$ the number of appearances of b and b^{-1} is even, therefore the number of sign changes is even, which gives us in the end the same exponent m that we started with.

In fact, the subgroup of G' generated by a^m gives all elements that commute with the whole group G':
Claim 4: The center $\mathcal{Z}(G')$ is given by $\langle a^m \rangle$ and $G'/\langle a^m \rangle \cong \mathbb{Z}/m\mathbb{Z} * \mathbb{Z}/m\mathbb{Z} * \mathbb{Z}$.
As in Claim 2, we consider the Bass-Serre tree T on which G acts. Since G' is a subgroup of G, it acts on T by left multiplication as well. This action is not transitive any more: As $g \in G'$ contains an even number of $b^{\pm 1}$, g can move vertices with an even number of $b^{\pm 1}$ only to vertices with an even number of $b^{\pm 1}$ and vertices with an odd number of $b^{\pm 1}$ only to vertices with an odd number of $b^{\pm 1}$. Hence, we have two orbits of vertices, given by the representatives $\langle a \rangle$ and $b\langle a \rangle$. Therefore, if we take the quotient of the tree T by the action of G' we obtain a graph of groups isomorphic to G' consisting of two vertices. Since

the action of G on T preserves the orientation of the edges, the quotient graph has two edges, one incoming and one outgoing for each of the two vertices. Now we need to determine the groups that are associated to the vertices and edges. Therefore we need to find the vertex and edge stabilizers of the G'-action, by intersecting the G-stabilizers with G':
First, consider the vertex stabilizers for the vertex orbit representatives of G'. By (29), we obtain for $\langle a\rangle$ the set $\langle a\rangle$ and for $b\langle a\rangle$ the set $b\langle a\rangle b^{-1}$. Since those are already subgroups of G', they are the G'-stabilizers and hence the groups associated to the vertices of our graph. Next, we want to find the edge stabilizers for the edge orbit representatives of G'. Clearly, those representatives are given by $e = (\langle a\rangle, b\langle a\rangle)$ and $be = (b\langle a\rangle, b^2\langle a\rangle)$. The G-stabilizer for an edge of the form $(g\langle a\rangle, ga^p b^{\pm 1}\langle a\rangle)$ for some $p \in \mathbb{Z}$ is given by

$$\{ga^{lm}g^{-1} \mid l \in \mathbb{Z}\}\,.$$

Therefore for e and be we obtain the sets $\langle a^m\rangle$ and $b\langle a^m\rangle b^{-1}$, which are already subgroups of G' and hence the groups associated to the edges of our graph of groups.
Lastly we want to determine the edge maps. Let $\langle x\rangle$ be the infinite cyclic group stabilizing the first edge and let $\langle y\rangle$ be the infinite cyclic group stabilizing the second. By the graph of groups description we know that $\langle x\rangle$ embeds into $\langle a\rangle$ as $\langle a^m\rangle$, so we obtain as the first map

$$\begin{aligned}\langle x\rangle &\to \langle a\rangle\\ x &\mapsto a^m\,.\end{aligned}$$

Moreover we know that $\langle x\rangle$ embeds into $b\langle a\rangle b^{-1}$ as $\langle ba^m b^{-1}\rangle =: \langle c^m\rangle$. Here we have two possibilities for the map: either $x \mapsto c^m$ or $x \mapsto c^{-m}$. But since we have the relation $a^m = ba^{-m}b^{-1}$ and we know $x \mapsto a^m$, we obtain that $c = ba^{-1}b^{-1}$ and get the second map

$$\begin{aligned}\langle x\rangle &\to \langle ba^m b^{-1}\rangle = \langle c^m\rangle\\ x &\mapsto c^m\,.\end{aligned}$$

Analogously we obtain for the other edge first the map

$$\begin{aligned}\langle y\rangle &\to \langle ba^m b^{-1}\rangle = \langle c^m\rangle\\ y &\mapsto c^m\end{aligned}$$

and deduce from the identities $a^m = ba^{-m}b^{-1}$ and $c = ba^{-1}b^{-1}$ that the second map is given by

$$\begin{aligned}\langle y\rangle &\to \langle a\rangle\\ y &\mapsto a^m\,.\end{aligned}$$

Now we use that any element in the center $\mathcal{Z}(G')$ fixes the whole graph of groups, in particular the edges. So any $z \in \mathcal{Z}(G')$ is in the edge stabilizer $\langle x\rangle = \langle a^m\rangle$, which implies that the center $\mathcal{G}'$ is a subgroup of $\langle a^m\rangle$. But now our previous observation $a^m \in \mathcal{Z}(G')$ implies $\mathcal{Z}(G') = \langle a^m\rangle$. Lastly, we need to show that $G'/\langle a^m\rangle \cong \mathbb{Z}/m\mathbb{Z} * \mathbb{Z}/m\mathbb{Z} * \mathbb{Z}$. Like in case (i), we let $G'/\langle a^m\rangle$ act on the Bass-Serre tree T via the restriction of the action (30):

$$\begin{aligned}\phi' : G'/\langle a^m\rangle \times V(T) &\to V(T)\\ (g\langle a^m\rangle, h\langle a\rangle) &\mapsto gh\langle a\rangle\,.\end{aligned}$$

The quotient graph of groups of T by this action consist of two vertices v_1 and v_2, which are connected by two edges e_1, e_2. Attached to v_1 is the group $\langle a\rangle/\langle a^m\rangle \cong \mathbb{Z}/m\mathbb{Z} = \langle x \mid x^m = 1\rangle$ and attached to v_2 is the group $b\langle a\rangle b^{-1}/\langle a^m\rangle \cong \mathbb{Z}/m\mathbb{Z} = \langle y \mid y^m = 1\rangle$. Furthermore we know that both edge groups $G_{e_1} = \langle a^m\rangle/\langle a^m\rangle = \{1\}$ and $G_{e_2} = b\langle a^m\rangle b^{-1}/\langle a^m\rangle = \{1\}$ are trivial. Thus we obtain that the fundamental group of the quotient, which is by the Structure Theorem 2.32 isomorphic to $G'/\langle a^m\rangle$, is given by

$$\langle x, y, t \mid x^m = y^m = 1\rangle \cong \mathbb{Z}/m\mathbb{Z} * \mathbb{Z}/m\mathbb{Z} * \mathbb{Z}\,.$$

Although this is some progress, it is not enough to consider $G'/\mathcal{Z}(G')$ to conclude that $m' = m$, similar to our argument in case (i), since isomorphic groups can have non-isomorphic index-2 subgroups.

For that reason we need to introduce two definitions:

Definition 3.3. Let G be a group. The subset of elements in $g \in G$, whose centralizer $\mathcal{C}(g) = \{h \in G \,|\, hg = gh\}$ has finite index in G is called the *virtual center* $\mathcal{V}(G)$ of G, i.e.

$$\mathcal{V}(G) \coloneqq \{g \in G \,|\, [G : \mathcal{C}(g)] < \infty\}\,.$$

Definition 3.4. Let G be a group and $H \leq G$ a subgroup. H is called a *characteristic subgroup* of G if every automorphism $\phi : G \to G$ maps H to within itself, i.e. $\phi(H) \leq H$.

In fact, those definitions are related in the following sense:

Claim 5: The virtual center $\mathcal{V}(G)$ of a group G is a characteristic subgroup. First, we need to check that $\mathcal{V}(G)$ is in fact a subgroup of G. So let $a, b \in \mathcal{V}(G)$ be two elements in the virtual center, i.e. their centralizers $\mathcal{C}(a)$ and $\mathcal{C}(b)$ are subgroups of finite index in G. Hence their intersection $\mathcal{C}(a) \cap \mathcal{C}(b)$ is a subgroup as well. Furthermore we can derive

$$\begin{aligned}
[G : \mathcal{C}(a) \cap \mathcal{C}(b)] &= [G : \mathcal{C}(a)][\mathcal{C}(a) : \mathcal{C}(a) \cap \mathcal{C}(b)] \\
&= [G : \mathcal{C}(a)][\mathcal{C}(a)\mathcal{C}(b) : \mathcal{C}(b)] \\
&\leq [G : \mathcal{C}(a)][G : \mathcal{C}(b)] < \infty \,,
\end{aligned}$$

hence $\mathcal{C}(a) \cap \mathcal{C}(b)$ is of finite index as well. But by the definition of the centralizers $\mathcal{C}(a) = \{g \in G \mid ga = ag\}$ and $\mathcal{C}(b) = \{g \in G \mid gb = bg\}$ we can observe that

$$\begin{aligned}
\mathcal{C}(a) \cap \mathcal{C}(b) &= \{g \in G \mid ga = ag, gb = bg\} \\
&= \{g \in G \mid g = aga^{-1}, g = bgb^{-1}\} \\
&= \{g \in G \mid aga^{-1} = bgb^{-1}\} \\
&= \{g \in G \mid ga^{-1}b = a^{-1}bg\} \leq \mathcal{C}(a^{-1}b) \,,
\end{aligned}$$

thus $\mathcal{C}(a) \cap \mathcal{C}(b)$ is a subgroup of $\mathcal{C}(a^{-1}b)$. That implies since $\mathcal{C}(a) \cap \mathcal{C}(b)$ has finite index in G, so does $\mathcal{C}(a^{-1}b)$. Hence, $a^{-1}b \in \mathcal{V}(G)$, and so $\mathcal{V}(G)$ is a subgroup of G.

For the virtual center to be a characteristic subgroup, we need to show that $\phi(\mathcal{V}(G)) \leq \mathcal{V}(G)$ for any automorphism $\phi : G \to G$. So we want to show that $\phi(v) \in \mathcal{V}(G)$ for every $v \in \mathcal{V}(G)$. That means, that we need to show that $[G : \mathcal{C}(\phi(v))] < \infty$ for every $v \in \mathcal{V}(G)$.

So let $v \in \mathcal{V}(G)$, i.e. $[G : \mathcal{C}(v)] < \infty$. Since this index is finite, it stays the same under any automorphism ϕ

$$\infty > [G : \mathcal{C}(v)] = [\phi(G) : \phi(\mathcal{C}(v))] = [G : \mathcal{C}(\phi(v))]$$

Therefore $\phi(v) \in \mathcal{V}(G)$ and $\mathcal{V}(G)$ is a characteristic subgroup.

Remark 3.3. In some literature (e.g. [40]), the virtual center is often referred to as *FC-center* and elements of the virtual center are called *FC-elements*.

Claim 5 will be crucial for the second case, together with the following map:

Definition 3.5. Let $G \coloneqq BS(m, n)$ be a Baumslag-Solitar group and let $a \in G$.

The map

$$\begin{aligned}\phi_a : G &\to \mathbb{Q}_+^* \\ g &\mapsto \tfrac{[g\langle a\rangle g^{-1} : g\langle a\rangle g^{-1} \cap \langle a\rangle]}{[\langle a\rangle : g\langle a\rangle g^{-1} \cap \langle a\rangle]}\end{aligned}$$

is called the *modular homomorphism of G associated to a.*

First, let us check that this is a well-chosen term:
Claim 6: ϕ_a is a homomorphism.
For that we think of ϕ_a in the following way: We have the relation $b^{-1}a^m b = a^n$ and with it $ba^n b^{-1} = a^m$ given in G, so we can conclude

$$b \in \{g \in G \mid g\langle a\rangle g^{-1} \cap \langle a\rangle \text{ has finite index in } g\langle a\rangle g^{-1} \text{ and in } \langle a\rangle\}\,.$$

Clearly

$$a \in \{g \in G \mid g\langle a\rangle g^{-1} \cap \langle a\rangle \text{ has finite index in } g\langle a\rangle g^{-1} \text{ and in } \langle a\rangle\}$$

as well and since G is generated by a and b we get that

$$G = \{g \in G \mid g\langle a\rangle g^{-1} \cap \langle a\rangle \text{ has finite index in } g\langle a\rangle g^{-1} \text{ and in } \langle a\rangle\}\,.$$

This means that for any $g \in G$ we can find $p, q \in \mathbb{Z}$ such that $ga^p g^{-1} = a^q$ and $\phi_a(g) = |\frac{p}{q}|$.
Now let $g, h \in G$. There are $p_g, q_g, p_h, q_h \in \mathbb{Z}$ such that

$$\begin{array}{lll} & g\, a^{p_g}\, g^{-1} = a^{q_g} & h\, a^{p_h}\, h^{-1} = a^{q_h} \\ \text{i.e.} & \phi_a(g) = |\frac{p_g}{q_g}| & \phi_a(h) = |\frac{p_h}{q_h}| \end{array}$$

Since for conjugate elements also any powers of them are conjugate (compare with condition 25) we can compute:

$$\begin{array}{ll} & gh\, a^{p_g p_h}\, h^{-1} g^{-1} = g\, a^{q_h p_g}\, g^{-1} = a^{q_g q_h} \\ \text{i.e.} & \phi_a(gh) = |\frac{p_g p_h}{q_p q_h}|\,. \end{array}$$

So we conclude with $\phi_a(gh) = |\frac{p_g p_h}{q_p q_h}| = |\frac{p_g}{q_g}||\frac{p_h}{q_h}| = \phi_a(g)\,\phi_a(h)$ that ϕ_a is a homomorphism.

Now, we can observe that for the generators $a, b \in G$, we obtain $\phi_a(a) = 1$ and by the group relation $\phi_a(b) = |\frac{n}{m}|$. Thus, since ϕ_a is a homomorphism, the image $|\frac{p_g}{q_g}|$ of any $g \in G$ under ϕ_a has to be a multiple of $|\frac{n}{m}|$, hence $n \mid p_g$ and $m \mid q_g$.

With the modular homomorphism we can show:

Claim 7: G has trivial center, but non-trivial virtual center if and only if $n = -m$. In this case, $\mathcal{V}(G) = \langle a^m \rangle$.

The Bass-Serre tree T on which G acts by left multiplication that was constructed in Claim 3 is not a line. Suppose that the virtual center $\mathcal{V}(G)$ is non-trivial. and let $z \in \mathcal{V}(G) \setminus \{1_G\}$. Define its centralizer to be $G'' := \mathcal{C}(z) \leq G$. Hence, z is in the center $\mathcal{Z}(G'')$ and acts therefore trivially on the Bass-Serre tree T. So in particular, z fixes all vertices, thus it is an element of all vertex stabilizers, which means

$$z \in \bigcap_{v \in V(T)} G_v \, .$$

Now suppose that $|m| \neq |n|$. Since the image of the modular homomorphism ϕ_a is a multiple of $|\frac{n}{m}|$, this means that in this case the image of ϕ_a is unbounded. So there is always a $g \in G$ such that the index $[\langle a \rangle : g\langle a \rangle g^{-1} \cap \langle a \rangle]$ is arbitrarily large. That means that we can find a vertex stabilizer $g\langle a \rangle g^{-1}$ that has arbitrarily little elements in common with the vertex stabilizer $\langle a \rangle$. So the intersection of the vertex stabilizers has to be trivial, i.e.

$$z \in \bigcap_{v \in V(T)} G_v = \{1_G\} \, .$$

That leads to the conclusion $z = 1_G$, in contradiction to the assumption $z \in \mathcal{V}(G) \setminus \{1_G\}$. So $|m|= |n|$.

Conversely suppose now $n = -m$. Since G' as the kernel of the line flipping homomorphism has index 2 in G, its center $\langle a^m \rangle = \mathcal{Z}(G') \subseteq \mathcal{V}(G)$ is contained in the virtual center, so the virtual center $\mathcal{V}(G)$ cannot be trivial.

It is left to show, that $\langle a^m \rangle$ is the whole virtual center: Suppose on the contrary there is a $z \in \mathcal{V}(G) \setminus \langle a^m \rangle$. That means that its centralizer $\mathcal{C}(z) =: G''$ has finite index in G. Hence, G'' acts on the Bass-Serre tree T with finitely many vertex orbits. So if we take the quotient of T by the action of G'' and construct from this graph of groups again a Bass-Serre tree we get T back. Thus, since $z \in \mathcal{C}(z)$, $z \in \mathcal{Z}(G'')$, by the same arguments as in Claim 4, z acts trivially on T and stabilizes in particular every edge. So z is contained in the edge stabilizer $\langle a^m \rangle$. So $\mathcal{V}(G) = \langle a^m \rangle$.

Now we can wrap up the second case:

(ii). $n = -m$: Since the virtual center $\mathcal{V}(G) \subseteq G$ is a characteristic subgroup, we know that its centralizer $\mathcal{C}(\mathcal{V}(G))$ and therefore the quotient $\mathcal{C}(\mathcal{V}(G))/\mathcal{V}(G)$ is an isomorphism invariant, i.e. for two isomorphic groups G and H their quotients $\mathcal{C}(\mathcal{V}(G))/\mathcal{V}(G)$ and $\mathcal{C}(\mathcal{V}(H))/\mathcal{V}(H)$ will be isomorphic as well. Now observe that the centralizer of the virtual center is precisely $\mathcal{C}(\mathcal{V}(G)) =$

$\mathcal{C}(\langle a^m \rangle) = G'$: By definition $\mathcal{C}(\langle a^m \rangle) = \{g \in G \mid gs = sg \quad \forall\, s \in \langle a^m \rangle\}$ and we already know that $G' \subseteq \mathcal{C}(\langle a^m \rangle)$ since any element in $\langle a^m \rangle$ commutes with all elements in G'. But also we already know that any element $g \notin G'$ does not commute, so G' is the whole centralizer of the virtual center. And by Claim 4, $G'/\langle a^m \rangle$ is isomorphic to $\mathbb{Z}/m\mathbb{Z} * \mathbb{Z}/m\mathbb{Z} * \mathbb{Z}$. Hence

$$\mathcal{C}(\mathcal{V}(H))/\mathcal{V}(H) \cong \mathcal{C}(\mathcal{V}(G))/\mathcal{V}(G) = G'/\langle a^m \rangle \cong \mathbb{Z}/m\mathbb{Z} * \mathbb{Z}/m\mathbb{Z} * \mathbb{Z}$$

and by Claim 7 we conclude that $H = BS(m,n)$, i.e. $m' = m$ and $n' = n = -m$.

That leaves us with the case where G has trivial center and trivial virtual center, so $m \neq |n|$. By Claim 2 it follows that $\tau(a) = 0$ for every action of G on a tree, which gives the subgroup $\langle a \rangle \subseteq G$ a special role: It is a maximal universally elliptic subgroup. By *universally elliptic* we mean, that all elements in $\langle a \rangle$ are elliptic for any arbitrary action of G on a tree, which we know to be true since $\tau(a) = 0$. The maximality condition refers to containment, which means there is no universally elliptic subgroup properly containing $\langle a \rangle$. This holds because we know that $\langle a \rangle$ is the maximal elliptic subgroup for the natural action of G on T. Therefore, it is a maximal universally elliptic subgroup. Then, since the minimal distance is invariant under conjugation as shown in the proof of Claim 2, conjugating $\langle a \rangle$ gives all the maximal universally elliptic subgroups of G, i.e. G has precisely one conjugacy class of maximal universally elliptic subgroups.

Now we want to find out how the modular homomorphism for any other representative of this conjugacy class than $\langle a \rangle$ looks like:
Claim 8: If we replace $\langle a \rangle$ by one of its conjugates $h\langle a \rangle h^{-1}$ for $h \in G$, the modular homomorphism stays the same, i.e. $\phi_a = \phi_{hah^{-1}}$.
First observe, that the index stays the same if we conjugate both groups by the same element: Let N be a group containing subgroups $M, K \leq N$ and let $M \leq K$ be a subgroup of K. Let $n \in N \setminus K$. Then

$$\begin{aligned}
[nKn^{-1} : nMn^{-1}] &= |\{k'nMn^{-1} \mid k' \in nKn^{-1}\}| \\
&= |\{nkn^{-1}nMn^{-1} \mid k \in K\}| \\
&= |\{nkMn^{-1} \mid k \in K\}| \\
&= |\{kM \mid k \in K\}| \\
&= [K : M]
\end{aligned}$$

Now let $g \in G$ and consider

$$\phi_{hah^{-1}}(g) = \tfrac{[gh\langle a\rangle h^{-1}g^{-1}:gh\langle a\rangle h^{-1}g^{-1}\cap h\langle a\rangle h^{-1}]}{[h\langle a\rangle h^{-1}:gh\langle a\rangle h^{-1}g^{-1}\cap h\langle a\rangle h^{-1}]}\,.$$

Conjugating the index in the numerator by h^{-1} gives

$$\begin{aligned}
&[gh\langle a\rangle h^{-1}g^{-1} : gh\langle a\rangle h^{-1}g^{-1} \cap h\langle a\rangle h^{-1}] = \\
&= [h^{-1}gh\langle a\rangle h^{-1}g^{-1}h : h^{-1}gh\langle a\rangle h^{-1}g^{-1}h \cap h^{-1}h\langle a\rangle h^{-1}h] \\
&= [h^{-1}gh\langle a\rangle h^{-1}g^{-1}h : h^{-1}gh\langle a\rangle h^{-1}g^{-1}h \cap \langle a\rangle] \\
&= [h^{-1}gh\langle a\rangle (h^{-1}gh)^{-1} : h^{-1}gh\langle a\rangle (h^{-1}gh)^{-1} \cap \langle a\rangle]
\end{aligned}$$

and conjugating the index in the denominator by h^{-1} gives

$$\begin{aligned}
&[h\langle a\rangle h^{-1} : gh\langle a\rangle h^{-1}g^{-1} \cap h\langle a\rangle h^{-1}] = \\
&= [h^{-1}h\langle a\rangle h^{-1}h : h^{-1}gh\langle a\rangle h^{-1}g^{-1}h \cap h^{-1}h\langle a\rangle h^{-1}h] \\
&= [\langle a\rangle : h^{-1}gh\langle a\rangle h^{-1}g^{-1}h \cap \langle a\rangle] \\
&= [\langle a\rangle : h^{-1}gh\langle a\rangle (h^{-1}gh)^{-1} \cap \langle a\rangle]\,.
\end{aligned}$$

By inserting that and using that ϕ_a is a homomorphism into the commutative group $\mathbb{Q}_+^*$ we can conclude

$$\begin{aligned}
\phi_{hah^{-1}}(g) &= \tfrac{[gh\langle a\rangle h^{-1}g^{-1}:gh\langle a\rangle h^{-1}g^{-1}\cap h\langle a\rangle h^{-1}]}{[h\langle a\rangle h^{-1}:gh\langle a\rangle h^{-1}g^{-1}\cap h\langle a\rangle h^{-1}]} \\
&= \tfrac{[h^{-1}gh\langle a\rangle (h^{-1}gh)^{-1}:h^{-1}gh\langle a\rangle (h^{-1}gh)^{-1}\cap\langle a\rangle]}{[\langle a\rangle:h^{-1}gh\langle a\rangle (h^{-1}gh)^{-1}\cap\langle a\rangle]} \\
&= \phi_a(h^{-1}gh) = \phi_a(h)^{-1}\phi_a(g)\phi_a(h) \\
&= \phi_a(h)^{-1}\phi_a(h)\phi_a(g) = \phi_a(g)
\end{aligned}$$

Actually, we can modify this homomorphism to be mapped into all multiplicative rationals. If we use the interpretation of ϕ_a described in the proof of Claim 6, that for any $g \in G$ we can use the $p, q \in \mathbb{Z}$ satisfying $ga^pg^{-1} = a^q$ to obtain $\phi_a(g) = |\frac{p}{q}|$, we see by very similar arguments as in Claim 6 and Claim 8 that we can lose the absolute value on $\frac{p}{q}$ and still have a homomorphism that is invariant under the conjugation of $\langle a\rangle$. So let us denote this homomorphism as

$$\begin{aligned}
\widetilde{\phi_a} : G &\to \mathbb{Q}^* \\
g &\mapsto \tfrac{p}{q}\,.
\end{aligned}$$

Finally, we see that $\widetilde{\phi_a}(g)$ is a multiple of $\frac{n}{m}$ for any $g \in G$, since that is true for generators of G. So the value $\frac{n}{m}$ is an isomorphism invariant.
That lets us draw the conclusion that $\frac{n}{m} = \frac{n'}{m'}$.

This still does not imply the desired result $m = m'$, $n = n'$, unless we can determine the $gcd(m, n)$ and show that it is an isomorphism invariant as well. And this is precisely what we want to do:
Let I be a representative of the only conjugacy class of maximal universally elliptic subgroups that is contained in G. Consider the subgroup $J \leq I$ that is generated by all elements $c \in I \cap gIg^{-1}$ for all $g \in G \setminus I$.
Claim 9: The index satisfies $[I : J] = gcd(m, n)$ and it is independent of the choice of I.
Suppose we would have chosen any other representative $\tilde{I}$ of the conjugacy class instead of I. That means that there is some $\tilde{g}$ such that $\tilde{g}I\tilde{g}^{-1} = \tilde{I}$. But since then the corresponding $\tilde{J}$ is generated by an element $\tilde{c} \in \tilde{g}I\tilde{g}^{-1} \cap g\tilde{g}I\tilde{g}^{-1}g^{-1}$ and we know that the index is invariant under conjugation of both groups, we can conclude that $[\tilde{I} : \tilde{J}] = [I : J]$. So the choice of I and J does not matter. Therefore pick the most convenient choice of $I = \langle a \rangle$.
Now consider the Bass-Serre tree T with $v \in V(T)$ the vertex that has the vertex stabilizer $G_v = \langle a \rangle$ associated to it. We want to describe J in terms of the tree: The conjugates of I are simply vertex stabilizers of other vertices. Let w be a vertex adjecent to v, denote the edge $(v, w) =: e_{vw} \in E(T)$. By definition, for the stabilizers holds $G_v \cap G_w = G_{e_{vw}}$. But since $J \leq \langle a \rangle \cap g\langle a \rangle g^{-1} = G_v \cap G_x$ for some vertex $x \in V(T)$ with stabilizer $G_x = g\langle a \rangle g^{-1}$ we conclude that J contains G_e for every edge $e \in N(v)$, i.e. for every edge e incident to v. But for any other vertex $u \notin N(v)$, we know that $G_v \cap G_u$ fixes the geodesic segment that connects v and u. Therefore it fixes in particular the very first edge $e_o \in E(T)$ of this segment. So $G_v \cap G_u$ is a subgroup of G_{e_o}. But since e_o is incident to v, we know that G_{e_o} and therefore $G_v \cap G_u$ is already contained in J. So J is simply generated by all the edge stabilizers for edges incident to v, which itself has stabilizer $G_v = \langle a \rangle$. So we conclude $J = \langle a^m, a^n \rangle = \langle a^{gcd(m,n)} \rangle$. But since $I = \langle a \rangle$ that implies $[I : J] = [\langle a \rangle : \langle a^{gcd(m,n)} \rangle] = gcd(m, n)$.
Summarizing those observations finishes the third case:

(iii). $m \neq |n|$: We know that in this setting since $\tau(a) = 0$ we only have one conjugacy class of maximal universally elliptic subgroups, with $\langle a \rangle$ as a representative. By Claim 8, we know that for any element of this conjugacy class, we will always get the same modular homomorphism $\tilde{\phi}_a$, which makes the quotient $\frac{n}{m}$ an isomorphism invariant. But by Claim 9 the index $[I : J] = gcd(m, n)$ is an isomorphism invariant as well. So for $G \cong H$ we obtain $\frac{n}{m} = \frac{n'}{m'}$ and $gcd(m, n) = gcd(m', n')$, which leads to the conclusion $m = m'$ and $n = n'$.

So putting all cases together we conclude that whenever G and H are isomorphic, there are only four possible relations between (m, n) and (m', n'), namley $(m', n') = (-m, -n)$, $(m', n') = (n, m)$ or $(m', n') = (-n, -m)$ by Claim 1 or $(m', n') = (m, n)$ by the case distinctions.

$\square$

3.2.3 Residual Finiteness and Hopf-Property

Example 3.1. $BS(2, 3)$ is not Hopfian and therefore not residually finite.

Proof. Let $G \coloneqq BS(2, 3) = \langle a, b \mid b^{-1}a^2b = a^3\rangle$. Define the map

$$\begin{aligned} \mu : G &\to G \\ a &\mapsto a^2 \\ b &\mapsto b\,. \end{aligned}$$

Clearly, μ induces a homomorphism. Since

$$\mu(b^{-1}aba^{-1}) = \mu(b)^{-1}\mu(a)\mu(b)\mu(a)^{-1} = b^{-1}a^2ba^{-2} = a^3a^{-2} = a$$

and

$$\mu(b) = b\,,$$

both generators of G are in the image of μ, therefore the homomorphism is surjective. We want to show that it is not injective, i.e. that $ker(\mu) \neq \{1_G\}$. Consider the word $(b^{-1}aba^{-1})^2a^{-1} = b^{-1}aba^{-1}b^{-1}aba^{-2} \in G$. By the Normal Form Theorem 2.17, this is not the trivial word. Now compute the image of $(b^{-1}aba^{-1})^2a^{-1}$ under μ:

$$\begin{aligned} \mu((b^{-1}aba^{-1})^2a^{-1}) &= \mu(b^{-1}aba^{-1})^2\mu(a)^{-1} = (b^{-1}a^2ba^{-2})^2a^{-2} \\ &= b^{-1}a^2ba^{-2}b^{-1}a^2ba^{-2}a^{-2} \\ &= a^3a^{-2}a^3a^{-4} = 1_G \end{aligned}$$

Thus, $(b^{-1}aba^{-1})^2a^{-1} \in ker(\mu)$, so the kernel is non-trvial. Therfore we have found an epimorphism with non-trivial kernel, which implies that $BS(2, 3)$ is non-Hopfian and hence by Proposition 1.20 not residually finite. [42] $\square$

This observation for the case $m = 2$ and $n = 3$ generalizes to the following proposition:

Proposition 3.15.

(i). The Baumslag-Solitar group $BS(m, n)$ is residually finite and thus Hopfian if $|m| = 1$ or $|n| = 1$ or $|m| = |n|$.

(ii). The Baumslag-Solitar group $BS(m,n)$ is not residually finite, but Hopfian, if $|m| \neq 1$, $|n| \neq 1$, $|m| \neq |n|$ and $\pi(m) = \pi(n)$, where $\pi(k)$ denotes the set of prime divisors of $k \in \mathbb{Z}$.

(iii). The Baumslag-Solitar group $BS(m,n)$ is not residually finite and not Hopfian, if $|m| \neq 1$, $|n| \neq 1$, $|m| \neq |n|$ and $\pi(m) \neq \pi(n)$, where $\pi(k)$ denotes the set of prime divisors of $k \in \mathbb{Z}$.

Remark 3.4. Originally, a characterization of residual finiteness and the Hopf-property of Baumslag-Solitar groups was given by Gilbert Baumslag and Donald Solitar themselves in [4] in 1962. But they claimed that the Baumslag-Solitar group $BS(m,n)$ is not only Hopfian, but also residually finite in case (ii)., i.e. when m and n have the same set of prime factors, which turns out to be wrong. The necessary correction was made by Meskin in 1972 and therefore the proof should be guided by his paper [32]. It divides the proof into six lemmata, whose proofs can be found in different sources, marked below.

Proof. To tackle the necessary case distinction we introduce a lemma per case:

Lemma 3.16. *If m and n are integers such that either $|m| = 1$ or $|n| = 1$, then $G = BS(m,n)$ is residually finite and thus Hopfian.*

The main idea of this proof was suggested by Derek Holt in [35].

Proof. Let $m = 1$ and $n \in \mathbb{Z} \setminus \{-1, 0, 1\}$ and define M to be the matrix group generated by the two matrices $A = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$ and $B = \begin{bmatrix} n & 0 \\ 0 & 1 \end{bmatrix}$. We want to show that the map induced by

$$\begin{aligned} \varphi : BS(1,n) &\to M \\ a &\mapsto A \\ b &\mapsto B \end{aligned}$$

defines an isomorphism.

First, we need to check that φ is a homomorphism, i.e. that the relation of

$BS(1,n)$ also holds in M:

$$\begin{aligned}
B^{-1}AB &= \begin{bmatrix} n & 0 \\ 0 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} n & 0 \\ 0 & 1 \end{bmatrix} \\
&= \begin{bmatrix} \frac{1}{n} & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} n & 0 \\ 0 & 1 \end{bmatrix} \\
&= \begin{bmatrix} \frac{1}{n} & 0 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} n & 0 \\ 0 & 1 \end{bmatrix} \\
&= \begin{bmatrix} 1 & 0 \\ n & 1 \end{bmatrix} \\
&= \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}^{n} = A^n \,.
\end{aligned}$$

Since the generators of $BS(1,n)$, are mapped to the generators of M, φ is surjective. That leaves injectivity to be shown: Observe that any element of the form

$$a^{\epsilon_1} b^{f_1} a^{\epsilon_2} \cdots a^{\epsilon_k} b^{f_k} \in BS(1,n)$$

with $\epsilon_l, f_l \in \mathbb{Z} \setminus \{0\}$ for $l \in \{2, \ldots, k-1\}$, $f_1, \epsilon_k \in \mathbb{Z} \setminus \{0\}$ and $\epsilon_1, f_k \in \mathbb{Z}$ can be written as

$$b^i a^j \in BS(1,n)$$

for some pair $i, j \in \mathbb{Z}$, since any occurrence of the subword ab can be replaced by ba^n. Fruthermore, any element $b^i a^j \in BS(1,n)$ is being mapped to $B^i A^j \in M$ for $i, j \in \mathbb{Z}$ via φ, but since

$$B^i A^j = \begin{bmatrix} n^i & 0 \\ j & 1 \end{bmatrix},$$

and $n \in \mathbb{Z} \setminus \{-1, 0, 1\}$ all $B^i A^j$ are distinct for distinct pairs $i, j \in \mathbb{Z}$. Hence φ is injective and $BS(1,n)$ is isomorphic to a linear group. Thus, by the Theorem 2.3 of Malcev, $BS(1,n)$ is residually finite.
Lastly, we note that by Claim 1 of Theorem 3.12, $BS(n,1)$ is isomorphic to $BS(1,n)$, and $BS(-1,n)$ and $BS(n,-1)$ are both isomorphic to $BS(1,-n)$. Thus $BS(m,n)$ has to be residually finite for all integers m and n where either $|m| = 1$ or $|n| = 1$. □

Lemma 3.17. *If m and n are integers such that $|m| = |n| > 1$, then $G = BS(m,n)$ is residually finite and thus Hopfian.*

The proof follows the idea suggested in [34].

Proof. First let $m = n$, i.e. we consider the group $BS(m, m)$. Define φ to be the homomorphism induced by

$$\begin{aligned} \varphi : BS(m,m) &\to \mathbb{Z} \\ a &\mapsto 1 \\ b &\mapsto 0\,. \end{aligned}$$

Considering the Cayley graph of $BS(m, m)$, we can interpret the map as measuring the height of a given point in regards of the a-level, independent of the b-level of the sheet the point is in. Hence, elements of the kernel H of φ have to be on the same a-height, independent of their b-sheet. Thus, H has to be generated by all the elements of the form $a^k ba^{-k}$ for $k \in \mathbb{Z}$. Since the given relation $b^{-1}a^m b = a^m$ implies $a^m ba^{-m} = b$, we can deduce that H has to be generated by elements of the form

$$a^k ba^{-k}$$

for $k \in \{0, \ldots, m-1\}$. Next, we need to show that H is a free group. In order to do so, consider the Bass-Serre tree T on which $BS(m, m)$ and therefore its subgroup H acts by left multiplication. Recall by (29) in the proof of Claim 3 of the proof of Theorem 3.12 that the G-vertex stabilizer of a vertex $g\langle a\rangle \in V(T)$ is given by $G_{g\langle a\rangle} = \{ga^p g^{-1} \mid p \in \mathbb{Z}\}$. Hence, by intersecting H with $G_{g\langle a\rangle}$, we see that there is no non-trivial element fixing any vertex $g\langle a\rangle \in V(T)$. Thus, H acts freely on the tree T and is therefore by Theorem 2.5 a free group.
Furthermore, we know by Claim 4 of the proof of Theorem 3.12 that the center of $BS(m, m)$ is precisely $\langle a^m\rangle$, which is isomorphic to $\mathbb{Z}$. So we can consider the direct product $H \times \langle a^m\rangle \cong H \times \mathbb{Z}$. We know by Theorem 3.1 that as a free group H is residually finite and that $\mathbb{Z}$ is residually finite. Hence, by Example 1.10, their direct product is residually finite. If we can lastly show that $H \times \langle a^m\rangle$ has finite index in $BS(m, m)$, we can deduce by Theorem 1.17 (ii) that $BS(m, m)$ is residually finite as well. Thus we want to show that $H \times \langle a^m\rangle$ is isomorphic to the kernel of the map

$$\begin{aligned} \psi : BS(m,m) &\to \mathbb{Z}/m\mathbb{Z} \\ a^k &\mapsto k + m\mathbb{Z} \\ b^k &\mapsto m\mathbb{Z}\,, \end{aligned}$$

which is of index m.

We know that $H \times \langle a^m \rangle$ is the special case of the free product $H * \langle a^m \rangle$ where H and $\langle a^m \rangle$ commute. And they commute indeed, since by the relation any element in $\langle a^m \rangle$ commutes with all $g \in BS(m,m)$, hence in particular with all $h \in H \leq G$. So we can think of any element $x \in H \times \langle a^m \rangle$ as an element of the form $x = ha^{lm}$ with $h \in H$ and $l \in \mathbb{Z}$.
So let first $x \in H * \langle a^m \rangle$. As the generators of H are of the form $a^k ba^{-k}$ with $k \in \{0, \dots, m-1\}$, they are all sent to $m\mathbb{Z}$ by ψ, as well as all elements a^{ml} for $l \in \mathbb{Z}$. So clearly, $x \in ker(\psi)$. Assume conversely that $x \in ker(\psi) \subseteq BS(m,m)$. Since $BS(m,m)$ is the HNN-extension of the infinite cyclic group $\langle a \rangle$ with associated subgroup $\langle a^m \rangle$ along the identity map, we know by the Normal Form Theorem 2.17 that x can be written as

$$x = a^{e_0} b^{\pm 1} a^{e_1} b^{\pm 1} \cdots a^{e_n}$$

with $e_0 \in \{0, 1, \dots, m-1\}$ and $e_i \in \{1, \dots, m-1\}$ for all $i \in \{1, \dots n\}$. But since $x \in ker(\psi)$, the exponent sum of the a's has to be divisible by m, i.e.

$$\sum_{i=0}^{n} e_i = 0 \quad mod\, m\,.$$

Now we apply the following: We start with $i = 0$ to insert in the subword $a^{e_i} b^{\pm 1} a^{e_{i+1}}$ the trivial element $1 = a^{-e_i} a^{e_i}$ to obtain

$$a^{e_i} b^{\pm 1} a^{-e_i} a^{e_i} a^{e_{i+1}}\,.$$

In case $e_i + e_{i+1} \geq m$, rewrite

$$a^{e_i} b^{\pm 1} a^{-e_i} a^{e_i} a^{e_{i+1}} = a^{e_i} b^{\pm 1} a^{-e_i} a^{e_i + e_{i+1}} = a^{e_i} b^{\pm 1} a^{-e_i} a^m a^{e_i + e_{i+1} - m}\,.$$

We see that this subsequence now starts with a generating element of H, possibly followed by an element of $\langle a^m \rangle$. To continue inductively with increasing i, if necessary we lastly substitute the original index e_{i+1} by $e_i + e_{i+1} - m$. So we get a sequence of subwords in H and $\langle a^m \rangle$. This works also for the suffix: We need to ensure that it is in $H * \langle a^m \rangle$ as well. But it is clear that our alterations do not change the divisibility of the exponent sum of the a's by m. Hence when entering the case $i+1 = n$ we will always obtain that $a^{e_{n-1}+e_n} \in \langle a^m \rangle$. Thus we have a word in elements in H and $\langle a^m \rangle$. So we conclude $x \in H * \langle a^m \rangle$, i.e. with the commuting property of $\langle a^m \rangle$ we obtain $ker(\psi) \cong H \times \langle a^m \rangle$ and $BS(m,m)$ is residually finite.
Let now $n = -m$, i.e. consider the group $BS(m,-m) = \langle a, b \mid b^{-1} a^m b = a^{-m} \rangle$.

Define ϕ to be the homomorphism induced by

$$\begin{aligned} \phi : BS(m,-m) &\to \mathbb{Z}/2\mathbb{Z} \\ a &\mapsto 0 \\ b &\mapsto 1\,. \end{aligned}$$

Let K be the kernel of ϕ, which is an index-2 subgroup of $BS(m,-m)$. K is given by the presentation

$$\langle a, x, y \mid a^m = x^m, y^{-1}a^m y = a^m, y^{-1}x^m y = x^m \rangle\,,$$

for $x = ba^{-1}b^{-1}$ and $y = b^2$. But K is isomorphic to a subgroup K' of $BS(m,m)$, given by the same homomorphism and presentation as K, where we define $x = bab^{-1}$ instead. Now we can consider the intersection of H with K', which has to be a finite index subgroup of K', since H is of finite index. Also, it has to be of the form $F \times \mathbb{Z}$, with F a free group, since H is of that form. But since K is isomorphic to K', there is a finite index subgroup of the form $F \times \mathbb{Z}$ in $BS(m,-m)$ as well. Hence, by the same argument as in the case $m = n$ before, $BS(m,-m)$ is residually finite.

□

Lemma 3.18. *Let m and n be integers such that $|m| \neq |n|$, $|m| \neq 1$, $|n| \neq 1$ and they do not have the same set of prime factors, i.e. $\pi(m) \neq \pi(n)$, then $G \coloneqq BS(m,n)$ is not Hopfian and thus not residually finite.*

Proof. [32, Lemma 2.1] Assume that m and n do not have the same set of prime factors, i.e. $\pi(m) \neq \pi(n)$. That means, without loss of generality there is a prime $p \in \mathbb{P}$ such that $p \mid m$ but $p \nmid n$. Define the map

$$\begin{aligned} \eta : G &\to G \\ a &\mapsto a^p \\ b &\mapsto b\,, \end{aligned}$$

which clearly induces a homomorphism. Furthermore we can check that it is surjective: Since b gets mapped to b by η anyways, we only need to make sure that there is an element $g \in G$ that gets mapped to the second generator $a \in G$ by η as well. Therefore observe that $gcd(n,p) = 1$, so by *Bézout's Identity*, there are integers $x, y \in \mathbb{Z}$ such that $xn + yp = 1$. Now consider the image of the element $(b^{-1}a^{\frac{m}{p}}b)^x a^y \in G$ under η:

$$\begin{aligned}\eta((b^{-1}a^{\frac{m}{p}}b)^x a^y) &= \eta(b^{-1}a^{\frac{m}{p}}b)^x\,\eta(a)^y \\ &= (b^{-1}a^m b)^x a^{yp} = a^{nx}a^{yp} \\ &= a^{xn+yp} \\ &= a^1 = a\end{aligned}$$

Hence, η is an epimorphism. To show that G is not Hopfian we need to check that η is not injective, i.e. there is a non-trivial element in $ker(\eta)$. Therefore we consider the element $[a^{\frac{m}{p}}, b]^p b^{m-n} \in G$. By the Normal form theorem 2.17

$$[a^{\frac{m}{p}}, b]^p b^{m-n} \neq 1_G\,,$$

but

$$\begin{aligned}\eta([a^{\frac{m}{p}}, b]^p b^{m-n}) &= [a^m, b]^p a^{p(m-n)} \\ &= (a^{-m}b^{-1}a^m b)^p b^{p(m-n)} \\ &= (a^{-m}a^n)^p a^{p(m-n)} \\ &= a^{-p(m-n)+p(m-n)} \\ &= 1_G\,.\end{aligned}$$

Hence G cannot be Hopfian and thus by Malcev's Theorem 1.20 not residually finite. ☐

Remark 3.5. Note that even without Proposition 1.20 at hand, the claim in the above proof of Lemma 3.18 that $G = BS(m,n)$ is not residually finite is an easy check: The kernel of any surjective endomorphism of a finitely generated group has to be contained in every normal subgroup of finite index of that group. Hence, the non-trivial element $[a^{\frac{m}{p}}, b]^p b^{m-n}$ has to be contained in every normal subgroup of finite index, contradicting the condition of residual finiteness.

Lemma 3.19. *Let m and n be integers such that $|m| \neq |n|$, $|m| \neq 1$ and $|n| \neq 1$. Furthermore let them have the same set of prime factors, i.e. $\pi(m) = \pi(n)$ and let it contain more than one prime, i.e. $|\pi(m)| = |\pi(n)| > 1$. Then $G := BS(m,n)$ is not residually finite.*

Proof. [32, Lemma 2.1] Let $\pi(m) = \pi(n)$ with $|\pi(m)| = |\pi(n)| > 1$. Then there has to be a common divisor $k \in \mathbb{Z} \setminus \{0\}$ of m and n, such that $|k| \neq |m|, |n|$ and $\pi(\frac{m}{k}) \neq \pi(\frac{n}{k})$. Now we can consider $BS(\frac{m}{k}, \frac{n}{k})$, to argue as before in the proof of Lemma 3.18 that it cannot be residually finite. But as $BS(\frac{m}{k}, \frac{n}{k})$ is isomorphic to the subgroup of $BS(m,n)$ generated by a^k and b, we conclude with Proposition 1.17 that $BS(m,n)$ cannot be residually finite either. $\square$

If $\pi(m) = \pi(n)$ and $|\pi(m)| = |\pi(n)| = 1$, so m and n are powers of the same prime, we use:

Lemma 3.20. *Let m and n be integers such that $m \mid n$ or $n \mid m$ and $|m| \neq |n|$ as well as $|m| \neq 1$ and $|n| \neq 1$. Then $G := BS(m,n)$ is not residually finite.*

Proof. [32, Lemma 2.2] Assume without loss of generality that $m \mid n$, i.e. there is an $h \in \mathbb{Z} \setminus \{0\}$ such that $n = h \cdot m$. Next, recall from Example 2.11 that the group

$$K := \langle x, y, z \mid x^{-1}zx = y^{-1}zy = z^h \rangle$$

is not Hopfian and thus not residually finite for $h = 2$. Analogously, as indicated in and Remark 2.8, we can check that the group is not Hopfian and thus not residually finite for any $h \in \mathbb{Z} \setminus \{-1, 0, 1\}$. Consider the map

$$\begin{aligned} \mu : K &\to K \\ x &\mapsto x \\ y &\mapsto y \\ z &\mapsto z^h\,, \end{aligned}$$

which clearly extends to a homomorphism. Again we can find a by the Normal Form Theorem 2.17 non-trivial element, namely again $[x^{-1}y, z] \in G$ that has trivial image under μ:

$$\begin{aligned} \mu([x^{-1}y, z]) &= \mu((x^{-1}y)^{-1}z^{-1}x^{-1}yz) \\ &= \mu(y^{-1}xz^{-1}x^{-1}yz) \\ &= y^{-1}xz^{-h}x^{-1}yz^h \\ &= y^{-1}xx^{-1}z^{-1}xx^{-1}yz^h \\ &= y^{-1}z^{-1}yy^{-1}zy \\ &= 1_K\,. \end{aligned}$$

Now we can make the substitution

$$x \mapsto b, \quad y \mapsto a^{-1}ba \text{ and } \quad z \mapsto a^m$$

and see that we obtain the group

$$\langle a, b, a^m \mid b^{-1}a^mb = (a^{-1}ba)^{-1}a^ma^{-1}ba = (a^m)^h \rangle\,,$$

which is clearly via Tietze transformations isomorphic to G. Hence, the element

$$[x^{-1}y, z] \mapsto [b^{-1}a^{-1}ba, a^m] = [[a, b], a^m]$$

is a non-trivial element in the kernel of an epimorphism as well. Thus G cannot be Hopfian and therefore not residually finite. $\square$

Lemma 3.21. *Let m and n be integers such that $|m| \neq 1$, $|n| \neq 1$, $|m| \neq |n|$ and $\pi(m) = \pi(n)$, then $G = BS(m, n)$ is Hopfian.*

Proof. [46, Proposition 4.9] Let $\varphi : G \to G$ be a surjective endomorphism. We want to show that it is injective. By the relation, we know that a^m and a^n are conjugates. Hence, so are $\varphi(a)^m$ and $\varphi(a)^n$.
First we claim that $\varphi(a)$ is conjugate to a power of a. Observe that $\varphi(a)$ is conjugate to an element $u \in G$, that is cyclically reduced, i.e. for which any cyclic permutation of the word is reduced as well. Thus, there is a $g \in G$ such that $\varphi(a) = g^{-1}ug$. As $u \in G$, by the Normal Form Theorem 2.16 it can be written as

$$u = a^{h_0}b^{\pm 1}a^{h_1} \cdots b^{\pm}a^{h_q} .$$

Moreover, with the relation we obtain

$$\begin{aligned} & \varphi(b)^{-1}\varphi(a)^m\varphi(b) = \varphi(a)^n \\ \Rightarrow \qquad & \varphi(b)^{-1}g^{-1}u^m g\varphi(b) = g^{-1}u^n g \\ \Rightarrow \qquad & g\varphi(b)^{-1}g^{-1}u^m g\varphi(b)g^{-1} = u^n . \end{aligned}$$

Hence, u^m and u^n are conjugates. By Collin's Lemma [29, p.185f, Theorem 2.5] this implies that u^m and u^n must have the same G-length. But since u is cyclically reduced, the G-length of $u^j = q|j|$ for every $j \in \mathbb{Z}$. Thus $q|m| = q|n|$, which implies with the assumption $|m| \neq |n|$ that $q = 0$. Thus u is a power of a. Hence, $\varphi(a)$ is conjugate to a^k for some $k \in \mathbb{Z}$. Since if necessary we could compose φ with an inner automorphism and still have a surjective endomorphism, we can without loss of generality assume that $\varphi(a) = a^k$ with $k \in \mathbb{N}$.
Next we want to show Lemma 1.2 of [9], which states that $gcd(k, m, n) = 1$. As φ is a surjective endomorphism we have $G = \langle \varphi(a), \varphi(b) \rangle$. Thus by the Normal Form Theorem 2.16 we can write a as

$$a = \varphi(a)^{h_0}\varphi(b)^{\pm 1} \cdots \varphi(b)^{\pm}\varphi(a)^{h_r} .$$

As $\varphi(a) = a^k$ we obtain

$$a = a^{kh_0}\varphi(b)^{\pm 1} \cdots \varphi(b)^{\pm 1}a^{kh_r} . \tag{31}$$

On the right hand side we have exponent sum $k\left(\sum_{i=0}^{r h_i}\right)$ and since G is not cyclic, we know $k \neq 0$. Now suppose that $gcd(k, m, n) \neq 1$, i.e. there is a prime $p \in \mathbb{P}$ such that $p \mid gcd(k, m, n)$. If there are two words u and v in G written in a and b with $u =_G v$, then their exponent sums have to be congruent modulo p. But the left hand side in (31) has exponent sum 1, while the right hand side has $k\left(\sum_{i=0}^{r} h_i\right)$. As p divides k we obtain

$$1 = 0 \quad mod\, p\,,$$

which is a contradiction. Hence we have $gcd(k, m, n) = 1$.
By assumption $\pi(m) = \pi(n)$, thus k is coprime to m as well as to n. Now we want to show that $\varphi(b)$ is of the form $a^{j_1} b a^{j_2}$ for some $j_1, j_2 \in \mathbb{Z}$. Again, by the Normal Form Theorem 2.16 we know that we can write $\varphi(b)$ as

$$\varphi(b) = a^{h_0} b^{\varepsilon_1} a^{h_1} \cdots b^{\varepsilon_r} a^{h_r}\,.$$

By the relation we get

$$\varphi(b)^{-1}\varphi(a)^m\varphi(b) = \varphi(a)^n \quad \Rightarrow \quad \varphi(b)^{-1} a^{km} \varphi(b) = a^{kn}\,.$$

Inserting the normal form of $\varphi(b)$ gives

$$\begin{aligned}
& (a^{h_0} b^{\varepsilon_1} a^{h_1} \cdots b^{\varepsilon_r} a^{h_r})^{-1} a^{km} (a^{h_0} b^{\varepsilon_1} a^{h_1} \cdots b^{\varepsilon_r} a^{h_r}) a^{-km} = 1_G \\
\Rightarrow \quad & a^{-h_r} b^{-\varepsilon_r} \cdots b^{-\varepsilon_1} a^{-h_0} a^{km} a^{h_0} b^{\varepsilon_1} a^{h_1} \cdots b^{\varepsilon_r} a^{h_r} a^{-km} = 1_G \\
\Rightarrow \quad & a^{-h_r} b^{-\varepsilon_r} \cdots b^{-\varepsilon_1} a^{km} b^{\varepsilon_1} a^{h_1} \cdots b^{\varepsilon_r} a^{h_r} a^{-km} = 1_G\,.
\end{aligned}$$

By Britton's Lemma 2.17, this expression is reducible. As $\varphi(b)$ was given in normal form, we cannot reduce it further. Hence we know that $b^{-\varepsilon_1} a^{km} b^{\varepsilon_1}$ is reducible. This implies that $\varepsilon_1 = 1$ and $b^{-\varepsilon_1} a^{km} b^{\varepsilon_1} = b^{-1} a^{km} b^1 = a^{kn}$. Thus we have the expression $b^{-\varepsilon_2} a^{kn} b^{\varepsilon_2}$ that is reducible. We obtain $\varepsilon_2 = -1$ and $b^{-\varepsilon_2} a^{kn} b^{\varepsilon_2} = a^{km}$. Continuing this procedure gives us that r is odd and that $\varepsilon_i = (-1)^{i+1}$. Hence $\varphi(b)$ is of the form

$$\varphi(b) = a^{h_0} b a^{h_1} b^{-1} \cdots b^{-1} a^{h_{r-1}} b a^{h_r}\,. \tag{32}$$

Applying the Normal Form Theorem 2.16 to b gives that it is of the form

$$b = \varphi(a)^{n_0} \varphi(b)^{\epsilon_1} \varphi(a)^{n_1} \cdots \varphi(a)^{n_s} = a^{kn_0} \varphi(b)^{\epsilon_1} a^{kn_1} \cdots a^{kn_s}\,.$$

Suppose that s is minimal. Assume that $s > 1$ and replace $\varphi(b)$ by its normal form (32). Again, by Britton's Lemma (2.17), we are able to reduce the right hand side. Thus, there has to be a sequence of the form

$$b^{-\varepsilon}a^d b^{\varepsilon}$$

with either $\varepsilon = 1$ and $n \mid d$ or $\varepsilon = -1$ and $m \mid d$. But since we know $\varepsilon_1 = \varepsilon_r = 1$, we deduce for a given r that $\epsilon_t = -\epsilon_{t+1}$ and $m \mid kn_t$. But since $gcd(k,m) = 1$, we conclude $m \mid n_t$, thus there is an $x \in \mathbb{Z}$ such that $n_t = mx$. We obtain

$$\varphi(b)^{\epsilon_t}a^{kn_t}\varphi(b)^{-\epsilon_t} = \varphi(b)^{\epsilon_t}a^{kmx}\varphi(b)^{-\epsilon_t} = a^{knx}\,,$$

which means that b is reducible in contradiction to the assumption that s is minimal. Thus $s = 1$ and $b = a^{kn_0}\varphi(b)a^{kn_1}$, which is equivalent to

$$\varphi(b) = a^{-kn_0}ba^{-kn_1}\,.$$

It remains to show that φ is injective. Therefore we suppose that it is not, i.e. that there is a non-trivial element v in the kernel of φ. Write v in Normal Form as

$$v = a^{l_0}b^{\varepsilon_1}\cdots b^{\varepsilon_z}a^{l_z}$$

with $z \geq 1$. Then we obtain for its image under φ

$$\begin{aligned}
1_G &= \varphi(v)\\
&= \varphi(a)^{l_0}\varphi(b)^{\varepsilon_1}\cdots\varphi(b)^{\varepsilon_z}\varphi(a)^{l_z}\\
&= a^{kl_0}(a^{-kn_0}ba^{-kn_1})^{\varepsilon_1}\cdots(a^{-kn_0}ba^{-kn_1})^{\varepsilon_z}a^{kl_z}\,.
\end{aligned}$$

Again, by Britton's Lemma 2.17 this expression is reducible, which means there is a reducible subsequence, for example $b^{-1}a^{mj}b$ with $b^{-1}a^{l_i}b$ a subsequence of its preimage. Thus, we have

$$\varphi(b^{-1}a^{l_i}b) = a^{kn_1}b^{-1}a^{kl_i}ba^{-kn_1}$$

where $kl_i = mj$. Since k and m are coprime, we have $m \mid l_i$. But v is given in normal form, hence $0 \leq l_i < |m|$, which is a contradiction. So we have that φ is injective and hence G is Hopfian. ☐

This concludes the proof of Proposition 3.15. ☐

3.3 Groups Without Finite Quotients

One way of finding an example for a non-residually finite group, is to find an infinite group that does not have any finite quotients, since then the property of residual finiteness trivially cannot be satisfied.

3.3.1 Higman-Group

The first and most well-known example of such a group was introduced by Graham Higman in 1951 [16]:

Proposition 3.22. *The group*

$$H = \langle a, b, c, d \mid [b, a] = b, [c, b] = c, [d, c] = d, [a, d] = a \rangle \,,$$

where $[x, y] \coloneqq x^{-1}y^{-1}xy$ *defines the commutator, is infinite, not simple and does not have any finite quotients.*

The absence of finite quotients was shown in [38], the rest of the proof is guided by [41].

Proof. First we need to observe that H is in fact an amalgamated product constructed the following way:

- Define the groups $H_i \coloneqq \langle a_i, b_i \mid a_i^{-1} b_i a_i = b_i^2 \rangle$ for $i \in \{1, 2, 3, 4\}$.
- Form amalgamated products K_{12} and K_{34}: Identify the isomorphic subgroups $\langle a_2 \rangle \cong \langle b_1 \rangle$ of H_1 and H_2 to obtain

  $$\begin{aligned} K_{12} &= H_1 *_{\langle a_2 \rangle \cong \langle b_1 \rangle} H_2 \\ &= \langle a_1, a_2, b_2 \mid a_1^{-1} a_2 a_1 = a_2^2, a_2^{-1} b_2 a_2 = b_2^2 \rangle \end{aligned}$$

  and analogously $\langle b_3 \rangle \cong \langle a_4 \rangle$ of H_3 and H_4 to obtain

  $$\begin{aligned} K_{34} &= H_3 *_{\langle b_3 \rangle \cong \langle a_4 \rangle} H_4 \\ &= \langle a_3, a_4, b_4 \mid a_3^{-1} a_4 a_3 = a_4^2, a_4^{-1} b_4 a_4 = b_4^2 \rangle \end{aligned}$$

- Form the amalgamated product K_{1234}: Identify the isomorphic subgroups $\langle a_1, b_2 \rangle \cong \langle a_3, b_4 \rangle$ of K_{12} and K_{34} via $b_2 = a_3$ and $a_1 = b_4$ to obtain

  $$K_{1234} = K_{12} *_{\langle a_1, b_2 \rangle \cong \langle a_3, b_4 \rangle} K_{34} =$$
  $$= \langle a_1, a_2, a_3, a_4 \mid a_1^{-1} a_2 a_1 = a_2^2, a_2^{-1} a_3 a_2 = a_3^2, a_3^{-1} a_4 a_3 = a_4^2, a_4^{-1} a_1 a_4 = a_1^2 \rangle$$

- Lastly observe that clearly $G \cong K_{1234}$.

Now we can check the three properties of G using the identity $G \cong K_{1234}$:

(i). G has no non-trivial finite quotient: Suppose that $N \trianglelefteq G$ is a normal subgroup of G with $|G/N| < \infty$. Then we need to deduce that $G/N = \{1_{G/N}\}$. So, since the relations of G have to hold in G/N as well, we need to deduce that $a_1N = a_2N = a_3N = a_4N = 1_{G/N}$. For that we use the following lemma:

Lemma 3.23. *Let $n \in \mathbb{N} \setminus \{1\}$, let $q \in \mathbb{P}$ be the smallest prime factor of n and let $p \in \mathbb{P}$ be the smallest prime factor of $2^n - 1$. Then $q < p$.*

Proof. Define $r := min\{k \in \mathbb{N} \mid p \mid 2^k - 1\}$. If $r \neq 1$, then there is a $q \in \mathbb{P}$ such that $q \mid r$. Furthermore, by Fermat's Little theorem, we know that $p \mid 2^{p-1} - 1$, hence $r \mid p - 1$. Thus we deduce that $q \leq p - 1$, and therefore $q < p$. But by the properties of Mersenne numbers, $r \mid n$, so with $q \mid r$ we deduce $q \mid n$, and the desired result is proven. ☐

Now we assume that the relations of $G \cong K_{1234}$ hold and that there is at least one $i \in \{1, 2, 3, 4\}$ such that $a_iN \neq 1_{G/N}$. Show that $a_1^{-n}a_2a_1^n = a_2^{2^n}$ by using $a_1^{-1}a_2 = a_2^2a_1^{-1}$ and induction on $n \in \mathbb{N}$:
For the case $n = 1$ the formula is apparent from the group relation. Now assume that $a_1^n a_2 a_1^n = a_1^{2^n}$ and show that $a_1^{-n-1}a_2a_1^{n+1} = a_2^{2^{n+1}}$ holds:

$$\begin{aligned}
a_1^{-n-1}a_2a_1^{n+1} &= a_1^{-1}a_1^{-n}a_2a_1^na_1 \\
&\overset{IH}{=} a_1^{-1}a_2^{2^n}a_1 \\
&= a_1^{-1}a_2a_2^{2^n-1}a_1 \\
&= a_2^2a_1^{-1}a_2^{2^n-1}a_1 \\
&= a_2^2a_1^{-1}a_2a_2^{2^n-2}a_1 \\
&= a_2^2a_2^2a_1^{-1}a_2^{2^n-2}a_1 \\
&= \cdots = \underbrace{a_2^2a_2^2\cdots a_2}_{2^n \text{ times}}a_1^{-1}a_1 \\
&= (a_2^2)^{2^n} = a_2^{2^{n+1}}.
\end{aligned}$$

Now define the order of the images of the generators a_i in G/N as $ord(a_iN) =: n_i$ and use in the above formula $n = n_1$ to obtain

$$\begin{aligned}
& a_1^{-n_1}a_2a_1^{n_1}N = a_2^{2^{n_1}}N \\
\Rightarrow \quad & a_1^{-n_1}N\,a_2N\,a_1^{n_1}N = a_2^{2^{n_1}}N \\
\Rightarrow \quad & (a_1N)^{-n_1}\,a_2N\,(a_1N)^{n_1} = a_2^{2^{n_1}}N \\
\Rightarrow \quad & a_2N = a_2^{2^{n_1}}N \\
\Rightarrow \quad & N = a_2^{2^{n_1}-1}N\,.
\end{aligned}$$

So we can conclude n_2 has to be a divisor of $2^{n_1} - 1$, hence if $n_1 = 1$, then $n_2 = 1$ and therefore $n_3 = n_4 = 1$ in contradiction to our assumption. Thus we can assume that $n_1 > 1$. Consider the smallest prime factor $p \in \mathbb{P}$ of the product $n_1 n_2 n_3 n_4$. Since the relations are symmetric, we can without loss of generality assume that $p \mid n_2$. But since $n_2 \mid 2^{n_1} - 1$, Lemma 3.23 implies that there is a prime $q \in \mathbb{P}$ such that $q \mid n_1$ and $q < p < n_2$ in contradiction to the minimality of p. So $n_1 \ngtr 1$ and therefore $n_1 = n_2 = n_3 = n_4$ and $a_1 N = a_2 N = a_3 N = a_4 N = 1_{G_N}$. Thus there is no proper normal subgroup $N \trianglelefteq G$ with finite index, so G does not have any proper finite quotient, and is therefore not residually finite.

(ii). G is infinite: By the construction of G via the amalgamated product we know that the element a_1 in H_1 has order $ord_{H_1}(a_1) = \infty$, so the element a_1 is a generator of K_{1234}, hence $\langle a_1 \rangle$ is an infinite cyclic subgroup of G, which implies that G has to be infinite as well.

(iii). G is not simple: We know that G is isomorphic to the amalgamated product $(K_{12} * K_{34})/M$, where M is the subgroup isomorphic to the normal closure of $\langle a_1, b_2 \rangle$. Now we consider the two groups

$$K := \langle a_1, a_2, b_2 \mid a_1^{-1} a_2 a_1 = a_2^2, a_2^{-1} b_2 a_2 = b_2^2, (a_1 b_2)^3 = (b_2 a_1)^3 \rangle$$
$$K' := \langle a_3, a_4, b_1 \mid a_3^{-1} a_4 a_3 = a_4^2, a_4^{-1} b_4 a_4 = b_4^2, (a_3 b_4)^3 = (b_4 a_3)^3 \rangle$$

and take the amalgamated product $(K * K')/M =: \tilde{K}$ with M as above. Define the group homomorphism

$$\begin{aligned} \varphi : G &\to \tilde{K} \\ a_i &\mapsto a_i \quad . \end{aligned}$$

We know that $(a_1 b_2)^3 (b_2 a_1)^{-3} \neq 1_G$, but

$$\varphi((a_1 b_2)^3 (b_2 a_1)^{-3}) = (a_1 b_2)^3 (b_2 a_1)^{-3} = 1_{\tilde{K}} \, ,$$

so $(a_1 b_2)^3 (b_2 a_1)^{-3} \in ker(\varphi)$ and φ is not injective. So the kernel of φ is a non-trivial normal subgroup of G and therefore G is not simple, if we can exclude the case that $ker(\varphi) = G$, which means $Im(\varphi) \neq \{1_G\}$ which is equivalent to $\tilde{K} \neq \{1_{\tilde{K}}\}$. But that $\tilde{K}$ is non-trivial is an easy check with GAP.

$\square$

3.3.2 Discrete Groups

The following two examples as constructed in [10] are of groups without any finite quotients. Therefore they are trivially not residually finite. We will only focus on the arguments essential for proving the non-existence of finite quotients:

Example 3.2. There is an infinite discrete subgroup $G \subseteq Isom(\mathbb{H}^3)$ of the isomorphism group of the hyperbolic space $\mathbb{H}^3$ without any finite quotient.

Proof. We want to give an algebraic description of a group and show that it has no finite quotients. Let T be the maximal rooted binary tree. Denote the root with $*$ and let $|v| := d(v, *)$ denote the length of the path from any vertex $v \in V(T)$ to the root $*$. Let the edges of T be oriented towards the root $*$ and denote the terminal vertex of any edge $e \in E(T)$ by $t(e)$. Lastly denote for any vertex $v \in V(T)$ the outgoing edge by $e_o(v)$ and the incoming edges by $e_1(v)$ and $e_2(v)$. (Note that any vertex can have at most two incoming and one outgoing edges as the tree T is binary.)

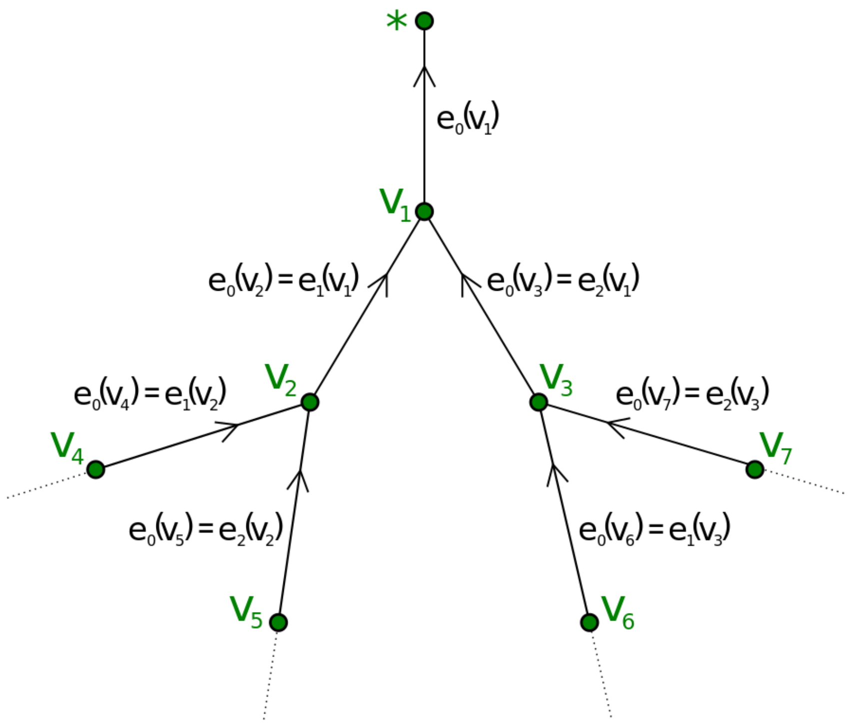


Figure 9: *This illustrates the first seven vertices of the infinite, maximal rooted binary T with root $*$. The two incoming edges of a vertex are labelled as e_1 and e_2, the outgoing edge is denoted as e_0.*

Now define a set $G_E := \{g_e \mid e \in E\}$, indexed by the edges in $E(T)$ and the group

$$G \coloneqq \langle G_E \mid \{g_{e_o(v)}^{3+|v|} = 1_G, g_{e_o(v)}g_{e_1(v)}^{-1}g_{e_2(v)}^{-1} = 1_G \mid v \in V \text{ with } |v| \geq 1\}\rangle \, .$$

Then G is the fundamental group $\pi_1(\mathcal{T})$ of the graph of groups $\mathcal{T}$ with underlying graph T. The vertex groups of $V(\mathcal{T})$ are for $v \in V(T) \setminus \{*\}$ given by

$$G_v = \langle g_{e_o(v)}, g_{e_1(v)}, g_{e_2(v)} \mid g_{e_1(v)}^{4+|v|} = 1, g_{e_2(v)}^{4+|v|} = 1, g_{e_o(v)}g_{e_1(v)}^{-1}g_{e_2(v)}^{-1} = 1\rangle$$

and

$$G_* = \langle g_{e_1(*)}, g_{e_2(*)} \mid g_{e_1(*)}^4 = 1, g_{e_2(*)}^4 \rangle \, .$$

The edge groups of $E(\mathcal{T})$ are given by

$$G_e = \langle g_e \mid g_e^{4+|t(e)|} \rangle \, .$$

Now we want to think about G as a nested union of a sequence of subgroups of G. Therefore we describe the subgroup $G^n \leq G$ for $n \in \mathbb{N}$ as the the fundamental group

$$G^n \coloneqq \pi_1(\mathcal{T}^n)$$

where $\mathcal{T}^n \subseteq \mathcal{T}$ is the subgraph of T corresponding to the ball of radius $n-1$ around the root $*$. So equivalently

$$G^n = \langle g_{e_o(v)} \mid v \in V(T) \text{ with } |v| \leq n \rangle \, .$$

By construction we have

$$G^1 \subseteq G^2 \subseteq G^3 \subseteq \dots$$

and

$$G = \bigcup_{n \in \mathbb{N}} G^n \, .$$

Now we can show:

Claim 1: G^n is generated by $S^n \coloneqq \{g_{e_o(v)} \mid |v| = n\}$ for any $n \in \mathbb{N}$.

We know that by construction G^n is generated by $S^1 \cup S^2 \cup \dots \cup S^n$, hence it is generated by $G^{n-1} \cup S^n$ for any $n \in \mathbb{N}$. So we need to show that $G^{n-1} \subseteq \langle S^n \rangle$. In order to do so we use induction on $n \in \mathbb{N}$:

Since the base case is trivial, let $v \in V$ with $|v| = n-1$ and define for $e_1(v)$ and $e_2(v)$ the initial vertices as v_1 and v_2 respectively, i.e. $e_o(v_1) = e_1(v)$ and

$e_o(v_2) = e_2(v)$. But we know by the relations in G that

$$\begin{aligned} g_{e_o(v)} &= g_{e_1(v)} g_{e_2(v)} \\ &= g_{e_o(v_1)} g_{e_o(v_2)} \in S^n \end{aligned}$$

and $g_{e_o(v_1)} g_{e_o(v_2)} \in S^n$ since $|v_1| = |v_2| = |v| + 1 = n$. So any element in G^{n-1} is also in $\langle S^n \rangle$.

This enables us to show:

Claim 2: G has no finite quotients.

Let H be a finite group and let $\pi : G \to H$ be a homomorphism. If π is trivial, that is $\pi(g) = 1_H$ for any $g \in G$, then the claim is true. So let $k, n \in \mathbb{N}$ such that $n > k$ and $3 + n \equiv 1 \pmod{|H|}$ and note that for any $g_{e_o(v)} \in S^n = \{g_{e_o(v)} \mid |v| = n\}$ we have $g_{e_o(v)}^{3+n} = 1_G$. But as $gcd(|H|, 3+n) = 1$, we obtain $\pi|_{S^n} : S^n \to H$ is trivial, thus $S^n \subseteq ker(\pi)$. Since S^n generates G^n this implies $G^n \subseteq ker(\pi)$, and with $k < n$ we conclude $G^k \subseteq ker(\pi)$. The choice of $k \in \mathbb{N}$ was arbitrary, and $\bigcup_{k \geq 1} = G$, hence $G \subseteq ker(\pi)$. So there is no non-trivial homomorphism $\pi : G \to H$ and therefore also no proper subgroup $K \leq G$ with finite index.

To complete the proof of the statement it is left to show that there is a discrete subgroup of $PSL_2(\mathbb{C})$ isomorphic to G. The proof is omitted, since it would go beyond the scope of the topics of interest. □

Example 3.3. There is a torsion-free discrete subgroup $G \subseteq Isom(\mathbb{H}^3)$ of the isomorphism group of the hyperbolic space $\mathbb{H}^3$ that does not have any finite quotients and is therefore not residually finite.

Proof. It is not possible to explain the construction of the group in question in all details with the tools at hand, as it is the fundamental group of a very special manifold. Therefore we will deal with it only heuristically and again focus on the argument that is crucial for the non-residual finiteness. For the construction we start with a one-holed torus T with two curves α and β on it that intersect exactly in one single point. Then we take a regular neighborhood U of $\beta \times \{0\}$ in the 3-manifold $T \times [-1, 1]$ and define two curves μ and β' with some special properties. Finally for $n \in \mathbb{N}$ we define the manifold M_n that we obtain by Dehn-filling the curve $n\mu + \beta'$. We get a presentation

$$\pi_1(M_n) = \pi_1(T) *_{\langle \beta \rangle} \pi_1(U) \cong \langle a, b, c \mid b = c^n \rangle .$$

The curve $\partial T \times \{0\}$ corresponds to the conjugacy class of $[a, b] = [a, c^n]$. By a claim in [10], $\pi_1(M_n)$ admits a geometrically finite representation

$$\begin{aligned} \rho_n : \pi_1(M_n) &\to PSL_2(\mathbb{C}) \\ [a,b] &\mapsto g \qquad\qquad \text{with } g(z) = z+1\,. \end{aligned}$$

Define for $t \in \mathbb{R}$ the element $h_t \in PSL_2(\mathbb{C})$ as the element $h_t(z) = z + ti$. It is shown in [10] that for $t_n \to \infty$ the group

$$G = \langle \bigcup_{n\in\mathbb{N}} h_{t_n} \rho_n(\pi_1(M_n)) h_{t_n}^{-1} \mid n \in \mathbb{N} \rangle$$

is discrete. Note that as $g(z) = z+1$ and $h_t(z) = z+ti$, we can write any $g \in G$ as

$$g = h_t g h_t^{-1} = h_t \rho_n([a,b]) h_t^{-1} = h_t \rho_n([a,c^n]) h_t^{-1}\,.$$

Now we claim that G is not residually finite: Let H be a finite group and let $\pi : G \to H$ be a homomorphism. If we can show that π is always trivial, then there is no finite quotient and therefore G is not residually finite. So let $g \in G$. Then

$$\begin{aligned} \pi(g) &= \pi(h_t \rho_n([a,c^n]) h_t^{-1}) \\ &= \pi(h_t)[\pi(\rho_n(a)), \pi(\rho_n(c))]\pi(h_t)^{-1} \\ &= \pi(h_t)[\pi(\rho_n(a)), \rho_{|H|}(c)^{|H|}]\pi(h_t)^{-1} \\ &= \pi(h_t)[\pi(\rho_n(a)), 1_H]\pi(h_t)^{-1} \\ &= \pi(h_t)\pi(h_t)^{-1} \\ &= 1_H \quad . \end{aligned}$$

Hence there is no proper subgroup $K \leq G$ with finite index, so G is not residually finite. □

4 Computational Approach: The Heineken Group

The following group will be the main interest and motivation in the forthcoming chapter:

Definition 4.1. The group H, given by the presentation

$$\mathsf{H} = \langle x, y, z \mid [x,[x,y]] = z, [y,[y,z]] = x, [z,[z,x]] = y \rangle\,, \tag{33}$$

where $[a,b] := a^{-1}b^{-1}ab$ denotes the commutator, is called the *Heineken group.*

Originally, the Heineken group H was of interest, because it is a generalization of the simpler presentation

$$\langle x, y, z \mid [x, y] = z, [y, z] = x, [z, x] = y \rangle ,$$

which was known to be trivial by computational methods very early, for example it was stated in [37]. It was investigated with methods that have proven themselves valuable to show that certain groups are finite. However, since the existence of an automatic structure and with it the infiniteness and the hyperbolicity of the Heineken group is proven, the research can be refocused on the question whether the Heineken group is residually finite. Especially, since on the one hand its presentation is similar to the one of the Higman group, which we proved to be non-residual finite by the absence of finite quotients and on the other hand since it turns out to be very hard to find finite quotients, it suggests itself that there might not exist a lot of them. Therefore it could be a feasible goal to determine whether the Heineken group is the first known example of a hyperbolic but non-residually finite group.
We want to retrace the findings about the Heineken group and test and explain different attempts to determine whether the Heineken group is residually finite.

Remark 4.1. With Tietze transformations we also obtain a presentation of the Heineken group H with only two generators, that is in general more useful for computations:

$$\mathsf{H} = \langle x, y \mid [y, [y, [x, [x, y]]]] = x, [[x, [x, y]], [[x, [x, y]], x]] = y \rangle . \tag{34}$$

All the computational concepts used to approach questions about the Heineken group in the forthcoming chapters are guided by Holt's Handbook of Computational Group Theory [19].

4.1 Finiteness & Hyperbolicity

The algorithms described in this chapter are available in GAP on Linux. They have been implemented by Derek Holt in an extra package called "KBMAG" (say "kay-bee-mag"), which stands for "Knuth-Bendix on Monoids and Automatic Groups". It is an interface to some C-programs that can be run stand-alone without GAP as well. However, it is much more convenient and user-friendly to use GAP. The package can be used after inputting

```
gap > LoadPackage("kbmag");
```

In case the package is not stored in the same directory as the GAP roots, we need to use the funtion `SetPackagePath("kbmag","...")` to be able to load the package with the above command.

In order to tackle the problem of Finiteness and Hyperbolicity and explain the algorithms needed, we need to introduce quite a lot of machinery:

4.1.1 Finite State Automata

Definition 4.2. A *finite state automaton (FSA) M* is a quintuple (Σ, A, τ, S, F), where

(i). Σ is a finite set, the set of *states*;

(ii). A is a finite set, the *alphabet*;

(iii). $S \subseteq \Sigma$ is the set of *start states*;

(iv). $F \subseteq \Sigma$ is the set of *final states*;

(v). $\tau \subseteq \Sigma \times A' \times \Sigma$ is the set of *transitions*, with $A' \coloneqq A \cup \{\varepsilon\}$ for $\varepsilon \notin A$.

Remark 4.2. We can think of a transition $(\sigma_1, a, \sigma_2) \in \tau$ as a labelled arrow

$$\sigma_1 \overset{a}{\to} \sigma_2 \,,$$

i.e. the element a sends state σ_1 to state σ_2. We often denote it as $\sigma_1^a = \sigma_2$.

Let us first define all the terminology needed:

Definition 4.3.

(i). A transition of the form $(\sigma_1, \varepsilon, \sigma_2) \in \tau$ for some $\sigma_1, \sigma_2 \in \Sigma$ is called *ε-transition.*

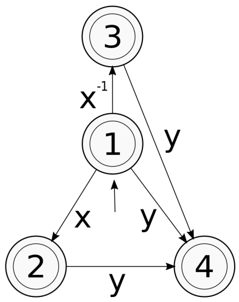
(ii). Let $w = a_1 a_2 \cdots a_r \in (A')^*$, then a sequence $((\sigma_{i-1}, a_i, \sigma_i))_{i \in \{1,\ldots,r\}} \subseteq \tau$ from state σ_0 to state σ_r is called *a path of arrows labelled w.*

(iii). Let $w \in (A')^*$. The word in A^* obtained by deleting all occurrences of ε in w is denoted by $\rho(w)$.

(iv). The *language* $L(M)$ of the FSA M is defined to be the subset of A^* containing $w \in A^*$ if there is an element $w' \in (A')^*$ such that $\rho(w') = w$ and there is a path of arrows labelled w' from a start state $\sigma_0 \in S$ to a final state $\sigma^{w'} \in F$.

Remark 4.3. We often refer to all elements of the language as to be accepted by the FSA and we sometimes refer to the final states as accepted states. States, for which no transition is defined, are often called failure or dead states.

Example 4.1. Let $A = \{x, y, x^{-1}, y^{-1}\}$, let $\Sigma = \{1, 2, 3, 4\}$, let $S = \{1\}$ and let $F = \{1, 2, 3, 4\}$. We usually have two options to illustrate the transitions τ. Either we use a diagram or a transition table:



	x	x^{-1}	y	y^{-1}
1	2	3	4	0
2	0	0	4	0
3	0	0	4	0
4	0	0	0	0

In the diagram the start state is marked by a small unlabelled arrow without a source. Circles represent the states, a double circle marks a final state. In the transition table every state has a row. In the columns it can be read off to which state it is sent by which element of the alphabet. If a transition is not defined, the corresponding spot in the table is filled with a 0.
In the given example the language $L(M)$ of the FSA $M = (\Sigma, A, \tau, S, F)$ is

$$L(M) = \{\varepsilon, x, xy, x^{-1}, x^{-1}y, y\}\,.$$

Definition 4.4. Let M be an FSA.

(i). M is called *(partial) deterministic finite state automaton (DFA)* if

- $|S| \leq 1$;
- There are no ε-transitions ;
- For all $\sigma_0 \in S$ and for all $a \in A$ there is at most one transition $(\sigma_0, a, \sigma) \in \tau$ for some $\sigma \in \Sigma$.

(ii). M is called *complete deterministic finite state automaton (complete DFA)* if

- $|S| = 1$;
- There are no ε-transitions ;
- For all $\sigma_0 \in S$ and for all $a \in A$ there is exactly one transition $(\sigma_0, a, \sigma) \in \tau$ for some $\sigma \in \Sigma$.

Example 4.2. Let $\mathfrak{C}$ be a coset table defined in the upcoming chapter 4.2 in Definition 4.29. Then M with $\Sigma = \Omega$, $S = \{1\}$, $F = \{i\}$ for some $i \in \Omega$, the alphabet given by A and the transitions given by the table $\mathfrak{C}$ defines a DFA. The language $L(M)$ is the subset of the coset Hg corresponding to i. If the coset table is complete, the DFA is complete and $L(M)$ is the whole coset Hg.

If a DFA has the following property, we can determine the cardinality of its language:

Definition 4.5. Let $M = (\Sigma, A, \tau, \{\sigma_0\}, F)$ be a DSA.

(i). A state $\sigma \in \Sigma$ is called *accessible* if there is a word $w \in A^*$ such that $\sigma_0^w = \sigma$ and it is called *coaccessible* if there is a word $w \in A^*$ such that $\sigma^w \in F$.

(ii). M is called *accessible/coaccessible* if all $\sigma \in \Sigma$ are accessible/coaccessible.

(iii). M is called *trim* if it is accessible and coaccessible.

If we have a trim DFA M given, we can determine the cardinality of its language $L(M)$ with the following algorithm, known as FSA-Count:
First observe that $L(M)$ is infinite if M contains a circuit. Suppose M has a circuit, then there is a non-empty word $w \in (A')^*$ and a state $\sigma \in \Sigma$ such that $\sigma^w = \sigma$. But since M is trim, there are also two words $w_1, w_2 \in (A')^*$ such that $\sigma_0^{w_1} = \sigma$ and $\sigma^{w_2} \in F$. Hence the word $w_1 w^n w_2 \in L(M)$ for every $n \in \mathbb{N}$, which implies that $L(M)$ is infinite.
Now we can distinguish three steps:

(i). Compute for every state $\sigma \in \Sigma$ the number δ_σ of arrows with target σ. If $\delta_{\sigma_0} > 0$, then we have a circuit starting and ending in σ_0, hence $L(M)$ has to be infinite.

(ii). Order the states in Σ in a list ζ such that $\zeta[1] = \sigma_0$ and for any arrow $\zeta[i] \xrightarrow{a} \zeta[j]$ we have $i < j$. This is done by first defining $\zeta[1] := \sigma_0$ and then considering $\sigma \in \Sigma$. For every arrow with target σ and some source $\zeta[i]$ for $i \in \{1, 2, \dots\}$ we reduce the number δ_σ by 1. If we obtain $\delta_\sigma = 0$ at some point, we know that all arrows with target σ have a source in the list. So we can append σ to the list. If δ_σ does not reduce to 0 for some $\sigma \in \Sigma$, the list ζ does not contain all elements in Σ. Define this set of not-contained elements to be Γ. Since $\delta_\gamma > 0$ for all $\gamma \in \Gamma$, there is at least one arrow with target $\gamma \in \Gamma$ and its source in Γ as well. Hence there is a circuit within Γ, implying that $L(M)$ is infinite.

(iii). If the list ζ contains all states of Σ, then we have ordered Σ with the property that for every arrow $\zeta[i] \xrightarrow{a} \zeta[j]$ we have $i < j$. Define β_σ to be the number of words $w \in A^*$ such that $\sigma^w \in F$. Clearly, we have for $\sigma \notin F$

$$\beta_\sigma = \sum_{(\sigma, a, \sigma') \in \tau} \beta_{\sigma'}$$

and $\beta_\sigma = 1$ in case $\sigma \in F$. Hence we can run through the list in reverse order and find $|L(M)| = \beta_{\zeta[1]}$.

Now we want to introduce an FSA for groups:

Definition 4.6. Let $G = \langle X \mid R \rangle$ be a group and define $A := X \cup X^{-1}$. An FSA W with alphabet A is called *word acceptor for* G if it accepts at least one word in A^* for every $g \in G$, which means that the language $L(W)$ contains at least one word of A^* for every $g \in G$. If the language $L(W)$ contains precisely one word of A^* for every $g \in G$, then W is called *unique*.

Remark 4.4. Since a unique word acceptor W for the group G has to be trim, we can use it by computing the cardinality of $L(W)$ with FSA-COUNT to determine the cardinality of G.

Hence, if we can find a word acceptor for the Heineken group, we will be able to answer the question for its cardinality. However, so far it is unclear how to find a unique word acceptor for a given group. It will be part of a so called automatic structure, but to be able to define this term, we need to define one last concept: We want to define an FSA that can read two words w_1 and w_2 simultaneously. It suggests itself to simply use $A \times A$ as an alphabet in an FSA, but one problem occurs: What to do when $|w_1| \neq |w_2|$?

Definition 4.7.

(i). Define $A^+ := A \cup \{\$\}$ for $\$ \notin A$ and let $w_1, w_2 \in A^*$ such that $w_1 = a_1 \cdots a_l$ and $w_2 = b_1 \cdots b_m$ with $a_i, b_j \in A$ for all $i \in \{1, \ldots, l\}$ and $j \in \{1, \ldots, m\}$. We define the word $(w_1, w_2)^+ \in (A^+ \times A^+)^*$ to be the word

$$(\alpha_1, \beta_1)(\alpha_2, \beta_2) \cdots (\alpha_n, \beta_n)$$

with $n := \max(l, m)$ and

- $\alpha_i = a_i$ for every $1 \leq i \leq l$ and $\alpha_i = \$$ for all $l < i \leq n$
- $\beta_j = b_i$ for every $1 \leq j \leq m$ and $\beta_j = \$$ for all $m < j \leq n$.

The pair $(w_1, w_2)^+$ is called *padded pair*, $\$$ is called the *padding symbol*.

(ii). A *2-variable FSA* M over the alphabet A is an FSA with alphabet $A^+ \times A^+$ where words in $L(M)$ are of the form of a padded pair $(w_1, w_2)^+$ for $w_1, w_2 \in A^*$.

Example 4.3. Let $A = \{a, b, c, d\}$ and let $w_1 = abc$ and $w_2 = d$. Then the padded pair $(w_1, w_2)^+ \in (A^+ \times A^+)^*$ is given by

$$(w_1, w_2)^+ = (a, d)(b, \$)(c, \$) .$$

4.1.2 Automatic Groups

For the following constructions, we need to interpret groups as monoids, thus recall:

Definition 4.8.

(i). A set M with an associative binary operation and an identity element is called *monoid*.

(ii). Let X be a finite set. The monoid M consisting of all possible words in X^*, i.e.

$$M = \{m \in M \mid m = x_1 \cdots x_r,\, x_i \in X \text{ for all } i \in \{1, \ldots, r\}\}$$

is called *finitely generated by X*.

(iii). Let M and N be two monoids. A *monoid homomorphism* is a map $f : M \to N$ satisfying $f(xy) = f(x)f(y)$ for all $x, y \in M$ and $f(1_M) = 1_N$.

(iv). An equivalence relation $\sim$ on a monoid M where $x \sim y$ implies $xz \sim yz$ and $zx \sim zy$ for all $x, y, z \in M$ is called *congruence*.

Proposition 4.1. *Let $f : M \to N$ be a monoid homomorphisn. Define the equivalence relation $\sim$ as*

$$x \sim y \quad \Leftrightarrow \quad f(x) = f(y)\,.$$

Then $\sim$ is a congruence, called the kernel of f.

Proof. The proof is straightforward. □

Theorem 4.2. *Let $f : M \to N$ be a monoid homomorphism with kernel $\sim$. Then the map*

$$\begin{aligned} \varphi : M/\sim &\to im(f) \\ [x] &\mapsto f(x) \end{aligned}$$

where $M/\sim$ is the quotient structure with elements given by equivalence classes under $\sim$ and the well-defined multiplication given by $[x][y] = [xy]$, is a monoid isomorphism.

Proof. The required properties can be easily checked by using the definitions. □

Definition 4.9.

(i). Let M be a monoid and $\mathcal{R} \subseteq M \times M$. The congruence σ *generated by* $\mathcal{R}$ is the intersection of all congruences on M containing $\mathcal{R}$.

(ii). Let X be a set and $\mathcal{R} \subseteq X^* \times X^*$. Then $Mon\langle X \mid \mathcal{R}\rangle$ is defined to be the monoid given by X^*/σ. If X and $\mathcal{R}$ are finite sets, we call $Mon\langle X \mid \mathcal{R}\rangle$ *finitely presentable.*

Forthcoming, we want to use the following fact:

Theorem 4.3. *[19, p. 416, Theorem 12.12] Let X be a set. Define the set $A := X \cup X^{-1}$ and let $\mathcal{R} \subseteq A^* \times A^*$. The group defined by the presentation $\langle X \mid \mathcal{R}\rangle$ is equal to the monoid defined by the presentation $Mon\langle A \mid \mathcal{I}_X \cup \mathcal{R}\rangle$, where $\mathcal{I}_X := \{(xx^{-1}, \varepsilon_A) \mid x \in A\} \subseteq A^* \times A^*$.*

With that at hand, we can return to groups:

Definition 4.10. Let $G = \langle X \mid R\rangle$ be a group, generated as a monoid by $A := X \cup X^{-1}$. G is called *automatic* if:

(i). There is a word acceptor W for G with alphabet A.

(ii). For every $x \in A \cup \{\varepsilon\}$ there is a 2-variable FSA over the alphabet A such that for every $v, w \in A^*$ we have

$$(v, w)^+ \in L(M_x) \quad \Leftrightarrow \quad v, w \in L(W) \text{ and } vx =_G w\,.$$

The M_x are called *multiplier automata.* The pair $(W, \{M_x\}_{x\in A\cup\{\varepsilon\}})$ is called *automatic structure for G.* In case the word acceptor W is unique, we say $(W, \{M_x\}_{x\in A\cup\{\varepsilon\}})$ is an automatic structure *with uniqueness.*

The definition of *the* automatic structure for the group G makes sense, since:

Proposition 4.4. *[12, p. 52, Theorem 2.4.1] Let G be a group generated as a monoid by the set A_1 and let $(W_1, \{M_{1x}\}_{x\in A\cup\{\varepsilon\}})$ be the corresponding automatic structure. If A_2 is another finite set generating the monoid presentation of the group, then there is an automatic structure $(W_2, \{M_{2x}\}_{x\in A\cup\{\varepsilon\}})$ corresponding to A_2.*

So now, we are aiming to show that the Heineken group is automatic, to be able to determine its cardinality by counting the elements of the language of the word acceptor of the automatic structure. To make matters easier, we first we limit our search to word acceptors only accepting minimal elements:

Definition 4.11. Let $G = \langle X \mid R\rangle$ be a group and define $A := X \cup X^{-1}$.

(i). Let $<_A$ be a well-ordering of A and $<_L$ the lexicographic order with respect to $<_A$. The ordering $<_s$ on A^* is defined for all $u, v \in A^*$ as

$$u <_s v \quad \Leftrightarrow \quad |u| < |v| \text{ or } |u| = |v| \text{ and } u <_L v.$$

and it is called *shortlex ordering*.

(ii). An automatic structure $(W, \{M_x\}_{x \in A \cup \{\varepsilon\}})$ is called *shortlex automatic structure* if W accepts precisely the minimal words under the shortlex ordering of A^*. That means

$$L(W) = \{w \in A^* \mid w <_s v \text{ for all } v \in A^* \text{ with } w =_G v\}.$$

(iii). The group G is called *shortlex automatic with respect to* X if it has a shortlex automatic structure for the presentation $G = \langle X \mid R \rangle$.

Remark 4.5. Shortlex automaticity is a stronger property than automatcity: A group can be automatic for a given set A, but not shortlex automatic for the same A, independet of the chosen well-ordering on A. Also, if a group is shortlex automatic for a specific well-ordering on A, it does not have to be shortlex automatic for any other well-ordering on A.

We are now aiming to define an algorithm that computes a shortlex automatic structure for a given group G in case G is shortlex automatic to apply it to the Heineken group in the end. In order to be able to do that we need to introduce some more terminology and concepts:

Definition 4.12. Let $G = \langle X \mid R \rangle$ be a group and let $A := X \cup X^{-1}$.

(i). Let $w = a_1 \cdots a_n \in A^*$. Then we denote the prefix of w of length $i \in \mathbb{N}_0$ as w_i, i.e.

$$\begin{aligned} w_0 &= \varepsilon_A \\ w_i &= a_1 \cdots a_i && \text{with } a_i \in A \text{ for } i \in \{1, \ldots n-1\} \\ w_i &= w && \text{for } i \geq n\,. \end{aligned}$$

(ii). Let $(w, v) \in A^* \times A^*$. The set

$$\{w_i^{-1} v_i \mid i \in \mathbb{N}_0\} \subseteq G$$

is called *the set of word-differences associated to* (w, v).

(iii). Let $P \subseteq A^* \times A^*$. Then the set

$$\bigcup_{(w,v)\in P} \{w_i^{-1}v_i \mid i \in \mathbb{N}_0\} \subseteq G$$

is called *the set of word-differences associated to* P.

Lemma 4.5. *Let $(W, \{M_x\}_{x\in A\cup\{\varepsilon\}})$ be an automatic structure for the group G and let D be the set of word-differences associated to the set*

$$P = \{(w,v) \in A^* \times A^* \mid (w,v)^+ \in L(M_x) \text{ for some } x \in A \cup \{\varepsilon\}\}\,.$$

Then D is finite.

Proof. Let $w_i^{-1}v_i \in D$ for some $i \in \mathbb{N}_0$ where $(w,v) \in A^* \times A^*$ such that $(w,v)^+ \in L(M_x)$ for some $x \in A \cup \{\varepsilon\}$. Then there are $w', v' \in A^*$ such that $w = w_i w'$ and $v = v_i v'$. Hence $(w_i,v_i)^+(w',v')^+ \in L(M_x)$, which means that there is a path of arrows in M_x from $\sigma_{0,x}^{(w_i,v_i)^+}$ to a final state, where $\sigma_{0,x}$ denotes the start state of M_x. By removing all circuits, we can choose this path to be of length at most equal to the number $|\Sigma_x|$ of states in M_x. Define now $k := \max\{|\Sigma_x| \mid x \in A \cup \{\varepsilon\}\}$. Thus, we can choose $w'', v'' \in A^*$ such that $|w''|, |v''| \leq k$ and $(w_i,v_i)^+(w'',v'')^+ \in L(M_x)$. That implies by definition of $L(M_x)$ that $w_i w'' x = v_i v''$ and hence $w_i^{-1}v_i = w''x(v'')^{-1}$. But $w''x(v'')^{-1}$ has at most length $2k+1$, hence all elements in D can be at most of this finite length. But there can only exist finitely many distinct words of a finite length and per given pair (w,v), the set D can only contain finitely many elements as well, therefore D has to be finite. $\square$

Geometrically interpreted, Lemma 4.5 gives that any pair $(w,v) \in A^* \times A^*$ with $(w,v)^+ \in L(M_x)$ for some $x \in A \cup \{\varepsilon\}$ has the following property:

Definition 4.13. A pair $(w,v) \in A^* \times A^*$ has *Fellow-Traveller Property* if the distance $d(w_i,v_i)$ in the word metric of the Cayley graph $Cay(G,X)$ of G is uniformly bounded, i.e. there is a fixed constant $k > 0$ such that $d(w_i,v_i) \leq k$ for all $i \in \{0,1,\ldots,\max(|w|,|v|)\}$.

Definition 4.14. Let Z be an accessible 2-variable DFA with $S = \{\sigma\}$. It is called *word-difference automaton* if there is a map $\delta : \Sigma_Z \to G$ satisfying

(i). $\delta(\sigma_0) = 1_G$;

(ii). For all $a, b \in A^+$ and all $\sigma \in \Sigma_Z$ such that if the transition $\sigma^{(a,b)}$ is defined it holds that

$$\delta(\sigma^{(a,b)}) = a^{-1}\delta(\sigma)b\,,$$

with the convention that \$ represents 1_G.

Lemma 4.6. *Let Z be a word-difference DFA and let there be a $g \in G$ such that $\delta(\sigma) = g$ for all $\sigma \in F_Z$. Then $wg =_G v$ for all $w, v \in A^*$ with $(w,v)^+ \in L(Z)$.*

Proof. Let $w, v \in A^*$ with $(w,v)^+ \in L(Z)$. Then there is a $\sigma \in F_Z$ such that $\sigma_0^{(w,v)^+} = \sigma$, where $\{\sigma_0\} = S_Z$. Hence we obtain

$$\delta(\sigma) = \delta(\sigma_0^{(w,v)^+}) = w^{-1}\delta(\sigma_0)v = w^{-1}v\,.$$

But $\delta(\sigma) = g$, thus $g =_G w^{-1}v$, which implies $wg =_G v$. □

The definition of a word-difference automaton can be adjusted to fit in the concept of an automatic structure:

Definition 4.15. Let $(W, \{M_x\}_{x \in A \cup \{\varepsilon\}})$ be an automatic structure of the group G. Let for every $x \in A \cup \{\varepsilon\}$ the *associated word-difference DFA Z_x* be a word-difference automaton defined as follows:

- Define D_L and D_R to be identical copies of the set of word-differences D and let $\Sigma_{Z_x} := D \cup D_L \cup D_R$ be the set of states of Z_x.
- Let 1_G be the unique start state.
- Let $\delta : \Sigma_{Z_x} \to G$ be the identity map.
- Define for $\sigma \in \Sigma_{Z_x}$ and $a, b \in A$ the transition $\sigma^{(a,b)} = \rho$ if $a^{-1}\sigma b =_G \rho \in D$ and let it undefined otherwise. Furthermore we define a transition $(a, \$) \in A \times A^+$ only when the source state is in $D \cup D_L$ and the target is in D_L. These are the only transitions with target in D_L. A transition $(\$, a) \in A^+ \times A$ is only defined when the source state is in $D \cup D_R$ and the target is in D_R. These are the only transitions with target in D_R.
- We make Z_x accessible by removing all inaccessible states.
- There are (up to) three final states, elements representing x in D, D_L and D_R.

Proposition 4.7. *Let $x \in A \cup \{\varepsilon\}$ arbitrarily. Then:*

(i). $wx =_G v$ *for all* $(w,v)^+ \in L(Z_x)$.

(ii). $(w,v)^+ \in L(Z_x)$ *for all* $(w,v)^+ \in L(M_x)$.

Proof.

(i). By definition, we know that all accepting states of Z_x represent x, hence $\delta(\sigma) = x$ for all $\sigma \in F_{Z_x}$ and we can apply Lemma 4.6. This tells us that for every $w, v \in A^*$ with $(w,v)^+ \in L(Z_x)$, $wx =_G v$.

(ii). Let $(w,v)^+ \in L(M_x)$, then $w, v \in L(W)$ and $wx =_G v$, which implies $x = w^{-1}v$. Consider $\sigma_0^{(w,v)^+}$, where $\{\sigma_0\} = S_{Z_x}$. Since the set of states in Z_x is precisely the set D of all word-differences associated to elements in $L(M_x)$, it follows that $\sigma_0^{(w,v)^+}$ is defined and that it is precisely

$$\sigma_0^{(w,v)^+} = w^{-1}v =_G x\,.$$

But x is represented by the accepting states of Z_x, hence $(w,v)^+ \in L(Z_x)$.

□

Definition 4.16. A collection $\{Z_x\}_{x\in A\cup\{\varepsilon\}}$ of word-difference DFA for a group G is called *correct* for an automatic structure $(W, \{M_x\}_{x\in A\cup\{\varepsilon\}})$ of G if every Z_x satisfies Proposition 4.7.

Proposition 4.8. *Let $\{Z_x\}_{x\in A\cup\{\varepsilon\}}$ be a collection of word-difference DFA for a group G. If it is correct for an automatic structure $(W, \{M_x\}_{x\in A\cup\{\varepsilon\}})$ of G then*

$$L(M_x) = \{(w,v)^+ \mid w, v \in L(W) \text{ and } (w,v)^+ \in L(Z_x)\}$$

for every $x \in A \cup \{\varepsilon\}$.

Proof. This follows directly by the definition of correctness of Z_x and the definition of the language of M_x as

$$L(M_x) = \{(w,v)^+ \mid w, v \in L(W) \text{ and } wx =_G v\}\,.$$

□

4.1.3 Algorithm to Compute Shortlex Automatic Structures

Before we can describe the full algorithm to find a shortlex automatic structure we need to introduce another algorithm essential for the process:

4.1.3.1 Knuth-Bendix Algorithm

Definition 4.17. Let A be a finite set and let A^* be the monoid consisting of all words over A. A *rewriting system on* A^* is a set $\mathcal{S}$ of ordered pairs $(w_1, w_2) \in A^* \times A^*$. An element $(w_1, w_2) \in \mathcal{S}$ is called *rewrite rule*, w_1 is referred to as the *left hand side (LHS)* and w_2 as the *right hand side (RHS)*.

Remark 4.6. We can interpret a rewrite rule (w_1, w_2) as the rule to replace an occurrence of w_1 in a word by w_2. Hence we want to assume that no distinct rules have the same LHS, since otherwise we would not be able to choose which

replacement to use.
For $u, v \in A^*$, we write

$$u \to_{\mathcal{S}} v$$

if there exist words $x, y, w_1, w_2 \in A^*$ such that

$$u = xw_1y, \quad v = xw_2y \quad \text{and} \quad (w_1, w_2) \in \mathcal{S}\,,$$

i.e. v is obtained from u by making a substitution using the rewrite rule $(w_1, w_2) \in \mathcal{S}$ of the rewriting system $\mathcal{S}$.
We write

$$u \to_{\mathcal{S}}^* v$$

if there is an $n \in \mathbb{N}$ such that there are $u = u_0, u_1, \ldots, u_n = v \in A^*$ with $u_i \to_{\mathcal{S}} u_{i+1}$ for every $0 \leq i \leq n$.
We write

$$u \leftrightarrow_{\mathcal{S}}^* v$$

if both $u \to_{\mathcal{S}}^* v$ and $v \to_{\mathcal{S}}^* u$.
Without a danger of ambiguity, we might omit the index $\mathcal{S}$ to clarify which rewriting system we are talking about.

Definition 4.18.

(i). A word $u \in A^*$ is called *irreducible* or *reduced* if there is no $v \in A^*$ such that $u \to v$.

(ii). The set of *descendents* of a word $u \in A^*$ is defined as

$$desc(u) \coloneqq \{w \in A^* \setminus \{u\} \mid u \to^* w\}$$

Definition 4.19. A rewriting system $\mathcal{S}$ is called

(i). *Noetherian* if there is no infinite chain of words, which means that for every $u \in A^*$ there is an irreducible $v \in A^*$ such that $u \to_* v$.

(ii). *confluent* if for every $u, v_1, v_2 \in A^*$ with $u \to^* v_1$ and $u \to^* v_2$ there is a word $w \in A^*$ such that $v_1 \to^* w$ and $v_2 \to^* w$.

(iii). *complete* if it is Noetherian and confluent.

(iv). *locally confluent* if for every $u, v_1, v_2 \in A^*$ with $u \to v_1$ and $u \to v_2$ there is a word $w \in A^*$ such that $v_1 \to^* w$ and $v_2 \to^* w$.

Lemma 4.9. *Let $\mathcal{S}$ be a rewriting system. Then:*

(i). If $\mathcal{S}$ is Noetherian, $|desc(u)| < \infty$ for all $u \in A^*$.

(ii). If $\mathcal{S}$ is Noetherian and locally confluent, $\mathcal{S}$ is confluent, hence complete.

(iii). If $\mathcal{S}$ is Noetherian and locally confluent, each equivalence class under $\leftrightarrow^*$ contains a unique $\mathcal{S}$-irreducible element.

(iv). $\mathcal{S}$ is locally confluent if and only if the following two conditions hold for all rewriting rules $(u_1, t_1), (u_2, t_2) \in \mathcal{S}$:

(a) If $u_1 = rs$ and $u_2 = st$ for some $r, t \in A^*$ and $s \in A^* \setminus \{\varepsilon\}$, then there exists a word $w \in A^*$ such that $t_1 t \to^* w$ and $r t_2 \to^* w$.

(b) If $u_1 = rst$ and $u_2 = s$ for some $r, t \in A^*$ and $s \in A^* \setminus \{\varepsilon\}$, then there exists a word $w \in A^*$ such that $t_1 \to^* w$ and $r t_2 t \to^* w$.

Definition 4.20. A pair $(u_1, t_1), (u_2, t_2) \in \mathcal{S}$ of rewriting rules of the rewriting system $\mathcal{S}$ is called a *crititcal pair* if either the condition (iv)a

$$u_1 = rs \text{ and } u_2 = st \text{ for some } r, t \in A^* \text{ and } s \in A^* \setminus \{\varepsilon\}$$

or the condition (iv)b

$$u_1 = rst \text{ and } u_2 = s \text{ for some } r, t \in A^* \text{ and } s \in A^* \setminus \{\varepsilon\}$$

of Lemma 4.9 (iv) holds.

Proof. (of Lemma 4.9) Since a complete proof can be found in [19, p. 418f.], we only want to outline the crucial arguments:

(i). By assuming that there is an $u \in A^*$ with $|desc(u)| = \infty$, we can construct an infinite chain $\{u_i\}_{i \in \mathbb{N}}$ such that $u_i \to u_{i+1}$ for all $i \in \mathbb{N}$, in contradiction to $\mathcal{S}$ being Noetherian.

(ii). Show by induction on $|desc(u)|$ that for every $u \in A^*$ there is a unique irreducible $w \in A^*$ such that $u \to^* w$. This is possible since (i) ensures that $|desc(u)|$ is finite.

(iii). Let $u, v \in A^*$ such that $u \leftrightarrow^* v$ and both u and v irreducible. Then there exist an $n \in \mathbb{N}$ and $u = u_0, u_1, \ldots, u_n = v$ with

$$u_0 \;{}^*\!\leftarrow u_1 \to^* u_2 \;{}^*\!\leftarrow \cdots \to^* u_n \,.$$

Induction on $n \in \mathbb{N}$ gives the result.

(iv). Suppose first that $\mathcal{S}$ is locally confluent. Suppose there is a pair of rewriting rules $(u_1, t_1), (u_2, t_2) \in \mathcal{S}$ such that

(a) $u_1 = rs$ and $u_2 = st$ for some $r, t \in A^*$ and $s \in A^* \setminus \{\varepsilon\}$. Consider the word $v = rst$. By applying the two rules it can be either rewritten to

$$v \to t_1 t \quad \text{or} \quad v \to rt_2 \,.$$

But then local confluency implies that there is a word $w \in A^*$ such that $t_1 t \to^* w$ and $rt_2 \to^* w$, so (iv)a holds.

(b) $u_1 = rst$ and $u_2 = s$ for some $r, t \in A^*$ and $s \in A^* \setminus \{\varepsilon\}$. Consider the word $v = rst$. By applying the two rules it can be either rewritten to

$$v \to t_1 \quad \text{or} \quad v \to rt_2 t \,.$$

But then local confluency implies that there is a word $w \in A^*$ such that $t_1 \to^* w$ and $rt_2 t \to^* w$, so (iv)b holds.

Assume conversely that (iv)a and (iv)b hold. Let $u \in A^*$ be a word with $u \to v_1$ and $u \to v_2$. Then u has two subwords u_1 and u_2 such that there are rewrite rules $(u_1, t_1), (u_2, t_2) \in \mathcal{S}$. There are two cases:

- If u_1 and u_2 do not overlap, then there are $r, s, t \in A^*$ such that $u = ru_1 su_2 t$. Hence $v_1 = rt_1 su_2 t$ and $v_2 = ru_1 st_2 t$. But then $v_1 \to w$ and $v_2 \to w$ with $w = rt_1 st_2 t$.
- If u_1 and u_2 overlap, then we are either in case (iv)a or (iv)b. Hence there is a word $w \in A^*$ such that $v_1 \to^* w$ and $v_2 \to^* w$.

Either way, $\mathcal{S}$ is locally confluent.

$\square$

Definition 4.21. The following process is called the *Knuth-Bendix completion process*: Take a rewriting system $\mathcal{S}$ and consider all critical pairs. For every critical pair $(u, v) \in A^* \times A^*$ for which neither the implication of (iv)a nor the implication (iv)b is true, determine the two distinct irreducible words w_1 and w_2 which are equivalent under $\leftrightarrow^*$ and adjoin the pair (w_1, w_2) or the pair (w_2, w_1) to $\mathcal{S}$.

One questions remains: How should the process decide whether to adjoin (w_1, w_2) or (w_2, w_1) to the rewriting system $\mathcal{S}$? It should be guided by an ordering:

Definition 4.22. Let $\leq$ be a well-ordering on A^*. It is called *reduction ordering* if $u \leq v$ for $u, v \in A^*$ implies that

$$uw \leq vw \quad \text{and} \quad wu \leq wv$$

for all $w \in A^*$.

Example 4.4. The shortlex ordering is a reduction ordering.

Remark 4.7. If we order all elements of A^* with a reduction ordering $\leq$, then $u \to^* v$ for $u, v \in A^*$ implies that $u \geq v$ and hence $\mathcal{S}$ is always Noetherian, since $\leq$ has to be a well-ordering. Thus we can forthcoming assume any rewriting system to be Noetherian.

It is natural to impose that the Knuth-Bendix completion process should, when adjoining a critical pair consisting of w_1 and w_2 to the rewriting system $\mathcal{S}$, choose the larger one of w_1 and w_2 with respect to a given reduction ordering to be the LHS of the new rule.

Definition 4.23. A rewriting rule (w_1, w_2) of the rewriting system $\mathcal{S}$ is called *irreducible* if w_2 and all proper subwords of w_1 are irreducible. The rewriting system $\mathcal{S}$ is called *irreducible* or *reduced* if all rules it contains are reduced.

Remark 4.8. Any non-irreducible rule in a rewriting system can be simplified. Any new rule adjoined by the Knuth-Bendix completion process is itself irreducible and it might imply further simplification of other rules, that have not been possible before. Hence we adjoin a new step to the Knuth-Bendix completion process: For any pair of rules, check if at least one of them can be removed or simplified. If not, check if it is a critical pair.

Definition 4.24. Let $\mathcal{S}$ be a rewriting system on A^*.

(i). An element $u \in A^*$ is called *($\mathcal{S}$-)minimal* if it is the least element with respect to a reduction ordering in its $\leftrightarrow^*$ equivalence class.

(ii). A pair $(w_1, w_2) \in A^* \times A^*$ is called *($\mathcal{S}$-)essential* if $w_1 \leftrightarrow^* w_2$ and w_2 and all proper subwords of w_1 are minimal.

Proposition 4.10. *[19, p. 421, Propostion 12.20] Let $\mathcal{S}$ be a rewriting system on A^*.*

(i). Let $w \in A^*$ and let $v \in A^*$ be $\mathcal{S}$-minimal such that $w \leftrightarrow^* v$. After running the Knuth-Bendix completion process on $\mathcal{S}$ sufficiently long, we have $w \to^*_{\mathcal{S}} v$.

(ii). Let (w_1, w_2) be an $\mathcal{S}$-essential rule. After running the Knuth-Bendix completion process on $\mathcal{S}$ sufficiently long, we have $(w_1, w_2) \in \mathcal{S}$.

Corollary 4.11. *[19, p. 422, Corollary 12.21] If there are only finitely many $\leftrightarrow^*$ equivalence classes, the Knuth-Bendix completion process on a rewriting system $\mathcal{S}$ halts with a finite complete set of rules.*

Now we are ready to use the Knuth-Bendix completion procedure also for group theoretical problems. We need the following observations:
Let $\mathcal{S}$ be a rewriting system on A^*. Then the equivalence relation $\leftrightarrow^*$ is a congruence on A, generated by $\mathcal{S}$. Hence we have in the monoid presentation that

$$A/\leftrightarrow^* = Mon\langle A \mid \mathcal{S}\rangle\,.$$

By Lemma 4.9 (iii), if $\mathcal{S}$ is complete, any $\mathcal{S}$-irreducible word is $\mathcal{S}$-minimal and they form a set of representatives of the $\leftrightarrow^*$ equivalence classes.
If $\mathcal{S}$ is not only complete, but also finite, we can determine for any pair $u, v \in A^*$, whether u and v reduce to the same irreducible word. Hence the word problem is solvable.
If $G = \langle X \mid R\rangle$ is now a group, we already know that G has a corresponding monoid presentation $Mon\langle A \mid \mathcal{I}_X \cup \mathcal{R}\rangle$. Hence we can interpret the set $\mathcal{I}_X \cup \mathcal{R}$ as a rewriting system on which we apply the Knuth-Bendix completion process. This might solve the word problem for specific groups.

Example 4.5. Let F_r be the free group of rank r. It corresponds to the monoid $Mon\langle\{x_1, \ldots, x_r, x_1^{-1}, \ldots, x_r^{-1}\} \mid \mathcal{I}_X\rangle$ where $\mathcal{I}_X$ consists of all pairs $(x_i x_i^{-i}, \varepsilon)$ and $(x_i^{-1}x_i, \varepsilon)$ for $i \in \{1, \ldots, r\}$, where we order with shortlex ordering. The pairs in $\mathcal{I}_X$ are critical, but any overlap of x_i^{-1} in $x_i x_i^{-1} x$ reduces to x in both cases for all $i \in \{1, \ldots, r\}$. Hence the rewriting system $\mathcal{I}_X$ is confluent by Lemma 4.9 (iv). Furthermore, we see that the $\mathcal{I}_X$-reduced words in F_r are those containing no adjacent mutually-inverses.

Example 4.6. Let $\mathcal{D}_3 := \langle x, y \mid x^3 = 1, y^2 = 1, (xy)^2 = 1\rangle$ be the dihedral group of order 6. The alphabet is given by $A := \{x, x^{-1}, y, y^{-1}\}$ and we impose the ordering $x < x^{-1} < y < y^{-1}$. We can start with the rules

$$\mathcal{S} = \{(xx^{-1}, \varepsilon), (x^{-1}x, \varepsilon), (yy^{-1}, \varepsilon), (y^{-1}y, \varepsilon), (xx, x^{-1}), (y^{-1}, y), (yx^{-1}, xy)\}\,,$$

where the LHS is larger than the RHS in the shortlex ordering.
Consider the critical pair $(x^{-1}x, \varepsilon), (xx, x^{-1})$, where there is an overlap in x. Hence the word $w = x^{-1}xx$ can be reduced in two different ways to x and to $x^{-1}x^{-1}$ and we adjoin the rule $(x^{-1}x^{-1}, x)$ to $\mathcal{S}$. After examining all critical pairs we end up with the complete set of rules

$$\begin{aligned}\mathcal{S} = \{&(xx^{-1}, \varepsilon), (x^{-1}x, \varepsilon), (yy^{-1}, \varepsilon), (y^{-1}y, \varepsilon), (xx, x^{-1}),\\ &(x^{-1}x^{-1}, x), (y^{-1}, y), (yy, \varepsilon), (yx^{-1}, xy), (yx, x^{-1}y)\}\,.\end{aligned}$$

A last observation gives that the rules (yy^{-1}, ε) and $(y^{-1}y, \varepsilon)$ are not reduced, since there is a rule containing y^{-1} as a LHS, hence y^{-1} is reducible. Substituting the occurrence of y^{-1} in (yy^{-1}, ε) and $(y^{-1}y, \varepsilon)$ with y gives two redundant rules, thus we obtain a set of essential rules:

$$\begin{aligned}\mathcal{S} = \{&(xx^{-1}, \varepsilon), (x^{-1}x, \varepsilon), (xx, x^{-1}), (x^{-1}x^{-1}, x),\\ &(y^{-1}, y), (yy, \varepsilon), (yx^{-1}, xy), (xy, x^{-1}, y)\} .\end{aligned}$$

We want to forge the bridge back to automatic groups:

Remark 4.9. Let $G = \langle X \mid R \rangle$ be a group with a monoid presentation $Mon\langle A \mid \mathcal{I}_X \cup \mathcal{R} \rangle$. Let W be word acceptor of a shortlex automatic structure for G. Define $\mathcal{S}$ to be the rewriting system $\mathcal{I}_X \cup \mathcal{R}$.

(i). A word $u \in A^*$ that is the $\mathcal{S}$-minimal representative of the equivalence relation $\leftrightarrow^*$ corresponding to $\mathcal{S}$, is the $\leq$-minimal representative of a group element in G. Hence the language of the unique word acceptor W is the set of all $\mathcal{S}$-minimal sets in A^*.

(ii). If (w, v) is an $\mathcal{S}$-essential pair, i.e. $w \leftrightarrow^* v$ and v and all proper subwords of w are minimal, then $w =_G v$ and all proper subwords of w are in the language of a unique word acceptor W.

While the Knuth-Bendix completion process seems to be a big asset, in its implementation a lot of time and space is needed - mostly for reducing words to equivalent irreducible words. This process was rendered more efficiently by the use of a DFA:

Definition 4.25. Let $\mathcal{S}$ be a finite rewriting process over A^*. Define the associated DFA $M_{\mathcal{S}}$ as follows:

- Σ is the set of all prefixes of all LHS of rules in $\mathcal{S}$.
- A is the alphabet.
- $S = \{\varepsilon\}$.
- F is the set of all proper prefixes of all LHS of rules in $\mathcal{S}$.
- For $\sigma \in F$ and $a \in A$, the transition σ^a is defined to be the longest suffix of $\sigma a \in \Sigma$ that lies in Σ. For $\sigma \in \Sigma \setminus F$, the transition σ^a is undefined. Hence complete LHS of rules are dead states.

If we assume that $\mathcal{S}$ is a reduced, finite rewriting process, hence no LHS of a rule is a proper substring of any other LHS of a rule, then the non-final states of $M_{\mathcal{S}}$ are precisely the LHS of the rules in $\mathcal{S}$. So let $w = w_1w_2w_3$ be a word in A^* with $(w_2, v_2) \in \mathcal{S}$, i.e. w_2 is a LHS of a rule in $\mathcal{S}$ (and let w_1 not contain a LHS of a rule). If we read w under $M_{\mathcal{S}}$, then after reading w_1w_2 a dead state is reached.

This means precisely that the language $L(M_{\mathcal{S}})$ is the set of $\mathcal{S}$-irreducible words. Hence, we can use $M_{\mathcal{S}}$ to reduce words in A^* with $\mathcal{S}$ in the following way:
Let $w \in A^*$. Read w into $M_{\mathcal{S}}$. If we find that w is of the form $w = w_1 w_2 w_3$ where $\varepsilon^{w_1 w_2} = w_2$, then we know there exists a rule $(w_2, v_2) \in \mathcal{S}$. Now we can replace the occurrence of w_2 by v_2 and restart the process.

Executing this process for the rewriting system of the dihedral group of order 6 we gave in Example 4.6 should illustrate it sufficiently:

Example 4.7. Let $D_3 \coloneqq \langle x, y \mid x^3 = 1, y^2 = 1, (xy)^2 = 1 \rangle$ be the dihedral group of order 6 with the complete, finite rewriting system

$$\mathcal{S} = \{(xx^{-1}, \varepsilon), (x^{-1}x, \varepsilon), (yy^{-1}, \varepsilon), (y^{-1}y, \varepsilon), (xx, x^{-1}),\\ (x^{-1}x^{-1}, x), (y^{-1}, y), (yy, \varepsilon), (yx^{-1}, xy), (yx, x^{-1}y)\}\,.$$

Number the states of $M_{\mathcal{S}}$ in the following way:

$$1 = \varepsilon, \quad 2 = x, \quad 3 = x^{-1}, \quad 4 = y;$$

$$-1 = y^{-1}, \quad -2 = xx^{-1}, \quad 3 = x^{-1}x, \quad -4 = xx,\\ -5 = x^{-1}x^{-1}, \quad -6 = yy, \quad -7 = yx^{-1}, \quad -8 = yx\,.$$

The set of final states is $F = \{1, 2, 3, 4\}$, while the dead states are ones with a negative number. The defined transitions are given by the longest suffix lying in the set of states, denoted in following table:

	x	x^{-1}	y	y^{-1}
1	2	3	4	-1
2	-4	-2	4	-1
3	-3	-5	4	-1
4	-8	-7	-6	-1

We want to use $M_{\mathcal{S}}$ to reduce the word $w = yxy^{-1}x^{-1}$. Scan the start state under w to obtain

$$1^w = 1^{yxy^{-1}x^{-1}} = 4^{xy^{-1}x^{-1}} = -8^{y^{-1}x^{-1}}\,,$$

which is undefined. Hence there has to be a rule in $\mathcal{S}$ with yx as a LHS. We find $(yx, x^{-1}y) \in \mathcal{S}$ and use it to replace the subword of w to obtain the word $w_1 = x^{-1}yy^{-1}x^{-1}$ equivalent to w. Now we scan the start state under w_1 and obtain

$$1^{w_1} = 1^{x^{-1}yy^{-1}x^{-1}} = 3^{yy^{-1}x^{-1}} = 4^{y^{-1}x^{-1}} = -1^{x^{-1}}\,.$$

Since $x^{-1}y$ was obtained by a replacement and $\mathcal{S}$ is reduced, we know that y^{-1} has to be the LHS of a rule in $\mathcal{S}$, namely $(y^{-1}, y) \in \mathcal{S}$. Replacing y^{-1} by y gives the word $w_2 = x^{-1}yyx^{-1}$, which is equivalent to w_1 and w. Scanning the start state under w_2 gives

$$1^{w_2} = 1^{x^{-1}yyx^{-1}} = 4^{yyx^{-1}} = -6^{yx^{-1}}\,.$$

Hence we find the rule $(yy, \varepsilon) \in \mathcal{S}$ to obtain the word $w_3 = x^{-1}x^{-1}$ equivalent to w_2, w_1 and w. The last scan of the start state under w_3 gives

$$1^{w_3} = 1^{x^{-1}x^{-1}} = 3^{x^{-1}} = -5\,.$$

Thus $x^{-1}x^{-1}$ is the LHS of a rule, the rule $(x^{-1}x^{-1}, x) \in \mathcal{S}$, which gives $w_4 = x$, equivalent to w_3, w_2, w_1 and w. Now

$$1^{w_4} = 1^x = 2$$

is accepted, hence $w = yxy^{-1}x^{-1}$ is equivalent to the reduced word $w_4 = x$.

Now we have all the necessary concepts at hand to describe the algorithm for finding a shortlex automatic structure:
As usual, we are dealing with a finitely presented group $G = \langle X \mid R \rangle$, where the set A is defined as $A := X \cup X^{-1}$. We need to have a total ordering on A given, which determines a the shortlex ordering $<$ on A^*. The aim is, in case G is shortlex automatic, that with sufficient time and space, we find a procedure that will eventually compute the shortlex automatic structure for G. If G is against our initial assumption not shortlex automatic, the procedure will simply not complete, hence no information is gained, since we cannot ensure that the procedure did not terminate for space or time deficiency reasons.

Definition 4.26. Let $(W, \{M_x\}_{x \in A \cup \{\varepsilon\}})$ be a shortlex automatic structure for the group G with the collection of associated word-difference DFA $\{Z_x\}_{x \in A \cup \{\varepsilon\}}$. In the algorithm computing this shortlex automatic structure with word-difference DFA we denote possible candidates as W', M'_x and Z'_x. We say a candidate is *correct* if its language agrees with the original one.

The algorithm is divided in five individual steps:

1. Compute a collection of candidates $\{Z'_x\}_{x \in A \cup \{\varepsilon\}}$ with the Knuth-Bendix completion process.

2. Use Z'_ε to compute a candidate W'.

3. Use Z'_ε and W' to compute candidates M'_x. If it turns out in this step that $L(W') \neq L(W)$, then we need to redefine Z'_x and repeat step 2.

4. Start testing the correctness of M_x'. If there is an $x \in A \cup \{\varepsilon\}$ with $L(M_x') \neq L(M_x)$, redefine Z_x' and repeat step 3.

5. Test the correctness of M_x' and W' fully. In case the test returns `false`, start the process from the beginning.

We will describe all steps in further detail now:

4.1.3.2 Step 1: Knuth-Bendix

Define $\mathcal{S}$ to be the rewriting system given by the monoid presentation associated to the group $G = \langle X \mid R \rangle$ using the shortlex ordering $<$ on A^*.

Lemma 4.12. *The set D_0 of all word-differences associated to the set $\mathcal{E} \subseteq \mathcal{S}$ of $\mathcal{S}$-essential rules is a subset of the set D of all word-differences associated to elements accepted by M_x for $x \in A \cup \{\varepsilon\}$. Thus, D_0 is finite.*

Proof. Let $(w, v) \in \mathcal{E}$, then we know by Remark 4.9 (ii) that $w =_G v$ and that all proper subwords of w and v are elements of $L(W)$. Hence there is an $u \in L(W)$ and an $x \in A$ such that $w = ux$. Thus the fact that $u, v \in L(W)$ and $ux =_G v$ implies that $(u, v)^+ \in L(M_x)$. But since by the shortlex ordering $|w| \geq |v|$, all word-differences associated with the pair $(w, v) = (ux, v)$ have to be associated with (u, v) as well. Hence D_0 has to be a subset of the set D of all word-differences associated to elements accepted by some M_x. But since D is finite by Lemma 4.5, the set $D_0 \subseteq D$ has to be finite as well. $\square$

So there can be infinitely many $\mathcal{S}$-essential rules, but the set of word-differences associated to them will always be finite.

Now we can describe the algorithm:

- Run the Knuth-Bendix completion process on $\mathcal{S}$: We denote at any stage of the procedure for a word $w \in A^*$ its current $\mathcal{S}$-reduced word obtained by the DFA-method described in Defintion 4.25 as $r(w)$.

- We regularly interrupt the Knuth-Bendix completion process to compute for each rule $(w, v) \in \mathcal{S}$ the set of reduced associated word-differences

$$\{r(a_i^{-1} r(w_{i-1}^{-1} v_{i-1}) b_i) \mid 0 \leq i \leq |w|\},$$

  where $w = a_1 \cdots a_{|w|}$ and $v = b_1 \cdots b_{|v|}$. Define D_0' to be the union of all those sets of word-differences over all rules currently in $\mathcal{S}$.

- Check if D_0' seems to have become stable. If so, compute the DFA Z_x' for all $x \in A \cup \{\varepsilon\}$ as follows:

- $\Sigma_{Z'_x} = D'_0 \cup D'_{0L} \cup D'_{0L}$;
- A is the alphabet;
- $S_{Z'_x} = \{\varepsilon\}$;
- $F_{Z'_x} = \{r(x)\}$
- For $\sigma \in D'_0$ and $a, b \in A^+$ in case $r(a^{-1}\sigma b) \in D'_0$ we define

  $$\sigma^{(a,b)} = r(a^{-1}\sigma b)$$

  if not, we leave it undefined. For padded pairs we use the additional transitions according to Definition 4.15.
- The map $\delta : \Sigma_{Z'_x} \to G$ sends a word in $\Sigma_{Z'_x}$ to its corresponding group element.

Proposition 4.13. *[19, p. 458, Proposition 13.14] If the Knuth-Bendix completion process runs sufficiently long, $(w, v)^+ \in L(Z'_\varepsilon)$ for any $\mathcal{S}$-essential rule $(w, v) \in \mathcal{E}$.*

4.1.3.3 Step 2: Construct the word acceptor

By Section 13.1.6 in [19], we can construct a DFA W' with a language defined as follows:
Let L be the set

$$L := \{w \in A^* \mid \exists\, v \in A^* : (w, v)^+ \in L(Z'_\varepsilon) \text{ and } v < w\}\,.$$

Then the language $L(W')$ of W' is given by $L(W') = \neg(A^* L A^*)$.

Definition 4.27. A set $P \subseteq A^*$ is called *prefix closed* if for all $w \in P$ all its proper prefixes w_i for $i \in \{0, \dots, |w| - 1\}$ are elements of P as well.

Proposition 4.14. *[19, p. 460, Proposition 13,15] The defined language $L(W')$ is prefixed closed, i.e. for every $w \in L(W')$, all prefixes $w_i \in L(W')$ for $i \in \{1, \dots, |w|\}$. Furthermore $L(W) \subseteq L(W')$ and if the DFA Z'_ε accepts all essential rules, then $L(W') \subseteq L(W)$ as well.*

Remark 4.10. It may happen that we need to replace a word $u \in A^*$ by a word $u' \in L(W')$ with $u =_G u'$. In general this is possible by executing the following procedure: Let $u \notin L(W')$, then we need to find a subword w of u which is the shortest subword that is not in $L(W')$. Then there is a $v \in A'$ such that $(w, v)^+ \in L(Z')$ and $v < w$. We can find this word v with an algorithm described in Section 13.1.7 of [19] and then substitute the occurrence of w in u by v. Then we repeat the process. Since u gets shorter and shorter in the shortlex ordering in each step, the process will terminate in finitely many steps with $u \in L(W')$. In fact, since $(w,v)^+ \in L(Z'_\varepsilon)$ implies that $(uw,uv)^+ \in L(Z'_\varepsilon)$ for all $u \in A^*$, it suffices to find the shortest prefix w of u with $w \notin L(W')$. That is an easy task: Since $L(W')$ is prefix-closed by Proposition 4.14, we simply read u into W' and we have found the shortest prefix not in $L(W')$ when we reach a failure state.

4.1.3.4 Step 3: Construct the multiplier automata

By Proposition 4.8 we know that

$$L(M_x) = \{(w,v)^+) \mid w,v \in L(W) \text{ and } (w,v)^+ \in L(Z_x)\}\,,$$

if Z_x is correct. Hence, since we want Z'_x to be correct, we need to define M'_x such that this set equivalence holds. This is ensured by constructing M'_x for $x \in A \cup \{\varepsilon\}$ as follows:

- $\Sigma_{M'_x} = \Sigma_{W'} \times \Sigma_{W'} \times \Sigma_{Z'_x}$;
- $S_{M'_x} = S_{W'} \times S_{W'} \times S_{Z'_x} := (\sigma_0, \sigma_0, \rho_0)\,.$
- $F_{M'_x} = F_{W'} \times F_{W'} \times F_{Z'_x}$;
- Since transitions for W' are not yet defined, we need to correct this in order to extend the definition of transitions on W' and Z'_x to get the transitions for M'_x. Hence we define $\sigma^\$ = \sigma$ if $\sigma \in F_{W'}$ and we leave it undefined otherwise. Then we define for a pair $(a,b) \in (A^+ \times A^+)^*$ and $(\sigma_1, \sigma_2, \rho) \in \Sigma_{M'_x}$ the transition

$$(\sigma_1, \sigma_2, \rho)^{(a,b)} = (\sigma_1^a, \sigma_2^b, \rho^{(a,b)})\,.$$

It is an easy check to see:

$$L(M'_x) = \{(w,v)^+ \mid w,v \in L(W'), (w,v)^+ \in L(Z'_x)\}\,.$$

Let us state two properties, we will need in Step 5:

Proposition 4.15. *[19, p. 461, Proposition 13.16]*

(i). For all $w,v \in A^*$ with $(w,v)^* \in L(M'_x)$ we have $w,v \in L(W')$ and $wx =_G v$.

(ii). If $a_1 a_2 \cdots a_n \in L(W')$ with $n > 0$, then

$$(a_1 \cdots a_{n-1}, a_1 a_2 \cdots a_n)^+ \in L(M'_{a_n})$$

.

Remark 4.11. Applying Step 3 directly after Step 2 guarantees that we cannot have $(w, v)^+ \in L(M'_\varepsilon)$ with $w > v$. But in Step 4 we will check the correctness of $L(M'_x)$ and if it turns out that it is not correct, we need to redefine that underlying Z'_x. Then we go back to Step 3, and we may find $(w, v)^+ \in L(M'_\varepsilon)$ with $w > v$. This implies that $w \notin L(W)$, hence $L(W')$ was not correct. Usually the reason for that is that the word difference associated to (w, v) is not yet in D'_0 used to construct Z'_x. Thus we need to compute the missing word differences, reduce them to words in $L(W')$ as described in Remark 4.10 and adjoin them to D'_0. Then we can recompute Z'_x and repeat Step 2.

4.1.3.5 Step 4: Checking the correctness of the multiplier automata

Throughout Step 4, we will assume that W' is correct.

Lemma 4.16. *If $L(M'_x)$ is not correct for some $x \in A \cup \{\varepsilon\}$, there is a $w \in L(W')$ such that there is no $v \in A^*$ with $(w, v)^+ \in L(M'_x)$.*

Proof. If $L(M'_x)$ is not correct for some $x \in A \cup \{\varepsilon\}$, then there are $w, v \in L(W')$ with $wx =_G v$ but $(w, v)^+ \notin L(M'_x)$. Suppose for this $w \in L(W')$ there is another $v' \in A^*$ with $(w, v') \in L(W') = L(W)$ with $(w, v') \in L(M'_x)$. Then $w, v' \in L(W') = L(W)$ and $wx =_G v'$, which implies that $v = v'$. Hence if $L(M'_x)$ is not correct, for $w \in L(W')$ such a $v \in A^*$ with $(w, v)^+ \in L(M'_x)$ cannot exist. □

Again, by the construction given in Section 13.1.6 of [19], we can define an FSA E_x for every $x \in A \cup \{\varepsilon\}$ with language

$$L(E_x) = \{w \in A^* \mid \exists\, v \in A^* : (w, v)^+ \in L(M'_x)\}.$$

There are two cases:

1. If $L(E_x) = A^*$ for all $x \in A \cup \{\varepsilon\}$, then $L(M'_x)$ cannot be incorrect by Lemma 4.16. Hence we can move on to Step 5.
2. If there is an $x \in A \cup \{\varepsilon\}$ with $L(E_x) \neq A^*$, then we can use an enumeration algorithm given in Section 13.1.2 of [19] to find $\neg L(E_x)$. With the procedure described in Remark 4.10 we can find for every $w \in \neg L(E_x)$ and element $v \in L(W')$ such that $wx =_G v$. Next we compute the word differences associated to (w, v), reduce them to words in $L(W')$ and adjoin them to D'_0. Then we can recompute the Z'_x and return to Step 3.

The following Theorem ensures that the algorithm does exactly what we are aiming for:

Theorem 4.17. *[19, p. 463, Theorem 13.17] If the Knuth-Bendix algorithm in Step 1 runs long enough and the group G is shortlex automatic, then the shortlex automatic structure will be successfully computed in Step 2 to Step 4.*

4.1.3.6 Step 5: Checking the correctness of the shortlex automatic structure

The following Theorem ensures that the shortlex automatic structure $(W', \{M'_x\}_{x\in A\cup\{\varepsilon\}})$ found in Step 1 to Step 4 is indeed a shortlex automatic structure for G.

Theorem 4.18. *[19, p.464, Theorem 13.18] Let $G = \langle X \mid R\rangle$ be a group with monoid presentation $Mon\langle X \mid \mathcal{I}_X \cup R\rangle$. Suppose W' is an FSA and M_x is a 2-variable FSA over A for all $x \in A \cup \{\varepsilon\}$ satisfying:*

(i). If $(w, v)^+ \in L(M_x)$ for some $w, v \in A^*$ and $x \in A\cup\{\varepsilon\}$, then $w, v \in L(W)$ and $wx =_G v$.

(ii). $L(W)$ is not empty.

(iii). If $a_1 \cdots a_n \in L(W)$ with $n > 0$, then $a_1 \cdots a_{n-1} \in L(W)$ and $(a_1 \cdots a_{n-1}, a_1 \cdots a_n)^+ \in L(M_{a_n})$.

(iv). Let $w = w_0 \in L(W)$ and $(a_1 \cdots a_n, \varepsilon) \in \mathcal{I}_X \cup R$. Then for $w_n \in L(W)$ there are $w_1, \ldots, w_{n_1} \in L(W)$ such that $(w_{i-1}, w_i)^+ \in L(M_{a_i})$ for $i \in \{1, \ldots, n\}$ if and only if $w = w_n$.

Then $(W, \{M_x\}_{x\in A\cup\{\varepsilon\}})$ forms an automatic structure for G with uniqueness.

So we only need to check that our structure $(W', \{M'_x\}_{x\in A\cup\{\varepsilon\}})$ satisfies the conditions (a precise verification is given in [19, p. 465]):

(i). This holds by Proposition 4.15 (ii).

(ii). Since by Proposition 4.14, $L(W')$ is prefix-closed, it contains ε and is therefore non-empty.

(iii). This holds by Proposition 4.15 (i).

(iv). To check this property, we need to use the construction of a so-called composite FSA M_u for $u = a_1 \cdots a_n \in A^*$, described in Section 13.1.7 of [19]. Its language is given by

$$L(M_u) = \{(w, v)^+ \mid \exists\, w = w_0, w_1, \ldots, w_n = v \in A^* : \\ (w_{i-1}, w_i)^* \in L(M'_{a_n}), i \in \{1, \ldots, n\}\}.$$

So if the property holds, we know

$$L(M_u) = \{(w, w) \mid w \in L(W')\} .$$

So in Step 5, we construct the language of M_u and check whether it is indeed equal to the set $\{(w, w) \mid w \in L(W')\}$.

If this above algorithm returns `true`, then we know by Theorem 4.18 that $(W', \{M'_x\}_{x \in A \cup \{\varepsilon\}})$ is a shortlex automatic structure for G. If it returns `false`, we need to start from the very beginning with Step 1.

Example 4.8. We want to use the shortlex automatic algorithm to determine whether the Heineken group is an automatic group. Therefore we use the following input in GAP:

```
gap > F := FreeGroup("a", "b", "c");;
gap > a := F.1;; b := F.2;; c := F.3;;
gap > H := F/[Comm(a, Comm(a, b)) * c^-1, Comm(b, Comm(b, c)) * a^-1,
>        Comm(c, Comm(c, a)) * b^-1];;
gap > R := KBMAGRewritingSystem(H);;
gap > AutomaticStructure(R, true);
```

The first line generates the free group F over the three generators a, b and c, the second line defines the names of the generators as a, b and c. The third line takes the quotient of F by the defining relations of the Heineken group, hence H is the Heineken group. In the forth line we compute a rewriting system R for H and in the last line we compute whether with this R we can find an automatic structure by using the shortlex automatic algorithm of H executed via the function `AutomaticStructure`. The input `true` states that we want to have large parameters. Since we suppressed the output of all but the last line with ;; , we only get the output for the last line, namely (depending on the info level we set) `true`. Hence we know that H is automatic. Furthermore we can read off that the procedure halts with 63832 equations, that there are 271 word differences and that the word-acceptor has 1106 states.

4.1.4 Use & Value of Automatic Structures

4.1.4.1 Finiteness

Let G be a group for which we have found a shortlex automatic structure $(W, \{M_x\}_{x \in A \cup \{\varepsilon\}})$. Since W is a word acceptor with uniqueness, we know as already mentioned in Remark 4.4 that we can determine whether G is finite or infinite by the FSA-COUNT algorithm.

Example 4.9. With the previously in Example 4.8 computed automatic structure for H, we can find the order of H. In GAP this is easily done, since the FSA-COUNT of the word-acceptor is implemented as the function `Size(R)`, where R is the rewriting system of H we used to compute the automatic structure with. So inputting

$$\texttt{gap} > \texttt{Size}(\texttt{R});$$

gives `infinity`, hence H is not finite, in contradiction to Heineken's conjecture.

4.1.4.2 Hyperbolicity

Definition 4.28. A group G is called *strongly geodesically automatic* if the set of all geodesic words form the language of the word acceptor of an automatic structure for G.

Theorem 4.19. *A strongly geodesically automatic group is hyperbolic.*

Proof. [39, p.1, Theorem 2] Suppose $G = \langle X \mid R \rangle$ is a strongly geodesically automatic group. Let $A = X \cup X^{-1}$. Then the set of all geodesic words form the language of the word acceptor W of a automatic structure $(W, \{M_x\}_{x \in A \cup \{\varepsilon\}})$ for G. Let now γ, γ' be a bigon in the Cayley graph $Cay(G, X)$ with l the length of γ and let $u_1, u_2 \in A^*$ be geodesic words in A^* such that $\gamma(l)\gamma'^{-1}(l) = u_1 x u_2^{-1}$ for $x \in A \cup \{\varepsilon\}$ and $u_1, u_2 \in V(Cay(G, X))$, i.e. the length of $u_1^{-1} u_2$ is at most 1. If $x = \varepsilon$, which means that the endpoint $\gamma(l)$ of γ is a vertex in $V(Cay(G, X))$, then $u_1 =_G u_2$ and thus $(u_1, u_2)^+ \in M_\varepsilon$. If $x \in A$, which means that the endpoint $\gamma(l)$ of γ is the midpoint of an egde in $E(Cay(G, X))$, then $u_1 x =_G u_2$ and thus $(u_1, u_2)^+ \in M_x$. Hence by the Fellow-Traveller Property in Lemma 4.5, there is a constant k such that $d(u_{1i}, u_{2i}) \leq k$ for for $i \in \mathbb{N}$ since all $(u_1, u_2)^+$ are in M_x for some $x \in A \cup \{\varepsilon\}$. Hence G is hyperbolic by Theorem 1.14. $\square$

The proof of this Theorem 4.19 suggests that we can verify hyperbolicity of a group G by finding the constant $k \geq 0$, for which all bigons are k-thin. This is exactly what the following algorithm that was introduced by Holt and Epstein in [11] does:
Assume that we have computed a shortlex automatic structure $(W, \{M_x\}_{x \in A \cup \{\varepsilon\}})$ We want to construct the following four sequences of finite state automata for $n > 0$:

- WD_n, which are the word-difference automata;
- GE_n, which are the automata, checking for geodesic equivalence;

- GW_n, which are the geodesic word acceptors;
- T_n, which are the automata testing whether the procedure has finished.

As above in Lemma 4.5, let D be the set of all word differences associated to the elements of the multiplier automata M_x with $x \in A \cup \{\varepsilon\}$. Moreover we introduce the set $\mathcal{WD}_n \subseteq G$, which is a subset of G containing 1_G and will be defined inductively for $n > 1$. For $n = 1$ we define $\mathcal{WD}_1 = D$.
Let WD_n be the word difference automaton associated to the set $\mathcal{WD}_n$ with unique accepting state 1_G. The other automata are constructed in a way that their languages are given by the following sets:

- $L(GE_n) = \{(u, v) \in A^* \times A^* \mid (u, v) \in L(WD_n), v \in L(W), l(u) = l(v)\}$;
- $L(GW_n) = \{u \in A^* \mid \exists\, v \in A^* : (u, v) \in L(GE_n)\}$;
- $L(T_n) = \{w \in A^* \setminus L(GW_n) \mid \exists u : (w, u) \in L(WD_n), u \in L(GW_n), l(u) = l(w)\}$.

We observe: For a pair $(u, v) \in L(GE_n)$, we have that $u =_G v$ and that u, v are both geodesics in $Cay(G, X)$ and v geodesic in G. So while they both have the same length, represent the same element and their associated word differences are in $\mathcal{WD}_n$, only v is the shortlex minimal representative of the corresponding element in G.
An element $u \in L(GW_n)$ is geodesic in $Cay(G, X)$, since it has a shortlex-minimal partner $v \in A^*$ that is geodesic in G, of the same length and representing the same element in G. Though, again u is not the shortlex-minimal representative of the corresponding element in G.
An element w of the test language is not in $L(GW_n)$, which means that it does not have a shortlex-minimal partner $v \in L(W)$ of same length, representing the same element and with the associated word. But it is geodesic in $Cay(G, X)$, since we want it to have a non-shortlex-minimal partner $u \in A^*$ that has a shortlex-minimal partner $v \in A^*$. Hence all three elements w, u and v illustrated in Figure 10 have the same length, correspond to the same element in G, and the word differences associated to (w, u) and (u, v) are in $\mathcal{WD}_n$, but only v is the shortlex-minimal representative of the corresponding word in G. So intuitively speaking, we can say that $u \in L(GW_n)$ is n steps away from being a shortlex-minimal element, while $w \in L(T_n)$ is $n + 1$ steps away from being a shortlex-minimal element.
If $L(T_n)$ is empty for some $n > 0$, the procedure halts. If for a given $n > 0$, we find $L(T_n)$ not to be empty, that means that there is a word $w \in L(T_n)$. We reduce this word with the procedure described in Remark 4.10 to its shortlex-representative $v \in L(W)$ and then we define the set $\mathcal{WD}_{n+1}$ as the union of $\mathcal{WD}_n$ and the set of word-differences associated to the pair (w, v).

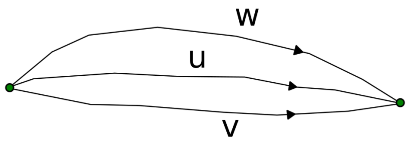


Figure 10: *While the geodesic paths v, u and w all represent the same element, only v is shortlex-minimal.*

Theorem 4.20. *[11, p. 7, Theorem 3.1] The above described procedure halts if and only if $G = \langle X \mid R \rangle$ is strongly geodesically automatic with respect to X.*

Remark 4.12. As by Theorem 4.19, any strongly geodesically automatic group is hyperbolic, by Theorem 4.20, the group G is hyperbolic if the procedure halts. This can be seen directly from the procedure as well: If it halts at step n then every geodesic element is paired with a shortlex-minimal element. Therefore geodesic bigons are $(2n+1) * l$-thin, where l is the maximal occurring length of the word differences, as illustrated in Figure 11 below.

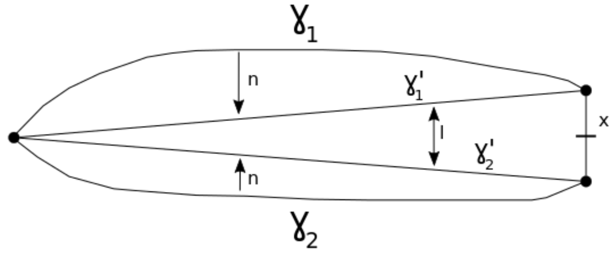


Figure 11: *The geodesic bigons γ_1 and γ_2 can be both reduced to their shortlex-minimal partners γ_1' and γ_2' in n steps of size l and their shortlex-representatives differ by one step of size l, thus their distance is at most $(2n+1)l$.*

Example 4.10. Unfortunately for the "lazy" user, there is no function in GAP like `IsHyperbolic(H)` to determine whether the Heineken group H is hyperbolic. So we have to start with the rewriting system R and the automatic structure we have computed for the Heineken group H in Example 4.8. We need to input the following steps to get the result we are aiming for:

```
gap > WriteRWS(R, "Heineken", "; "); ;
gap > progname := Filename(_KBExtDir, "gpgeowa"); ;
gap > callstring := Concatenation(progname, "Heineken"); ;
gap > Exec(callstring);
```

The function `WriteRWS(R,"Heineken",";")` writes the rewriting system R to a file with name `Heineken` in external format. The symbol `;` is the endsymbol. `_KBExtDir` is an external variable representing the directory of external programs in the KBMAG package available and `"gpgeowa"` is a program, computing whether bigons are ε-thin with the above described algorithm. Hence we use the function `Filename()` to name the program `"gpgeowa"` of the KBMAG directory, we choose the name `progname`. Then we need to concatenate the calling of the two external files, `progname` and `Heineken`. In the end we execute the program with the function `Exec()`, hence after that the program `gpgeowa` is run on the file `Heineken`. If the process completes, then we know bigons to be ε-thin for some $\varepsilon > 0$, i.e. the group to be hyperbolic. In the case of the Heineken group H it completes with the information

```
#Geodesic word-acceptor with 54 states computed.
#Geodesic pairs machine with 114 states computed.
#Geodesic difference machine with 31 states computed.
```

Thus, we have a verification that the Heineken group H is hyperbolic indeed.

4.2 Residual Finiteness

4.2.1 Using the Tools

While the tools described in Chapter 2 may be very valuable for other (well-known) examples as illustrated in Chapter 3, for the Heineken group H they have not been of great help so far. Since to apply the Theorem 2.3 of Malcev, we would need to check the Heineken group for linearity, the problem is simply shifted to another very tough open question of how to check whether a group is linear. Also, it is not apparent at all how to find a space or even a tree on which the Heineken group acts to be able to approach it with Bass-Serre theory. At least in Remark 4.1 with Tietze transformations we found another presentation of H that has not yet given us any advantage when applying the KBMAG-algorithms, but will be valuable in the forthcoming chapter dedicated to the task of finding finite quotients.

4.2.2 Finite Quotients

Since we do not have any tool at hand to deal with the question of residual finiteness of the Heineken group H, we choose to take one step back and start by trying to find finite quotients of the Heineken group since this has been the crucial idea for proving that the Higman group is not residually finite for example. Unfortunately, as opposed to the Higman group, the Heineken group has finite quotients, but it would be helpful if we could verify that at least it does not have a lot of them.

4.2.2.1 Brute Force

Of course, there is a naive approach to find a finite quotient: We choose any finite group F and search for a non-trivial homomorphism from the Heineken group H onto F. To check for a given group F whether there exists such an epimorphism, we can use GAP: There is an algorithm, called GQuotients, which finds all epimorphisms between a finitely presented group $G = \langle X \mid R \rangle$ and a finite group F up to automorphisms of F. The idea of the procedure is the following: We consider all r-tuples $(g_1, \ldots, g_r) \in F^r$, where $r := |X|$ is the number of generators of G. Then we test whether the map

$$\begin{aligned} \varphi : X &\to F \\ x_i &\mapsto g_i \qquad\qquad \text{for all } i \in \{1, \ldots, r\} \end{aligned}$$

extends to a homomorphism $\varphi : G \to F$ by checking if $\varphi(a_1) \cdots \varphi(a_m) = 1_G$ for all $w = a_1 \cdots a_m \in R$ with $a_i \in A := X \cup X^{-1}$ for all $i \in \{1, \ldots, m\}$. That this is a sufficient property for φ to be a homomorphism is ensured by Theorem 2.52 of [19, p. 37]. Of course this process is only efficient enough for small groups F and a small number of generators r. But we can prune the search tree by the following adjustment: We first try to rule out k-tuples $(g_1, \ldots, g_k) \in G^k$ for some $k < r$ as initial subsequences of a possible image $(g_1, \ldots, g_r) \in G^r$. Suppose there is a relator $w \in R$ only involving the first k generators $x_1, \ldots, x_k \in X$. Then we check if the map

$$\begin{aligned} \varphi : X &\to F \\ x_i &\mapsto g_i \qquad\qquad \text{for all } i \in \{1, \ldots, k\} \end{aligned}$$

extends to a homomorphism. If not we can rule out the initial subsequence $(g_1, \ldots, g_k) \in G^k$, if it is succesfull we consider a possible image $(g_1, \ldots, g_r) \in G^r$. Two further observations are also helpful: If we reorder the generators such that the subsets $R_k \subseteq R$ of R containing all relators involving only the first k elements are as large as possible, the algorithm becomes more efficient. Also, if there is a generator $x_i \in X$ of order $n \in \mathbb{N}$ for some $i \in \{1, \ldots, r\}$, then we know that it can only be mapped to an element with an order dividing n, which helps us to rule out some images.

In case of the Heineken group we first want to apply the algorithm to the finite, alternating group A_5 and to the finite, linear group $SL(2,5)$ of 2×2-matrices over the field of size 5, since [21, p. 345] suggests that those two groups are isomorphic to finite quotients of H. We obtain the two epimorphisms:

$$\begin{aligned} \varphi_1 : \mathsf{H} &\to A_5 \\ x &\mapsto (1,2,5,3,4) \\ y &\mapsto (1,2,4,5,3) \\ z &\mapsto (1,2,3,4,5) \end{aligned} \qquad \begin{aligned} \varphi_2 : \mathsf{H} &\to SL(2,5) \\ x &\mapsto \begin{bmatrix} 4 & 1 \\ 0 & 4 \end{bmatrix} \\ y &\mapsto \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix} \\ z &\mapsto \begin{bmatrix} 1 & 4 \\ 4 & 3 \end{bmatrix} . \end{aligned}$$

Furthermore, we can observe the following restrictions we can make before trying any other finite group F:
Denote the images of the generators of x, y and z by $\bar{x}, \bar{y}$ and $\bar{z}$.

- F cannot be abelian: Suppose it was. Since the relations of H have to hold in F as well, we have for example

$$\bar{x}^{-1}\bar{y}^{-1}\bar{x}^{-1}\bar{y}\bar{x}\bar{y}^{-1}\bar{x}\bar{y} = \bar{z} \, .$$

  But since the number of appearances of $\bar{x}$ and its inverse and of $\bar{y}$ and is inverse is the same, we conclude $\bar{z} = 1_F$. Analogously we get with the other two relations $\bar{y} = 1_F$ and $\bar{x} = 1_F$. Hence the epimorphism is trivial.

- No two generators of H can be mapped to generators of F that commute: Suppose $\bar{x}$ and $\bar{y}$ commute. Then we conclude as above that $\bar{z} = 1_F$. But then the third relation gives

$$\bar{z}^{-1}\bar{x}^{-1}\bar{x}^{-1}\bar{x}\bar{z}\bar{x}^{-1}\bar{z}\bar{x} = \bar{x}^{-1}\bar{x}\bar{x}^{-1}\bar{x} = 1_F$$

  and hence

$$\bar{y} = 1_F \, .$$

  Analogously we obtain with the second relation $\bar{x} = 1_F$. So the epimorphism is trivial again.

- No generator can be mapped to the trivial element 1_F: This follows by the same argument as above.

Unfortunately, these observation still have not been helpful enough to find a proper finite quotient by coincidence.
Also taking finite groups and adding the relations that have to hold in the

Heineken group has not been a successful brute force attempt.
Moreover, we can try to use the GAP-function `PQuotient(H,p)`. It searches for quotients of H, whose orders are a power less or equal 24 of the prime number p. Trying this with all prime numbers less than 500000 does not give any finite quotient though.

4.2.2.2 Coset Enumeration & the Low Index Method

Another way to determine whether a quotient is finite, is by counting the cosets it contains. Therefore we need to introduce the procedure to enumerate all the cosets efficiently that was formerly known as *Todd-Coxeter algorithm* and has been called *Coset Enumeration* since its implementation on a computer. We will not give the precise algorithm for the procedure, nor the pseudo code, but the main interest of this Chapter is to heuristically explain and illustrate with examples how the enumeration works.
First we introduce a tool to keep track of the current cosets in the procedure:

Definition 4.29. Let $G = \langle X \mid R \rangle$ be a group and $H = \langle Y \rangle$ a finitely generated subgroup. Define the set $A \coloneqq X \cup X^{-1}$.

(i). The *coset table* $\mathfrak{C}$ is the quintuple (τ, χ, p, n, M) where

- $M \in \mathbb{N}$, is the largest number of different cosets we allow;
- $n < M$, $n \in \mathbb{N}$ is the number of different cosets that we have found so far;
- τ is a map $\tau : [1, \dots, n] \to A^*$ that gives the representative $x \in A^*$ for the coset with number $\alpha \in [1, \dots, n]$;
- p is a map $p : [1, \dots, n] \to [1, \dots, n]$ with $p(\alpha) \le \alpha$ for all $\alpha \in [1, \dots, n]$ recording if two unequal numbers represent the same coset;
- χ is a partial map

$$\begin{aligned} \chi : [1, \dots, n] \times A &\to [1, \dots, n] \\ (\alpha, x) &\mapsto \alpha^x \, . \end{aligned}$$

(ii). The set $\Omega \coloneqq \{\alpha \in [1, \dots, n] \mid p(\alpha) = \alpha\}$ is called the *set of live cosets*.

(iii). A coset table $\mathfrak{C}$ is called *complete* if it does not have undefined values, i.e. χ is defined for all $\alpha \in [1, \dots, n]$ and for all $x \in A$.

Remark 4.13.

(i). The notation $[1, \ldots, n]$ is used for an ordered list of all consecutive elements in $\mathbb{N}$ from 1 to n. We need it because it we need to define the coset numbers consecutively for the enumeration to make sense.

(ii). Apart from the value M, all components of the coset table C usually change throughout the procedure.

(iii). We will see that our procedure has a flaw: It can happen that the cosets with the numbers α and β in $[1, \ldots n]$ coincide, although $\alpha < \beta$. That is why we need the map p. We simply define $p(\beta) = \alpha$, since then β is removed from the set of live cosets Ω.

(iv). We require in the coset table before starting the procedure the initial values $n = 1$, $p(1) = 1$ and $\tau(1) = \varepsilon$.

As the name suggests, coset tables are usually visualized in a table. The following example will illustrate how and furthermore heuristically explain how the procedure works, before defining it in general.

Example 4.11. Let $G = \langle x, y \mid x^3 = y^3 = 1, xy = yx \rangle$ and let $H = \langle x \rangle \leq G$ be a subgroup. We want to show that $|G : H| = 3$.
We always start by representing the trivial coset H with the number 1, hence the coset representative given by $\tau(1)$ is the trivial element, as required. Now we need to *scan* 1 under all the elements of $A = \{x, x^{-1}, y, y^{-1}\}$, which means that we use the partial map χ and evaluate the image of the form α^x by multiplying the element from the right to the coset that is represented by the value α. For 1 we get: 1^x represents Hx, which is simply H, since $x \in H$. Hence $1^x = 1$. Analogously, since $x^{-1} \in H$, we obtain $1^{x^{-1}} = 1$. Next, we scan 1 under y and under y^{-1}. 1^y represents Hy, which is a coset appearing for the first time. Thus, it does not have a number yet and we represent it by 2, i.e. $1^y = 2$. Analogously we obtain that $1^{y^{-1}}$ represents the coset Hy^{-1}, which we represent by 3. So for now, we obtain the coset table

	x	x^{-1}	y	y^{-1}
1	1	1	2	3
2				
3				

Next, we see that $2^{y^{-1}} = 1$, since $Hyy^{-1} = H$ and that $3^y = 1$, since $Hy^{-1}y = H$.

But 2^y and $3^{y^{-1}}$ is undefined:

	x	x^{-1}	y	y^{-1}
1	1	1	2	3
2				1
3			1	

Now, our second tool, called *deduction* can be used. Since we know that $1^{y^3} = 1$, as $Hy^3 = H$, we can use the scans that we know so far to deduce

$$1 \xrightarrow{y} 2 \xrightarrow{\underline{y}} 3 \xrightarrow{y} 1$$

The underlined scan is the one we deduced, which is $2^y = 3$. Furthermore we obtain the inverse, which is $3^{y^{-1}} = 2$. Hence the current coset table is

	x	x^{-1}	y	y^{-1}
1	1	1	2	3
2			3	1
3			1	2

Lastly, we can use the fact that $1^{x^{-1}y^{-1}xy} = 1$ and $2^{x^{-1}y^{-1}xy} = 2$ as the relations imply that $x^{-1}y^{-1}xy$ is the trivial word and make the following deductions:

$$1 \xrightarrow{x^{-1}} 1 \xrightarrow{y^{-1}} 3 \xrightarrow{\underline{x}} 3 \xrightarrow{y} 1$$

which implies $3^x = 3$ and thus $3^{x^{-1}} = 3$ and

$$2 \xrightarrow{\underline{x^{-1}}} 2 \xrightarrow{y^{-1}} 1 \xrightarrow{x} 3 \xrightarrow{y} 2$$

which implies $2^x = 2$ and thus $2^{x^{-1}} = 2$. So we obtain the complete coset table

	x	x^{-1}	y	y^{-1}
1	1	1	2	3
2	2	2	3	1
3	3	3	1	2

which we know to be correct after checking that for every $\alpha \in [1, \ldots, n]$ and for every relator $w \in R$, we get $\alpha^w = \alpha$. Hence, the index $|G : H|$ of H in G is equal to the number of cosets defined, which is 3.

Studying the introductory example carefully lets us quickly filter out four properties that the procedure might or should have:

Definition 4.30. Let $\mathfrak{C}$ be the coset table for a group G with respect to the subgroup H. We define the following properties for $\mathfrak{C}$

Property 1 $1 \in \Omega$ and $\tau(1) = \varepsilon$.

Property 2 $\alpha^x = \beta \Leftrightarrow \beta^{x^{-1}} = \alpha$.

Property 3 If $\alpha^x = \beta$, then $H\tau(\alpha)x = H\tau(\beta)$.

Property 4 For all $\alpha \in \Omega$, $1^{\tau(\alpha)}$ is defined and it is equal to α.

We are ready to describe the procedure for the general case in detail now: Since we start the procedure with an, besides the initial values, empty table, we first need to determine how to define new elements in the set of live coset Ω. So we run through the table and check for empty entries until we have found the first $\alpha \in \Omega$ and $x \in A$ such that α^x is undefined. Then we simply choose the first β that has so far not been an element of Ω, adjoin it to Ω and define it to be equal to α^x, i.e. $\alpha^x = \beta$. Moreover we put $p(\beta) := \beta$ and we define $\tau(\beta) \coloneqq \tau(\alpha)x$.

Clearly, Property 1 and Property 3 are satisfied. To ensure that Property 2 holds, we simply define $\beta^{x^{-1}} \coloneqq \alpha$. So we are left to check Property 4: Suppose $1^{\tau(\beta)} \neq \beta$. That implies that $1^{\tau(\alpha)x} \neq \alpha^x$, thus $\tau(\alpha)$ has to end with x^{-1}, i.e. there is a word v such that $\tau(\alpha) = vx^{-1}$. But then we obtain, since Property 4 holds for α by assumption, $\alpha = 1^{\tau(\alpha)} = 1^{vx^{-1}}$ and hence by Property 2 $\alpha^x = 1^v$ in contradiction to the fact that α^x has been undefined. So Property 4 has to hold as well.

Next, we want to define the scanning and the deduction: There are two settings in which it makes sense to scan $\alpha \in \Omega$ under $w \in A^*$, since we know what outcome to require:

1) $\alpha = 1$ and $w \in Y$: If $w \in Y$, then we know that $Hw = H$. So, since H is always represented by 1, we know that $1^w = 1$ has to hold.

2) $\alpha \in \Omega$ and $w \in R$: If $w \in R$, then we know that $Hgw = Hg$ for all $g \in G$. Hence we know for any $\alpha \in \Omega$ that $\alpha^w = \alpha$ has to hold.

Suppose we have a pair α and w given in one of the two settings. Then the scanning has two steps:

(i). Locate the largest prefix $s \in A^*$ of w such that α^s is defined. The word w can be written as $w = sv$ with $v \in A^*$.

(ii). Locate the largest suffix $t \in A^*$ of v such that $\alpha^{t^{-1}}$ is defined. The word v can be written as $v = ut$ for $u \in A^*$, hence $w = sut$.

There are three possible cases:

(1) $t = v$: In this case $w = st$ and we say that the scan *completes*. There are two possibilities:

   (a) $\alpha^s = \alpha^{t^{-1}}$: We we say the scan completes *correctly*.

   (b) $\alpha^s \neq \alpha^{t^{-1}}$: We say the scan completes *incorrectly*.

(2) $v = ut$ and $|u| = 1$: Then there is an element $x \in A$ such that x represents the word u. We say that the scan is *incomplete*, but a *deduction* is possible: We can compute

$$\alpha^s =: \beta \quad \text{and} \quad \alpha^{t^{-1}} =: \gamma$$

and define

$$\beta^x := \gamma \quad \text{and} \quad \gamma^{x^{-1}} = \beta\,.$$

(3) $v = ut$ and $|u| \neq 1$: We say the scan is *incomplete*.

Proposition 4.21. *Let the assignments $\beta^x = \gamma$ and $\gamma^{x^{-1}} = \beta$ result from a deduction of scanning an element $\alpha \in \Omega$ under a word $w \in A^*$ described above. Then the Properties 1, 2, 3 and 4 remain true after making the deductions.*

Proof. It is obvious that the deductions do not have any effect on Property 1 and 4 and that the Property 2 is satisfied by construction. So we are left to show that Property 3 holds, i.e. we need to prove that $H\tau(\beta)x = H\tau(\gamma)$. Since we know that $\alpha^x = \beta$ and $\gamma^t = \alpha$ before the deduction, Property 3 implies at that stage that $H\tau(\alpha)x = H\tau(\beta)$ and $H\tau(\gamma)t = H\tau(\gamma)\alpha$. We need to distinguish between the two possible settings:

1) Let $\alpha = 1$ and $w \in Y$: Then $H\tau(\alpha) = H$ and we obtain

$$H\tau(\beta) = H\tau(\alpha)s = Hs = H(sxt)^{-1}s = Ht^{-1}x^{-1}s^{-1}s =$$
$$Ht^{-1}x^{-1} = H\tau(\alpha)t^{-1}x^{-1} = H\tau(\gamma)tt^{-1}x^{-1} = H\tau(\gamma)x^{-1}\,.$$

Thus, $H\tau(\beta)x = H\tau(\gamma)$.

2) Let $\alpha \in \Omega$ and $w \in R$: Since $1_G = w = sxt$, we obtain that $tsx = 1_G$ and hence

$$H\tau(\beta)x = H\tau(\alpha)sx = H\tau(\gamma)tsx = H\tau(\gamma)\,.$$

So either way, Property 3 holds after the deductions as well. $\square$

Example 4.12. Let $G = \langle x, y \mid x^2 = 1, y^3 = 1, (xy)^3 = 1\rangle$ be a group with a subgroup H generated by $Y := \{xy\}$. Then $|G : H| = 4$.

Proof. We obtain the set $A = \{x, x^{-1}, y, y^{-1}\}$, but since the relation $x^2 = 1$ implies $x = x^{-1}$, the column of x and x^{-1} will be exactly the same and therefore we can omit x^{-1} in our table. As usual, let 1 represent the trivial coset H. Now we define $1^x = 2$, hence 2 represents the coset Hx and with it we get $1 = 2^{x^{-1}} = 2^x$. So the current table is

	x	y	y^{-1}
1	2		
2	1		

Now we scan 1 under $w = xy \in Y$:

(i). Since $1^x = 2$ is defined, and $1^{xy} = 2^y$ is not, we get $s = x$ and $v = y$.

(ii). Since $1^{y^{-1}}$ is not defined, we get $t = \varepsilon$ and $u = y$.

Thus, we are in Case (2) and obtain from $1^s = 1^x = 2$ and $1^{t^{-1}} = 1^{\varepsilon} = 1$ the deductions

$$2^y = 1 \quad \text{and} \quad 1^{y^{-1}} = 2\,.$$

In a diagram we can illustrate the deduction as

$$\begin{array}{ccccc} & x & & y & \\ 1 & & 2 & & 1 \\ & & & - & \end{array}$$

Hence, current coset table is

	x	y	y^{-1}
1	2		2
2	1	1	

Since there is no other useful scan possible, we next define $1^y = 3$ and thus $3^{y^{-1}} = 1$. Now we scan 1 under the word $y^3 \in R$:

(i). Since $1^y = 3$ is defined, and $1^{y^2} = 3^y$ is not, we get $s = y$ and $v = y^2$.

(ii). Since $1^{y^{-1}} = 2$ is defined and $1^{y^{-2}} = 2^{y^{-1}}$ is not, we get $t = y$ and $u = y$.

Thus, we are in Case (2) and obtain from $1^s = 1^y = 3$ and $1^{t^{-1}} = 1^{y^{-1}} = 2$ the deductions

$$3^y = 2 \quad \text{and} \quad 2^{y^{-1}} = 3\,.$$

In a diagram we can illustrate the deduction as

$$\begin{array}{ccccccc} & x & & y & & y & \\ 1 & & 3 & & 2 & & 1 \end{array}$$

—

The current table is

	x	y	y^{-1}
1	2	3	2
2	1	1	3
3		2	1

Next, we define $3^x = 4$ and thus $4^x = 3$ and scan 2 under $(xy)^3 \in R$:

(i). Since $2^{xyx} = 1^{yx} = 3^x = 4$ is defined, and $1^{xyxy} = 4^y$ is not, we get $s = xyx$ and $v = yxy$.

(ii). Since $2^{y^{-1}x^{-1}} = 3^{x^{-1}} = 3^x = 4$ is defined and $2^{y^{-1}x^{-1}y^{-1}} = 4^{y^{-1}}$ is not, we get $t = xy$ and $u = y$.

Thus, we are in Case (2) and obtain from $2^s = 2^{xyx} = 4$ and $2^{t^{-1}} = 2^{y^{-1}x^{-1}} = 4$ the deductions

$$4^y = 4 \quad \text{and} \quad 4^{y^{-1}} = 4\,.$$

In a diagram we can illustrate the deduction as

$$\begin{array}{ccccccccccccc} & x & & y & & x & & y & & x & & y & \\ 2 & & 1 & & 3 & & 4 & & 4 & & 3 & & 2 \end{array}$$

—

The current table is

	x	y	y^{-1}
1	2	3	2
2	1	1	3
3	4	2	1
4	3	4	4

and complete. □

Theorem 4.22. *Let $G = \langle X \mid R \rangle$ be a group with subgroup H generated by the set Y and let $\mathfrak{C}$ be a coset table. If*

(i). Property 1 to 4 are satisfied,

(ii). the coset table $\mathfrak{C}$ is complete,

(iii). 1 scans correctly under all $w \in Y$,

(iv). all $\alpha \in \Omega$ scan correctly under all $w \in R$,

then $|G : H| = |\Omega|$.
Furthermore for every $x \in A = X \cup X^{-1}$, the map

$$\begin{aligned} \varphi(x) : \Omega &\to \Omega \\ \alpha &\mapsto \alpha^x \end{aligned}$$

is a permutation of Ω and φ extends to a homomorphism $\bar{\varphi}(x) : G \to Sym(\Omega)$, which is equivalent to the action of G on the cosets of H by right multiplication.

Proof. Since $\alpha^x = \alpha^x$ trivially holds, Property 2 implies that $\alpha^{xx^{-1}} = \alpha$, hence the maps $\varphi(x)$ and $\varphi(x^{-1})$ are inverse to each other and φ is a permutation of Ω.
Next, take $w = x_1 \cdots x_r \in R$. Clearly $\varphi(w) = \varphi(x_1 \cdots x_r) = \varphi(x_1) \cdots \varphi(x_r)$ and by (iv), $\alpha^w = \alpha$. Hence $\varphi(x_1) \cdots \varphi(x_r) = 1_{Sym(\Omega)}$. Thus, φ extends by Theorem 2.52 of [19, p. 37] to a group homomorphism and defines therefore a group action

$$\begin{aligned} \varphi : G &\to Sym(\Omega) \\ x &\mapsto \varphi(x)\,. \end{aligned}$$

To show that this action is equivalent to the action ρ of G on the set $C = \{Hg \mid g \in G\}$ of right cosets of H given by

$$\begin{aligned} \rho : G &\to Sym(C) \\ x &\mapsto \rho(x) : C \to C \\ &\qquad\qquad Hg \mapsto Hgx \end{aligned}$$

by Definition 2.8 we need to find a bijection $t : \Omega \to C$ such that

$$t(\varphi(x)(\alpha)) = \rho(x)(t(\alpha))$$

holds for every $\alpha \in \Omega$ and every $g \in G$. Let us define t and check the properties: Let t be the map

$$\begin{aligned} t : \Omega &\to C \\ \alpha &\mapsto H\tau(\alpha)\,. \end{aligned}$$

Then by (ii) and Property 3, we have for every $v \in A^*$

$$t(1^v) = H\tau(1^v) = H\tau(1)v = Hv\,.$$

Hence, t hits every coset in C, and is therefore surjective. For injectivity suppose that $t(\alpha) = t(\beta)$. We need to show that $\alpha = \beta$. First deduce from $t(\alpha) = t(\beta)$ that $H\tau(\alpha) = H\tau(\beta)$, hence $t(\alpha)t(\beta)^{-1} \in H$. Therefore there are $w_1, w_2, \ldots, w_r \in Y \cup Y^{-1}$ such that $t(\alpha)t(\beta)^{-1} = w_1 w_2 \cdots w_r$. But by (iii), $1^{w_i} = 1$ for all $w_i \in Y \cup Y^{-1}$, thus $1^{\tau(\alpha)\tau(\beta)^{-1}} = 1$. By Property 2 and 4 we obtain $\alpha = 1^{\tau(\alpha)} = 1^{\tau(\beta)} = \beta$, hence t is injective. So we obtain

$$|\Omega| = |C| = |G : H|\,.$$

Lastly, we see by Property 3 that

$$t(\alpha^g) = H\tau(\alpha^g) = H\tau(\alpha)g = t(\alpha)^g$$

holds for every $\alpha \in \Omega$ and for every $g \in G$, hence the actions φ and ρ are equivalent. □

One problem has been unaddressed so far: How to proceed when a scan completes incorrectly. This means, if we scan $\alpha \in \Omega$ under w that w is of the form $w = st$, but $\alpha^s =: \beta \neq \gamma := \alpha^{t^{-1}}$. We can show that that simply means that β and γ represent the same coset:

Proposition 4.23. *Suppose that either $\alpha \in \Omega$ and $w \in R$ or $\alpha = 1$ and $w \in Y$ and that w is of the form $w = st$ with $\beta = \alpha^s$ and $\gamma = \alpha^{t^{-1}}$. Then $H\tau(\beta) = H\tau(\gamma)$.*

Proof. Since $\alpha^s = \beta$ we have by Property 3 $H\tau(\alpha)s = H\tau(\beta)$ and since $\gamma^t = \alpha$ we have $H\tau(\gamma)t = H\tau(\alpha)$. We need to distinguish between the two cases:

1) If $w = st \in H$ and $\alpha = 1$, then $H\tau(\alpha) = H$. Hence

$$H\tau(\beta) = H\tau(\alpha)s = Hs = H(st)^{-1}s = Ht^{-1}s^{-1}s = \\ Ht^{-1} = H\tau(\alpha)t^{-1} = H\tau(\gamma)tt^{-1} = H\tau(\gamma)\,.$$

2) If $w = st \in R$, then $ts = 1$. Hence

$$H\tau(\beta) = H\tau(\alpha)s = H\tau(\gamma)ts = H\tau(\gamma)\,.$$

Thus, either way, the claim is true. □

Definition 4.31. Let $G = \langle X \mid R \rangle$ be a group with subgroup H generated by the set Y. Let $\alpha \in \Omega$ scan incorrectly under the word $w = st \in A^*$ such that $\alpha^s = \beta$ and $\alpha^{t^{-1}} = \gamma$. Then we say there is a *coincidence* between β and γ. We denote it by $\beta \sim \gamma$.

Definition 4.32.

(i). Let G be a group acting on a set Ω by

$$\begin{aligned}\varphi : \Omega \times G &\to \Omega\\ (\alpha, g) &\mapsto \alpha^g\end{aligned}$$

and let $\sim$ be an equivalence relation on Ω. $\sim$ is called a *G-congruence* if for all $\alpha, \beta \in \Omega$ and for all $g \in G$, $\alpha \sim \beta$ implies that $\alpha^g = \beta^g$.

(ii). Let $S \subseteq \Omega \times \Omega$ be a set. The *G-congruence generated by S* is the intersection of all G-congruences containing S.

Remark 4.14.

(i). Being coincident is an equivalence relation.

(ii). Consider β^x and γ^x, both defined for some $x \in A$ where $\beta \sim \gamma$. Then, by Property 3 we have $H\tau(\beta)x = H\tau(\gamma)x$ and hence $H\tau(\beta^x) = H\tau(\gamma^x)$. By induction on the word length, we obtain that if $\beta \sim \gamma$, then $H\tau(\beta^w) = H\tau(\gamma^w)$ for any word $w \in G$. So, putting aside that α^w is in general not defined for all $w \in A^*$ and $\alpha \in \Omega$, $\sim$ can be considered to be a G-congruence.

We still need to implement a way to eliminate coincidences in our procedure: If we find a coincidence by an incorrectly completed scan, we need to compute all by the G-congruence implied coincidences. That is, if there is a coincidence $\beta \sim \gamma$, we need to compute β^z and γ^z for all $z \in A$ and deduce that $\beta^z \sim \gamma^z$. Next, we need to make sure to not lose any information. So we substitute every occurrence of the redundant elements in the coset table by their coinciding elements. Lastly, we will eliminate all redundant elements of Ω by using the function p: As indicated in Remark 4.13 (iii), if we have the coincidence $\alpha \sim \beta$ and $\alpha < \beta$, then we set $p(\beta) = \alpha$, since then by definition of the set Ω, the element β is eliminated from it.
An example should illustrate how the producure works:

Example 4.13. Let $G = \langle x, y \mid x^2y^2, x^3y^5\rangle$ be a group with $H = \{1_G\}$ being the trivial subgroup, i.e. its generating set Y is empty. Then $|G : H| = 4$.

Proof. As usual, $A = \{x, x^{-1}, y, y^{-1}\}$ and 1 represents the trivial coset H. Define $1^x = 2$ and thus $2^{x^{-1}} = 1$, $2^x = 3$ and thus $3^{x^{-1}} = 2$ and $3^y = 4$ and thus $4^{y^{-1}}$.

So the current coset table is

	x	x^{-1}	y	y^{-1}
1	2			
2	3	1		
3		2	4	
4				3

We can make the following deduction by scanning 1 under $x^2y^2 \in R$:

$$\begin{array}{ccccccccc} & x & & x & & y & & x & \\ 1 & & 2 & & 3 & & 4 & & 1 \\ & & & & & & & - & \end{array}$$

hence $4^y = 1$ and $1^{y^{-1}} = 4$. Next define $3^x = 5$ and thus $5^{x^{-1}} = 3$ and $5^y = 6$ and thus $6^{y^{-1}} = 5$. The current coset table is

	x	x^{-1}	y	y^{-1}
1	2			4
2	3	1		
3	5	2	4	
4			1	3
5		3	6	
6				5

Scanning 2 under $x^2y^2 \in R$ gives with

$$\begin{array}{ccccccccc} & x & & x & & y & & x & \\ 2 & & 3 & & 5 & & 6 & & 2 \\ & & & & & & & - & \end{array}$$

the deduction $6^y = 2$ and $2^{y^{-1}} = 6$,

scanning 1 under $x^3y^5 \in R$ gives with

$$\begin{array}{ccccccccccccccccc} & x & & x & & x & & y & & y & & y & & y & & y & \\ 1 & & 2 & & 3 & & 5 & & 6 & & 2 & & 3 & & 4 & & 1 \\ & & & & & & & & & & & - & & & & & \end{array}$$

the deduction $2^y = 3$ and $3^{y^{-1}} = 2$

and scanning 3 under $x^2y^2 \in R$ gives with

$$3 \xrightarrow{x} 5 \xrightarrow[-]{x} 6 \xrightarrow{y} 2 \xrightarrow{x} 3$$

the deduction $5^x = 6$ and $6^{x^{-1}} = 5$. The current coset table is

	x	x^{-1}	y	y^{-1}
1	2			4
2	3	1	3	6
3	5	2	4	2
4			1	3
5	6	3	6	
6		5	2	5

But if we now scan 2 under $x^3y^5 \in R$ we see that the scan completes incorrectly

$$2 \xrightarrow{x} 3 \xrightarrow{x} 5 \xrightarrow{x} 6 \xrightarrow{y} 2 \xrightarrow{y} 3 \xrightarrow{y} 4 \xrightarrow{y} 1 \mid 6 \xrightarrow{y} 2$$

so we have detected a coincidence between 1 and 6. Computing 1^z and 6^z for all $z \in A$ gives us a for y^{-1} the second coincidence

$$5 = 6^{y^{-1}} \sim 1^{y^{-1}} = 4\,.$$

Thus, we need to substitute every occurrence of 6 in the coset table by 1 and every occurrence of 5 by 4 and finally remove 5 and 6 from Ω. Hence we get the coset table

	x	x^{-1}	y	y^{-1}
1	2	4	2	4
2	3	1	3	1
3	4	2	4	2
4	1	3	1	3

$\square$

Although the procedure for an occurrence of a coincidence is just described heuristically, for full disclosure we want to give the following statement insuring that the procedure works:

Theorem 4.24. *[19, p. 158, Proposition 5.4] After executing the procedure to get rid of the coincidence of α and β, only the least representatives of the equivalence classes of the G-congruence $\sim$ generated by α and β remain in Ω and*

$p(v) = v$ for every $v \in [1, \dots, n]$ if and only if $v \in \Omega$. Furthermore, if for some $x \in A$ we have $\gamma^x = \delta$ before executing the procedure, then the smallest equivalent representative of γ to the x is equal to the smallest equivalent representative of δ. Also, Property 1 to 4 all continue to hold after executing the procedure.

To get more uniform results when running the procedure we introduce:

Definition 4.33. A *Standardized Coset Table* $\mathfrak{C}$, is a coset table that is reordered such that if we scan the table first by elements of Ω and then by elements of A, the cosets occur in ascending numerical order.

Next, we want to deal with a problem that can be interpreted as the converse of Coset Enumeration: We want to find all subgroups up to a given index. Since the time required for solving a problem of that form grows more than exponentially in the given index N, it is called *Low-Index Subgroups Problem.* A classic example that illustrates that fact will be the Heineken group H.
There are two well-known fundamentally different algorithms that attempt to solve the Low Index Subgroup Problem. Both were first introduced by Charles Sims. In general it is dependent on the group and its presentation, which algorithm is more efficient, but there is no proper rule stating which one should be applied. In dealing with standard examples of groups, Holt claims in [19] that the second algorithm is more efficient. It is implemented in MAGMA. Nonetheless, the widely used open source program GAP still uses the first algorithm. Therefore we want to introduce both heuristically.

The setting is the same as for the Coset Enumeration: Let forthcoming $G = \langle X \mid R \rangle$ be a group and let A be the set $A := X \cup X^{-1}$.

Definition 4.34. We call a (standardized) coset table $\mathfrak{C}$ *complete* for G if

- all columns are indexed by A,
- $\alpha^x = \beta$ if and only if $\beta^{x^{-1}} = \alpha$ for all $\alpha, \beta \in \Omega$ and for all $x \in A$, i.e. Property 2 holds,
- all $\alpha \in \Omega$ scan correctly under all $w \in R$.

The following statement ensures that with a standardized complete coset table we can really find a subgroup up to a given index:

Proposition 4.25. *[19, p. 189, Proposition 5.6] For any $n > 0$ there is a one-to-one-correspondence between the standardized complete coset tables for G with $n = |\Omega|$ and the subgroups of G with index n. In particular the subgroup H corresponding to a coset table $\mathfrak{C}$ is the stabilizer G_1 in the group action corresponding to $\mathfrak{C}$.*

Both algorithms search for a standardized complete coset table:
The first algorithm was invented in 1974 by Sims and fully formulated and presented by Joachim Neubüser in 1981 in [36]. Its idea is to use Coset Enumeration, but let the procedure run until more than the desired N cosets are found and then force coincidences. Concretely, we define a number $f(N) \in (N, 2N) \cap \mathbb{N}$, which is the maximal number of cosets we allow in a coset table. We start with an empty coset table and fill it by enumerating the cosets of the trivial group $\{1_g\}$ with the Coset Enumeration. The procedure is executed until

1) The coset table is complete.

2) All possible numbers for cosets up until $f(N)$ are used and there are no further possible deductions and no more coincidence occurring.

We have reached the first level, level 0, of our search tree. There are three options

1A) The maximal coset number used is less or equal N. Then G itself is of order less or equal N. We need to display G as the first subgroup of index less or equal N, that we have found. Then we can continue by forcing a coincidence.

1B) The maximal coset number used is greater than N. We force a coincidence.

2) We need to force a coincidence.

Forcing a coincidence means, we choose two different numbers α and β in our coset table to represent the same coset of some subgroup H. That means that $H\tau(\alpha) = H\tau(\beta)$ and hence we can use $\tau(\alpha)\tau(\beta)^{-1} \in H$ as a generator.
We need to work through all possible forced coincidences - each represents a branch of the search tree at level 0. After a forced coincidence, we make all possible deductions and eliminate all further occurring coincidences. Then we rename the cosets, such that only consecutive numbers starting from 1 are used in the coset table and start executing the Coset Enumeration again until case 1) or case 2) is reached. Then we continue as before:
We have reached the second level, level 1, of our search tree. There are three options

1A) The maximal coset number used is less or equal N. We have found a subgroup of index less or equal N and we need to display it. Then we can continue by forcing a coincidence.

1B) The maximal coset number used is greater than N. We need to force a coincidence.

2) We need to force a coincidence.

Remark 4.15.

(i). In fact, further coincidences occurring with a forced coincidence can make the procedure in practice way easier. If we impose an ordering on the pairs of coset numbers that are possibly coinciding (often lexicographically in combination with the length of the coset representatives that belong to the coset numbers), we can do the following: If a forced coincidence implies a coincidence where the second pair is in this ordering less than the first pair, we can skip this case, i.e. branch of the search tree, since it will be covered by another branch in the search anyways. Also, the ordering ensures that the results we obtain are standardized. Further refinements also give us only one representative per conjugacy class of subgroups.

(ii). If we can continue this search it will eventually terminate: Since we know that per given finite index there are only finitely many subgroups of a finitely generated group, the search tree has to be finite.

We want to illustrate the procedure by means of the following example.

Example 4.14. Determine all subgroups of $G = \langle x, y \mid x^2, y^3 \rangle$ up to index $N = 4$.

Remark 4.16. To make matters easier we will neglect stating the ordering used. We will give, guided by [23], the search tree first, and then simply follow it to detect the subgroups. The fact that we have found all such subgroups in the end can be easily checked in GAP.

Proof. Since $x^2 = 1_G$, we can omit x^{-1} in the set A, thus $A = \{x, y, y^{-1}\}$. As usual 1 represents the trivial coset H. We start with the definition $H := \{1_G\}$, throughout the procedure H will change by the forced coincidences. Our search tree is the following, given in Figure 12:

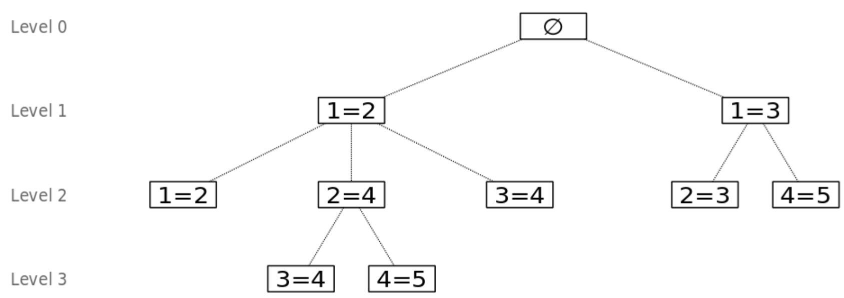


Figure 12: *This search tree gives all necessary forced coincidences.*

We let $f(N) = 7 \in (4,8) \cap \mathbb{N}$ and start the Coset Enumeration of H in G to obtain the following coset table $\mathfrak{C}_0$ at level 0:

$\mathfrak{C}_0$	x	y	y^{-1}
1	2	3	4
2	1	5	6
3	7	4	1
4		1	3
5		6	2
6		2	5
7	3		

We follow the search tree and exercise the first forced coincidence $1 \sim 2$. It implies $H = Hx$, hence $x \in H$ is a generator. We obtain by setting $1 \sim 2$ the coincidences $3 \sim 5$ and $4 \sim 6$, which lead to the left table. Then, by renaming and executing the Coset Enumeration again, the first coset table $\mathfrak{C}_{01}$ at level 1 can be determined:

	x	y	y^{-1}
1	1	3	4
3	7	4	1
4		1	3
7	3		

$\longrightarrow$

$\mathfrak{C}_{01}$	x	y	y^{-1}
1	1	2	3
2	4	3	1
3	5	1	2
4	2	6	7
5	3		
6		7	4
7		4	6

Since $\mathfrak{C}_{01}$ is incomplete, we force again the coincidence $1 \sim 2$. It implies $H = Hy^{-1}$, hence $y \in H$ is a generator. We obtain by setting $1 \sim 2$ the coincidences $1 \sim 4$, $2 \sim 4$, $1 \sim 3$, $1 \sim 5$, $3 \sim 6$ and $1 \sim 7$, hence all numbers represent the same coset and the first coset table $\mathfrak{C}_{011}$ at level 2 is complete and trivial:

$\mathfrak{C}_{011}$	x	y	y^{-1}
1	1	1	1

Thus, the coset table $\mathfrak{C}_{011}$ corresponds to the subgroup H_1 of G of index 1, which is generated by x and y, hence $H_1 = G$.
Since we have found a complete table, we return to $\mathfrak{C}_{01}$ at the previous level 1, and force a new coincidence, namely $2 \sim 4$. It implies $Hy = Hyx$, hence $yxy^{-1} \in H$ is a generator. The coincidence $2 \sim 4$ implies the coincidences $3 \sim 6$ and $1 \sim 7$, which lead to the left table. Renaming and executing the Coset Enumeration determines the second coset table $\mathfrak{C}_{012}$ at level 2:

	x	y	y^{-1}
1	1	2	3
2	2	3	1
3	5	1	2
5	3		

$\longrightarrow$

$\mathfrak{C}_{012}$	x	y	y^{-1}
1	1	2	3
2	2	3	1
3	4	1	2
4	3	5	6
5	7	6	4
6		4	5
7	5		

Since $\mathfrak{C}_{012}$ is incomplete, we impose with $3 \sim 4$ another coincidence. It implies $Hy^{-1} = Hy^{-1}x$, hence $y^{-1}xy \in H$. The coincidence $3 \sim 4$ implies the coincidences $1 \sim 5$, $2 \sim 6$ and $1 \sim 7$. Thus, we obtain the first coest table $\mathfrak{C}_{0121}$ at level 3, which is complete:

$\mathfrak{C}_{0121}$	x	y	y^{-1}
1	1	2	3
2	2	3	1
3	3	1	2

So the complete coset table $\mathfrak{C}_{0121}$ corresponds to the subgroup H_2 of G of index 3, which is generated by x, yxy^{-1} and $y^{-1}xy$.
We have found another complete coset table, so we return to $\mathfrak{C}_{012}$ at level 2 and force the coincidence $4 \sim 5$. It implies $Hy^{-1}x = Hy^{-1}xy$, hence $y^{-1}xy^{-1}xy \in H$ is a generator. The forced coincidence implies $3 \sim 7$, $5 \sim 6$ and $4 \sim 6$. We obtain the second coset table $\mathfrak{C}_{0122}$ at level 4 and it is also complete:

$\mathfrak{C}_{0122}$	x	y	y^{-1}
1	1	2	3
2	2	3	1
3	4	1	2
4	3	4	4

The coset table $\mathfrak{C}_{0122}$ corresponds to the subgroup H_3 of G of index 4, which is generated by x, yxy^{-1} and $y^{-1}xy^{-1}xy$.
Since we have found a complete coset table, we need to return to a previous level. Our search tree suggests to continue with $\mathfrak{C}_{01}$ at level 1. We force the coincidence $3 \sim 4$, which implies $Hy^{-1} = Hyx$, hence $y^{-1}xy^{-1} \in H$ is a generator. We obtain the implied coincidences $2 \sim 5$, $1 \sim 6$ and $2 \sim 7$. They determine the

third coset table $\mathfrak{C}_{013}$ at level 2, which turns out to be complete:

$\mathfrak{C}_{013}$	x	y	y^{-1}
1	1	2	3
2	3	3	1
3	2	1	2

The coset table $\mathfrak{C}_{013}$ corresponds to the subgroup H_4 of G of index 3, which is generated by x and $y^{-1}xy^{-1}$.

Now we return to $\mathfrak{C}_0$ at level 0 and force another coincidence, namely $1 \sim 3$. It implies that $H = Hy$, hence $y^{-1} \in H$ is a generator and that $2 \sim 7$, $3 \sim 4$ and $4 \sim 1$. These coincidences lead to the left table. After renaming and executing the Coset Enumeration, we obtain the second coset table $\mathfrak{C}_{02}$ at level 1:

	x	y	y^{-1}
1	2	1	1
2	1	5	6
5		6	2
6		2	5

$\longrightarrow$

$\mathfrak{C}_{02}$	x	y	y^{-1}
1	2	1	1
2	1	3	4
3	5	4	2
4	6	2	3
5	3	7	
6	4		
7			5

$\mathfrak{C}_{02}$ is incomplete, hence we impose $2 \sim 3$, which implies $Hx = Hxy$, i.e. $xy^{-1}x \in H$ is a generator. Furthermore we deduce the coincidences $1 \sim 5$, $3 \sim 4$, $4 \sim 2$, $1 \sim 6$ and $1 \sim 7$. So we obtain the forth coset table $\mathfrak{C}_{021}$ at level 2, which is complete:

	x	y	y^{-1}
1	2	1	1
2	1	2	2

The complete coset table $\mathfrak{C}_{021}$ corresponds to the subgroup H_5 of G of index 2, which is generated by y^{-1} and $xy^{-1}x$.

We return to $\mathfrak{C}_{02}$ at the previous level 1 and force the coincidence $4 \sim 5$. It implies $Hxy^{-1} = Hxyx$, thus $xy^{-1}xy^{-1}x \in H$ is a generator. We obtain with the implied coincidences $2 \sim 7$ and $3 \sim 6$ the fifth coset table $\mathfrak{C}_{022}$ at level 2 and it turns out to be complete:

$\mathfrak{C}_{02}$	x	y	y^{-1}
1	2	1	1
2	1	3	4
3	4	4	2
4	3	2	3

So the complete coset table $\mathfrak{C}_{022}$ corresponds to the subgroup H_6 of G of index 4, which is generated by y^{-1} and $xy^{-1}xy^{-1}x$. □

The second algorithm was introduced more recently in 1984 in Sims' book [44]. Its idea is to use incomplete coset tables and systematically search for a possible way to fill them by running through all elements of $\{1, \ldots, N\}$ and checking whether a contradiction occurs.

Remark 4.17. It is not efficient to store and output whole tables, so we can save some space and time by only returning coset tables corresponding to representatives of conjugacy classes of subgroups of G, i.e. per conjugacy class we only get one coset table. This works since the conjugates of a subgroup are then given by the stabilizers $G_\alpha = \{g \in G \mid \alpha^g = \alpha\}$ of the associated group action for the different $\alpha \in \Omega$.

Since the algorithm consists of five layers, we will not give a detailed explanation, but will just introduce it heuristically by means of an example:

Example 4.15. Determine all subgroups of $G = \langle x, y \mid x^2, y^3, (xy)^4 \rangle \cong S_4$ up to index $N = 4$.

Proof. Since $x = x^{-1}$, we can omit x^{-1} in A, which is thus given by $A = \{x, y, y^{-1}\}$.

- Start by imposing an ordering on the elements of A to determine in which order elements from A are used:

We impose

$$x < y < y^{-1}.$$

- Split the set R of relators into two disjoint sets R_1 and R_2 dependent on the length of the relators. R_1 should contain all "shorter" relators, since they are being used for the scaning procedure and R_2 should contain all "longer" relators, since they are being used for checking if the coset table is correct. It is not fully determined, what qualifies for short and long.

We define $R_1 = \{x^2, y^3\}$ and $R_2 = \{(xy)^4\}$.

- Set the default coset table $\mathfrak{C}_0$ to be empty.

We have

$\mathfrak{C}_0$	x	y	y^{-1}
1			

- Run through the coset table line by line until the first undefined, i.e. empty position is found.

The first undefined position in $\mathfrak{C}_0$ is 1^x.

- Define this position as the smallest available element $\alpha \in [1, \ldots, N] = [1, 2, 3, 4]$. Check that this definition is not a contradiction to Property 2 and there is no incorrect scan under any element of R_1. If so, put the new definition in the coset table and name the new coset table $\mathfrak{C}_1$.

Define $1^x =: 1$. Property 2 is not contradicted and there is no incorrect scan, hence we obtain

$\mathfrak{C}_1$	x	y	y^{-1}
1	1		

- Check if by Property 2 or scanning α under all words of R_1 any deductions can be made. Put them in the coset table.

There are no possible deductions.

- Check if the coset table is incomplete. If so, run through the coset table line by line until the first undefined, i.e. empty position is found.

The first undefined position in $\mathfrak{C}_1$ is 1^y.

- Define this position as the smallest available element $\alpha \in [1, \ldots, N] = [1, 2, 3, 4]$. Check that this definition is not a contradiction to Property 2 and there is no incorrect scan under any element of R_1. If so, put the new definition in the coset table and name the new coset table $\mathfrak{C}_{11}$.

Define $1^y = 1$. Property 2 is not contradicted and there is no incorrect scan, hence we obtain

$\mathfrak{C}_{11}$	x	y	y^{-1}
1	1	1	

- Check if by Property 2 or scanning α under all words of R_1 any deductions can be made. Put them in the coset table.

By Property 2 we deduce $1^{y^{-1}} = 1$. Hence we obtain

$\mathfrak{C}_{11}$	x	y	y^{-1}
1	1	1	1

- Check if the coset table is incomplete. If not, scan all elements of Ω under all elements of R_2 to make sure it is correct.

The table is complete. $1^{(xy)^4} = 1$, hence the table is correct. With $\mathfrak{C}_{11}$, we have found the first coset table.

- Return to the coset before the last new definition. This we can do by simply losing the last index of the coset table.

The coset table we return to is $\mathfrak{C}_1$.

- Run through the coset table line by line until the first undefined, i.e. empty position is found.

Again, the first undefined position in $\mathfrak{C}_1$ is 1^y.

- Define this position as the smallest available element $\alpha \in [1, \ldots, N] = [1, 2, 3, 4]$. Check that this definition is not a contradiction to Property 2 and there is no incorrect scan under any element of R_1. If so, put the new definition in the coset table and add to the index of the name of the current coset table the number of how often this position has been validly defined so far to name the new coset table.

Since we have used $1 \in [1, 2, 3, 4]$ already it is not available anymore. Thus, define $1^y =: 2$. Property 2 is not contradicted and there is no incorrect scan, hence we obtain

$\mathfrak{C}_{12}$	x	y	y^{-1}
1	1	2	
2			

- Check if by Property 2 or scanning α under all words of R_1 any deductions can be made. Put them in the coset table.

By Property 2 we decuce $2^{y^{-1}} = 1$. Hence we obtain

$\mathfrak{C}_{12}$	x	y	y^{-1}
1	1	2	
2			1

- Check if the coset table is incomplete. If so, run through the coset table line by line until the first undefined, i.e. empty position is found.

The first undefined position in $\mathfrak{C}_{12}$ is 1^y.

- Define this position as the smallest available element $\alpha \in [1, \ldots, N] = [1, 2, 3, 4]$. Check if this definition is a contradiction to Property 2 or if there is an incorrect scan under any element of R_1. If so, move on and define this position as the next smallest available element $\alpha \in [1, 2, 3, 4]$. Check again that this definition is not a contradiction to Property 2 and there is no incorrect scan under any element of R_1. Repeat this until there is an element $\alpha \in [1, 2, 3, 4]$ such that Property 2 is not contradicted and there is no incorrect scan under any element of R_1. Put this new definition in the coset table and add to the index of the name of the current coset table the number of how often this position has been validly defined so far to name the new coset table.

Define $1^y =: 1$. Then by Property 2, we get that $1^{y^{-1}} = 1$ must hold, in contradiction to $1^{y^{-1}} = 2$. So the definition is invalid and we define $1^{y^{-1}} = 2$. Property 2 is not contradicted and gives $2^y = 1$. Now we can scan 1 under y^3 and know that the result must be 1. But

$$1^{y^3} = 2^{y^2} = 1^y = 2\,,$$

hence we obtain a contradiction again. So the definition is invalid and we define $1^y =: 3$. Property 2 is not contradicted and there is no incorrect scan, hence we obtain

$\mathfrak{C}_{121}$	x	y	y^{-1}
1	1	2	3
2			1
3			

- Check if by Property 2 or scanning α under all words of R_1 any deductions can be made. Put them in the coset table.

By Property 2, we deduce $3^y = 1$. By scanning 1 under y^3 we deduce

$$\begin{array}{ccccccc} & y & & y & & y & \\ 1 & & 2 & & 3 & & 1 \\ & & & - & & & \end{array}$$

i.e. $2^y = 3$ and $3^{y^{-1}} = 2$. So the coset table $\mathfrak{C}_{121}$ is

$\mathfrak{C}_{121}$	x	y	y^{-1}
1	1	2	3
2		3	1
3		1	2

- Check if the coset table is incomplete. If so, run through the coset table line by line until the first undefined, i.e. empty position is found.

The first undefined position in $\mathfrak{C}_{121}$ is 2^x.

- Define this position as the smallest available element $\alpha \in [1, \ldots, N] = [1, 2, 3, 4]$. Check if this definition is a contradiction to Property 2 or if there is an incorrect scan under any element of R_1. If so, move on and define this position as the next smallest available element $\alpha \in [1, 2, 3, 4]$. Check again that this definition is not a contradiction to Property 2 and there is no incorrect scan under any element of R_1. Repeat this until there is an element $\alpha \in [1, 2, 3, 4]$ such that Property 2 is not contradicted and there is no incorrect scan under any element of R_1. Put this new definition in the coset table and add to the index of the name of the current coset table the number of how often this position has been validly defined so far to name the new coset table.

Define $2^x = 1$ Then by Property 2, we know that $1^x = 2$ must hold, in contradiction to $1^x = 1$. So the definition is invalid and we define $2^x =: 2$. Property 2 is not contradicted and there is no incorrect scan, hence we obtain

$\mathfrak{C}_{1211}$	x	y	y^{-1}
1	1	2	3
2	2	3	1
3		1	2

- Check if by Property 2 or scanning α under all words of R_1 any deductions can be made. Put them in the coset table.

There are no possible deductions.

- Check if the coset table is incomplete. If so, run through the coset table line by line until the first undefined, i.e. empty position is found.

The first undefined position in $\mathfrak{C}_{1211}$ is 3^x.

- Define this position as the smallest available element $\alpha \in [1, \ldots, N] = [1, 2, 3, 4]$. Check if this definition is a contradiction to Property 2 or if there is an incorrect scan under any element of R_1. If so, move on and define this position as the next smallest available element $\alpha \in [1, 2, 3, 4]$. Check again that this definition is not a contradiction to Property 2 and there is no incorrect scan under any element of R_1. Repeat this until there is an element $\alpha \in [1, 2, 3, 4]$ such that Property 2 is not contradicted and there is no incorrect scan under any element of R_1. Put this new definition in the coset table and add to the index of the name of the current coset table the number of how often this position has been validly defined so far to name the new coset table.

Define $3^x = 1$. Then by Property 2, we know that $1^x = 3$ must hold, in contradiction to $1^x = 1$. So the definition is invalid and we define $3^x = 2$. Then by Property 2, we know that $2^x = 3$ must hold, in contradiction to $2^x = 2$. So the definition is invalid and we define $3^x =: 3$. Property 2 is not contradicted and there is no incorrect scan, hence we obtain

$\mathfrak{C}_{12111}$	x	y	y^{-1}
1	1	2	3
2	2	3	1
3	3	1	2

- Check if the coset table is incomplete. If not, scan all elements of Ω under all elements of R_2 to make sure it is correct.

The coset table $\mathfrak{C}_{12111}$ is complete. Scan all $\alpha \in \Omega = \{1, 2, 3\}$ under all words $w \in R_2$: First scan 1 under $(xy)^4$. The result must be 1, but

$$1^{(xy)^4} = 1^{xyxyxyxy} = 1^{yxyxyxy} = 2^{xyxyxy} = 2^{yxyxy} = 3^{xyxy} = 3^{yxy} = 1^{xy} = 1^y = 2\,,$$

hence we obtain a contradiction. So the coset table $\mathfrak{C}_{12111}$ is incorrect.

- Return to the coset table before the last new definition. This we can do by simply losing the last index of the name of the coset table.

The coset we return to is $\mathfrak{C}_{1211}$.

- Run through the coset table line by line until the first undefined, i.e. empty position is found.

Again, the first undefined position in $\mathfrak{C}_{1211}$ is 3^x.

- Define this position as the smallest available element $\alpha \in [1, \ldots, N] = [1, 2, 3, 4]$. Check that this definition is not a contradiction to Property 2 and there is no incorrect scan under any element of R_1. If so, put the new definition in the coset table and add to the index of the name of the current coset table the number of how often this position has been validly defined so far to name the new coset table.

Since we have tried $1, 2, 3 \in [1, 2, 3, 4]$ already they are not available anymore. Thus, define $3^x =: 4$. Property 2 is not contradicted and there is no incorrect scan, hence we obtain

$\mathfrak{C}_{12112}$	x	y	y^{-1}
1	1	2	3
2	2	3	1
3	4	1	2
4			

- Check if by Property 2 or scanning α under all words of R_1 any deductions can be made. Put them in the coset table.

By Property 2, we deduce $4^x = 3$, hence we obtain

$\mathfrak{C}_{12112}$	x	y	y^{-1}
1	1	2	3
2	2	3	1
3	4	1	2
4	3		

- Check if the coset table is incomplete. If so, run through the coset table line by line until the first undefined, i.e. empty position is found.

The first undefined position in $\mathfrak{C}_{12112}$ is 4^y.

- Define this position as the smallest available element $\alpha \in [1, \ldots, N] = [1, 2, 3, 4]$. Check if this definition is a contradiction to Property 2 or if there is an incorrect scan under any element of R_1. If so, move on and define this position as the next smallest available element $\alpha \in [1, 2, 3, 4]$. Check again that this definition is not a contradiction to Property 2 and there is no incorrect scan under any element of R_1. Repeat this until there is an element $\alpha \in [1, 2, 3, 4]$ such that Property 2 is not contradicted and there is no incorrect scan under any element of R_1. Put this new definition in the coset table and add to the index of the name of the current coset table the number of how often this position has been validly defined so far to name the new coset table.

Define $4^y =: 1$. Then by Property 2, we know that $1^{y^{-1}} = 4$ must hold, in contradiction to $1^{y^{-1}} = 3$. Also for $4^y = 2$ and $4^y = 3$ we obtain a contradiction. So we define $4^y =: 4$. Property 2 is not contradicted and there is no incorrect scan, hence we obtain

$\mathfrak{C}_{121121}$	x	y	y^{-1}
1	1	2	3
2	2	3	1
3	4	1	2
4	3	4	

- Check if by Property 2 or scanning α under all words of R_1 any deductions can be made. Put them in the coset table.

By Property 2, we deduce $4^{y^{-1}} = 4$, hence we obtain

$\mathfrak{C}_{121121}$	x	y	y^{-1}
1	1	2	3
2	2	3	1
3	4	1	2
4	3	4	4

- Check if the coset table is incomplete. If not, scan all elements of Ω under all elements of R_2 to make sure it is correct.

The coset table $\mathfrak{C}_{121121}$ is complete. All $\alpha \in \Omega$ scan correctly under $(xy)^4$, hence the table is correct. With $\mathfrak{C}_{121121}$, we have found the second coset table.

- Return to the coset table before the last new definition. This we can do by simply losing the last index of the coset table.

The coset we return to is $\mathfrak{C}_{12112}$.

- Run through the coset table line by line until the first undefined, i.e. empty position is found.

Again, the first undefined position in $\mathfrak{C}_{12112}$ is 4^y.

- Define this position as the smallest available element $\alpha \in [1, \ldots, N] = [1, 2, 3, 4]$. If there is no element available, return to the coset table before the last new definition, by losing the last index of the coset table.

We have tried all $\alpha \in [1, 2, 3, 4]$ for 4^y, hence we need to return to $\mathfrak{C}_{1211}$.

- Run through the coset table line by line until the first undefined, i.e. empty position is found.

Again, the first undefined position in $\mathfrak{C}_{1211}$ is 3^x.

- Define this position as the smallest available element $\alpha \in [1, \ldots, N] = [1, 2, 3, 4]$. If there is no element available, return to the coset table before the last new definition, by losing the last index of the coset table.

We have tried all $\alpha \in [1, 2, 3, 4]$ for 3^x, hence we need to return to $\mathfrak{C}_{121}$.

- Run through the coset table line by line until the first undefined, i.e. empty position is found.

Again, the first undefined position in $\mathfrak{C}_{121}$ is 2^x.

- Continue the process as described before.

The next available definition is $2^x =: 3$. Property 2 is not contradicted and there is no incorrect scan. By Property 2 we can deduce $3^x = 2$. Hence we obtain

$\mathfrak{C}_{1212}$	x	y	y^{-1}
1	1	2	3
2	3	3	1
3	2	1	2

The coset table $\mathfrak{C}_{1212}$ is complete and all $\alpha \in \Omega$ scan correctly under all words of R_2, hence we have found the third coset table.

Return to $\mathfrak{C}_{121}$, the last coset table before the last definition. The next available definition for 2^x is $2^x =: 4$. It is not a contradiction to Propery 2 and there is no incorrect scan. By Property 2 we deduce $4^x = 2$. Hence we obtain

$\mathfrak{C}_{1213}$	x	y	y^{-1}
1	1	2	3
2	4	3	1
3		1	2
4	2		

The next undefined position is 3^x. Since $3^x = 1$ and $3^x = 2$ contradicts Property 2, we define $3^x = 3$. There is no contradiction to Property 2 and there is no incorrect scan.

- Check if the algorithm FIRSTINCLASS, which is described in [19, p.194] returns `true`: This algorithm checks whether the subgroup corresponding to the coset table is the canonical representative of the conjugacy class of subgroups. If it is not, i.e. FIRSTINCLASS returns `false`, the definition is invalid. Move on to the next available definition.

For the definition $3^x = 3$, FIRSTINCLASS returns `false` (in fact the table is in the same conjugacy class as $\mathfrak{C}_{121121}$ as stated in [19, p. 195, Example 5.10]), hence the definition is invalid. We are left with the possible definition $3^x = 4$, which is a contradiction to Property 2. Thus, we need to return to the coset table before the last new definition.

Return to $\mathfrak{C}_{12}$, where the next undefined position is 1^y. The next available $\alpha \in [1, 2, 3, 4]$ would be 3.

- Check if every $\alpha \in [1, 2, 3, 4]$ such that $\alpha < 3$ has occured in the coset table at least once. If not, the definition is unvalid. Return to the coset table before the last definition.

If we define $1^y =: 3$, then 2 would not occur in the coset table. Hence the definition is invalid and we need to return to C_0.
Define $1^x =: 2$ and deduce $2^x = 1$:

$\mathfrak{C}_2$	x	y	y^{-1}
1	2		
2	1		

Define $1^y = 1$ and deduce $1^{y^{-1}} = 1$:

$\mathfrak{C}_{21}$	x	y	y^{-1}
1	2	1	1
2	1		

Since $2^y = 1$ implies $1^{y^{-1}} = 2$ in contradiction to $1^{y^{-1}} = 1$, define $2^y =: 2$ and deduce $2^{y^{-1}} = 2$:

$\mathfrak{C}_{211}$	x	y	y^{-1}
1	2	1	1
2	1	2	2

The coset table $\mathfrak{C}_{211}$ is complete and all $\alpha \in \Omega$ scan correctly under all words of R_2, hence we have found the forth coset table.

Return to the coset table $\mathfrak{C}_{21}$, define $2^y =: 3$ and deduce $3^{y^{-1}} = 2$:

$\mathfrak{C}_{212}$	x	y	y^{-1}
1	2	1	1
2	1	3	
3			2

Since $2^{y^{-1}} = 1$ and $2^{y^{-1}} = 2$ both imply contradictions to Property 2, we try $2^{y^{-1}} = 3$. But scanning 2 under y^{-3} gives

$$2^{y^{-3}} = 3^{y^{-2}} = 2^{y^{-1}} = 3\,,$$

in contradiction to the desired result 2. Hence we define $2^{y^{-1}} =: 4$ and deduce $4^y = 2$. Furthermore by scanning 2 under y^3 we deduce

$$\begin{array}{ccccccc} & y & & y & & y & \\ 2 & & 3 & & 4 & & 2 \\ & & & - & & & \end{array}$$

i.e. $3^y = 4$ and $4^{y^{-1}} = 3$:

$\mathfrak{C}_{2121}$	x	y	y^{-1}
1	2	1	1
2	1	3	4
3		4	2
4		2	3

Since $3^x = 1$ and $3^x = 2$ both imply contradictions to Property 2, we try $3^x = 3$. But then FIRSTINCLASS returns `false`. Hence we define $3^x =: 4$ and deduce $4^x = 3$:

$\mathfrak{C}_{21211}$	x	y	y^{-1}
1	2	1	1
2	1	3	4
3	4	4	2
4	4	2	3

The coset table $\mathfrak{C}_{21211}$ is complete and all $\alpha \in \Omega$ scan correctly under all words of R_2, hence we have found the fifth coset table.

Return to the coset table $\mathfrak{C}_2$, since this is the last one where a new definition is still available. Define $1^y =: 2$ and deduce $2^{y^{-1}} = 1$:

$\mathfrak{C}_{22}$	x	y	y^{-1}
1	2	2	
2	1		1

Since $1^{y^{-1}} = 1$ implies a contradiction to Property 2, we try $1^{y^{-1}} = 2$. But scanning 1 under y^3 gives

$$1^{y^3} = 2^{y^2} = 1^{y^1} = 2\,,$$

in contradiction to the desired result 1. Hence we define $1^{y^{-1}} =: 3$ and deduce $3^y = 1$. Furthermore, by scanning 1 under y^3 we deduce

$$\begin{array}{ccccccc} & y & & y & & y & \\ 1 & & 2 & & 3 & & 1 \\ & & & - & & & \end{array}$$

i.e. $2^y = 3$ and $3^{y^{-1}} = 2$:

$\mathfrak{C}_{221}$	x	y	y^{-1}
1	2	2	3
2	1	3	1
3		1	2

Since $3^x = 1$ and $3^x = 2$ both imply contradictions to Property 2, we try $3^x = 3$. But then FirstInClass returns `false`. Hence we define $3^x =: 4$ and deduce $4^x = 3$:

$\mathfrak{C}_{221}$	x	y	y^{-1}
1	2	2	3
2	1	3	1
3	4	1	2
4	3		

Since $4^y = 1$, $4^y = 2$ and $4^y = 3$ all imply contradictions to Property 2, we try $4^y = 4$. But then FirstInClass returns `false`. Hence we need to return to coset table with the last definition, where a new one is available.

Return to $\mathfrak{C}_2$. Define $1^y =: 3$ and deduce $3^{y^{-1}} = 1$:

$\mathfrak{C}_{23}$	x	y	y^{-1}
1	2	3	
2	1		1
3			1

Since $1^{y^{-1}} = 1$ implies a contradiction to Property 2, we try $1^{y^{-1}} = 2$. But then FirstInClass returns `false`. Hence we try $1^{y^{-1}} = 3$ and deduce $3^y = 1$. But scanning 1 under y^3 gives

$$1^{y^3} = 3^{y^2} = 1^y = 3\,,$$

in contradiction to the desired result 1. Hence we define $1^{y^{-1}} =: 4$ and deduce $4^y = 1$. Furthermore by scanning 1 under y^3 we deduce

$$\begin{array}{ccccccc} & y & & y & & y & \\ 1 & & 3 & & 4 & & 1 \\ & & & - & & & \end{array}$$

i.e. $3^y = 4$ and $4^{y^{-1}} = 3$:

$\mathfrak{C}_{231}$	x	y	y^{-1}
1	2	3	4
2	1		
3		4	1
4		1	3

Since $2^y = 1$ implies a contradiction to Property 2, we try $2^y = 2$. But then FirstInClass returns `false`. Hence we try $2^y = 3$ and $2^y = 4$, which both give a contradiction to Property 2 as well.

Hence we return to $\mathfrak{C}_2$, but we cannot define $1^y = 4$, since then 3, which is less then 4, would not occur in the coset table. Thus, we need to return to $\mathfrak{C}_0$. But we cannot define $1^x = 3$, since then 2, which is less then 3, would not occur in the coset table. Hence with $\mathfrak{C}_{11}$, $\mathfrak{C}_{121121}$, $\mathfrak{C}_{1212}$, $\mathfrak{C}_{211}$ and $\mathfrak{C}_{21211}$ we have found all possible coset tables corresponding to a canonical representative of a subgroup of G up to index 4. □

Example 4.16. It turns out that the Heineken group H is an example of a group for which the Low Index Method becomes quickly unfeasible with increasing index n. Computing any subgroup of large index is almost impossible. The only chance we stand is by using the 2-generator presentation (34). Searching for all subgroups up to index 10 with the GAP-function `LowIndexSubgroupsOfFpGroup(H,n)`, where H still has the 3-generator presentation (33) takes already about 280 seconds on a standard desktop computer. Hence we use `SimpliedFpGroup(H)` to obtain the 2-generator presentation, with which the computation is finished in 11 seconds. We obtain a list of five subgroups up to index 10, given in the following table. None of them is (properly) normal.

	Index	Generators	Normal	#Relators
H_1	1	a, b	/	2
H_2	6	$a, b^{-1}a^2b^{-1}, b^{-1}a^{-1}b^{-2}, b^{-2}a^{-1}b^{-1}, ba^2b^{-2}, bab^{-3}$	no	12
H_3	5	$ba^{-1}, a^2b, a^{-2}b^{-1}a^{-1}, a^{-1}b^2, a^{-1}b^{-2}a^{-1}, b^{-1}ab^{-1}a^{-1}$	no	10
H_4	10	$ba^{-1}, b^{-1}ab^{-1}a^{-1}, b^{-1}a^{-1}b^{-1}a, b^{-2}a^{-2}, a^5, a^2b^{-2}a^{-1}, a^{-2}ba^{-1}b^{-1}a^{-1}, (a^{-1}b)^2a$	no	20
H_5	10	$ba^{-1}, a^{-1}ba^{-2}, b^{-1}ab^{-1}a^{-1}, a^5, a(ab)^2, (ab)^2a, ab^4, a^{-2}b^2a, a^{-2}b^{-1}ab$	no	20

A computation to verify with the Low Index Method whether the quotient isomorphic to A_5 is the smallest proper finite one, took more than 125 days on my desktop computer provided by the University and was therefore aborted. To my knowledge, no further computations can be done in the time and space frame of the devices accessible to me.
Nonetheless it seems like the Low Index Method has some useful application for the Heineken group: Joachim Neubüser proves in [37] using the Low Index Method applied together with the so-called "Schreier-Sims Algorithm", the "Reidemeister-Schreier Algorithm", the "Method of $SL(2, \cdots)$-Representations" and the "Nilpotent Quotient Algorithm" that there is a finite quotient of the Heineken group H of size $60 \cdot 2^{24}$. Furthermore Neubüser's student Hoppe suggests in [22] that it has finite quotients of size 960 and 3840 besides the ones of size 60 and 120. Although he gives a set of generators of those subgroups, it is not feasible for me to verify their index in GAP. But it is apparent that all four numbers 60, 120, 960, 3840 are of the form $60 \cdot 2^x$ where x is a natural number, in those cases we have $x \in \{0, 1, 4, 6\}$. It might be interesting to investigate further, whether those results can be obtained with similar methods as the quotient of size $60 \cdot 2^{24}$.

References

[1] Ian Agol, Daniel Groves, and Jason Fox Manning. "Residual finiteness, QCERF, and fillings of hyperbolic groups". In: *arXiv preprint arXiv:0802.0709* (2008).

[2] Asuman G. Aksoy and Sixian Jin. "The apple doesn't fall far from the (metric) tree: the equivalence of definitions". In: *arXiv preprint arXiv:1306.6092* (2013).

[3] Goulnara Arzhantseva. "Asymptotic approximations of finitely generated groups". In: *Extended Abstracts Fall 2012*. Springer, 2014, pp. 7–15.

[4] Gilbert Baumslag and Donald Solitar. "Some two-generator one-relator non-Hopfian groups". In: *Bulletin of the American Mathematical Society* 68.3 (1962), pp. 199–201.

[5] Oleg Bogopolski. *Introduction to group theory*. eng. EMS textbooks in mathematics. 2008.

[6] Brian H. Bowditch. "A course on geometric group theory". In: (2005). URL: `https://www.math.ucdavis.edu/~kapovich/280-2009/bhb-ggtcourse.pdf`.

[7] Martin R. Bridson and André Haefliger. *Metric spaces of non-positive curvature*. Vol. 319. Springer Science & Business Media, 2013.

[8] Matt Clay and Max Forester. "On the isomorphism problem for generalized Baumslag–Solitar groups". In: *Algebraic & Geometric Topology* 8.4 (2008), pp. 2289–2322. URL: `https://doi.org/10.2140/agt.2008.8.2289`.

[9] Donald J. Collins and Frank Levin. "Automorphisms and Hopficity of certain Baumslag-Solitar groups". In: *Archiv der Mathematik* 40.1 (1983), pp. 385–400.

[10] Tommaso Cremaschi and Juan Souto. "Discrete groups without finite quotients". In: *Topology and its Applications* 248 (2018), pp. 138–142.

[11] David B. A. Epstein and Derek F. Holt. "Computation in word-hyperbolic groups". In: (1998).

[12] David BA Epstein. *Word processing in groups*. AK Peters/CRC Press, 1992.

[13] Gil Goffer. *Bass-Serre theory*. 2018. URL: `https://www.youtube.com/watch?v=21Py6CmS5iw` (visited on 11/15/2018).

[14] Mikhael Gromov. "Hyperbolic groups". In: *Essays in group theory*. Springer, 1987, pp. 75–263.

[15] Allen Hatcher. *Algebraic topology.* eng. 1. publ.. Cambridge: Cambridge Univ. Press, 2002.

[16] Graham Higman. "A finitely generated infinite simple group". In: *Journal of the London Mathematical Society* 1.1 (1951), pp. 61–64.

[17] Graham Higman. "A finitely related group with an isomorphic proper factor group". In: *Journal of the London Mathematical Society* 1.1 (1951), pp. 59–61.

[18] Graham Higman, Bernhard Hermann Neumann, and Hanna Neuman. "Embedding theorems for groups". In: *Journal of the London Mathematical Society* 1.4 (1949), pp. 247–254.

[19] Derek F Holt, Bettina Eick, and Eamonn A O'Brien. *Handbook of computational group theory.* Chapman and Hall/CRC, 2005.

[20] Math Overflow: Dan Sălăjan, Derek Holt and Misha. *The role of the Automatic Groups in the history of Geometric Group Theory.* 2013. URL: `https://mathoverflow.net/questions/130306/the-role-of-the-automatic-groups-in-the-history-of-geometric-group-theory` (visited on 10/19/2019).

[21] Asian Mathematical Conference (1st : 1990 : University of Hong Kong) et al. *Proceedings of Asian Mathematical Conference, 1990, University of Hong Kong, August 14-18, 1990.* English. Conference Proceedings. "First Asian Mathematical Conference was jointly organized by the Southeast Asian Mathematical Society and Hong Kong Mathematical Society"–Pref. 1992.

[22] Andreas Hoppe. "Effiziente Algorithmen zur Berechnung von Elementarteilern ganzzahliger Matrizen-Implementation in GAP". PhD thesis. 1994.

[23] Marston Conder, Fields Institute. *Lecture 3: Computational and group-theoretic methods.* 2011. URL: `https://www.google.com/search?client=ubuntu&channel=fs&q=regmapspolytopes+lecture+3&ie=utf-8&oe=utf-8` (visited on 07/17/2019).

[24] D. L. Johnson. *Presentations of Groups.* Vol. 2nd ed. London Mathematical Society Student Texts Vol. 15. Cambridge University Press, 1997. URL: `http://search-ebscohost-com.uaccess.univie.ac.at/login.aspx?direct=true&db=nlebk&AN=570482&site=ehost-live`.

[25] Ilya Kapovich and Daniel T. Wise. "The equivalence of some residual properties of word-hyperbolic groups". In: *Journal of Algebra* 223.2 (2000), pp. 562–583.

[26] Aditi Kar. *A Mathematical Toolkit for Group Theorists*. 2019. URL: http://www.personal.rhul.ac.uk/udah/001/LectureNotes.pdf (visited on 09/23/2019).

[27] Andras Kriegl. *Algebraic topology*. 2008. URL: https://www.mat.univie.ac.at/~kriegl/Skripten/alg-top.pdf (visited on 05/14/2019).

[28] Gilbert Levitt. "On the automorphism group of generalized Baumslag–Solitar groups". In: *Geometry & Topology* 11.1 (2007), pp. 473–515. URL: https://doi.org/10.2140/gt.2007.11.473.

[29] Roger C. Lyndon and Paul E. Schupp. *Combinatorial group theory*. Springer, 2015.

[30] Wilhelm Magnus, Abraham Karrass, and Donald Solitar. *Combinatorial group theory: Presentations of groups in terms of generators and relations*. Courier Corporation, 2004.

[31] Math Stack Exchange: Max. *Proving that a group acting freely on a tree is a free group*. 2017. URL: https://math.stackexchange.com/questions/2298992/proving-that-a-group-acting-freely-on-a-tree-is-a-free-group (visited on 05/24/2019).

[32] Stephen Meskin. "Nonresidually finite one-relator groups". In: *Transactions of the American Mathematical Society* 164 (1972), pp. 105–114.

[33] DI Moldavanskii. "Isomorphism of the Baumslag-Solitar groups". In: *Ukrainian Mathematical Journal* 43.12 (1991), pp. 1569–1571. URL: https://doi.org/10.1007/BF01066696.

[34] Math Overflow: Edgar Ndie and YCor. *Are the Baumslag-Solitar groups BS(n,n) and BS(n,-n) automata groups?* 2015. URL: https://mathoverflow.net/questions/215126/are-the-baumslag-solitar-groups-bsn-n-and-bsn-n-automata-groups/215628 (visited on 09/11/2019).

[35] Math Stack Exchange: Edgar Ndie and Derek Holt. *Is Bs(1,-1) linear?* 2015. URL: https://math.stackexchange.com/questions/1414538/is-bs1-1-linear (visited on 08/17/2019).

[36] J. Neubüser. "An elementary introduction to coset table methods in computational group theory". eng. In: *Groups - St Andrews 1981*. Cambridge University Press, 1982, pp. 1–45.

[37] Joachim Neubüser and Said Sidki. "Some computational approaches to groups given by a finite presentation". In: (1988).

[38] B. H. Neumann. “An Essay on Free Products of Groups with Amalgamations”. In: *Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences* 246.919 (1954), pp. 503–554. URL: http://www.jstor.org/stable/91573.

[39] P. Papasoglu. “STRONGLY GEODESICALLY AUTOMATIC GROUPS ARE HYPERBOLIC”. English. In: *Inventiones Mathematicae* 121.2 (1995), pp. 323–334.

[40] Derek JS Robinson. *Finiteness conditions and generalized soluble groups*. Vol. 63. Springer Science & Business Media, 2013.

[41] Math Overflow: Mark Sapir. *The Higman Group*. 2012. URL: https://mathoverflow.net/questions/87347/the-higman-group (visited on 04/23/2019).

[42] The Berstein Seminar. *Exploring the works of Gilbert Baumslag - Residual Finiteness and Hopfian Groups*. 2015. URL: https://berstein2015.wordpress.com/2015/03/02/residual-finiteness-and-hopfian-groups/ (visited on 03/27/2019).

[43] Jean Pierre Serre. *Trees*. eng. 1980.

[44] Charles C. Sims. *Computation with finitely presented groups*. Vol. 48. Cambridge University Press, 1994.

[45] Alessandro Sisto. *Lecture notes on Geometric Group Theory*. 2014. URL: https://people.math.ethz.ch/~alsisto/LectureNotesGGT.pdf (visited on 05/07/2019).

[46] Estelle Souches. “Groupes de Baumslag-Solitar”. In: (1995). Rapport.

[47] Wikipedia, the free encyclopedia, Ipipipourax. *A triangle in a real tree*. (Online; visitied on 05/06/2019). 2011. URL: https://en.wikipedia.org/wiki/Real_tree#/media/File:Y_property.png.

[48] Henry R. Wilton. *392C Geometric Group Theory*. 2009. URL: https://392c.wordpress.com/2009/ (visited on 09/28/2018).

[49] Henry R. Wilton. *Group actions on trees*. 2004. URL: https://www.dpmms.cam.ac.uk/~hjrw2/Talks/trees.pdf (visited on 09/23/2019).

[50] Daniel T. Wise. “The residual finiteness of negatively curved polygons of finite groups”. In: *Inventiones mathematicae* 149.3 (2002), pp. 579–617.