

# A topological proof of the Nielsen-Schreier theorem

Benjamin Debeerst

May 23, 2010

This article is a project work for the course “Topology and Algebraic Topology” in the school year 2009/2010 at the University of Ghent. It treats the topological proof for the Nielsen-Schreier theorem, which states that every subgroup of a free group is free.

For this, some theory about free groups and infinite graph theory is given first in sections 1 and 2. The reader familiar with finite graphs will see known structures in a broader context. The following sections 3, 4 and 5 work out the basics for the topological proof in section 6. For this, the *Cayley graph* of a group and the *fundamental group* of a graph are declared. The reader familiar with topology will recognise the topological arguments, although none of the classic topologic theory is directly used. Topologic notions as *path*, *path equivalence* and *fundamental groups* are newly defined. The results on subgroups and the corresponding graphs lead directly to the proof of Nielsen-Schreier. This short and elegant proof is followed by a sketch of the algebraic proof in section 7, which is quite technical and therefore not given completely. The interested reader can find the details in [Kna06]. Some consequences of the Nielsen-Schreier theorem are given in section 8. The article ends with a justification of the use of topologic terms in section 9. Since all notions are newly defined within this article, the connection to classical topology is not directly clear. A sketch on how the connection can be made is given here. For this section, a basic knowledge of topology is needed, for example the definition of topological spaces and the terms of quotient space, paths, homotopy and covering spaces.

The general scope and way of treating the subject is taken from [Sti80].

## Contents

1	Free Groups	3
2	Basic Notions of Graph Theory	3
3	The Cayley Graph	6
4	The Fundamental Group of a Graph	7
5	Graph Coverings	9
6	Subgroups of Free Groups	13
7	Sketch of the Algebraic Proof	14
8	Consequences	16
9	Connection to Topology	17
	References	19

## 1 Free Groups

**Definition 1.1** (Generators, words, relations). A *generator* is a symbol  $x_i$  with a formal inverse symbol  $x_i^{-1}$ . A collection of generators is called *alphabet*.

A *word* is a finite sequence of some generators or their inverses

$$x_{i_1}^{\varepsilon_1} x_{i_2}^{\varepsilon_2} \cdots x_{i_n}^{\varepsilon_n}, \quad n \in \mathbb{N}_0, \quad \varepsilon_j \in \{-1, 1\}.$$

The product of two words  $w_1$  and  $w_2$  is the concatenation of the corresponding sequences  $w_1 \cdot w_2 := w_1 w_2$ . This product is trivially associative. There is a neutral element for this product, namely the empty word, denoted by 1 (in literature sometimes also  $\varepsilon$ ). We obviously have  $1 \cdot w = w \cdot 1 = w$  for any word  $w$ .

A *relation* is an equation  $r = 1$ , where the word  $r$  is called *relator*. Two words  $w_1$  and  $w_2$  are said to be equivalent with respect to some set of relators  $R$  if  $w_1$  can be converted to  $w_2$  by a sequence of the following operations:

- (i) insertion or deletion of a relator  $r \in R$ ,
- (ii) insertion or deletion of a subword  $xx^{-1}$  or  $x^{-1}x$ .

**Lemma 1.2.** *The equivalence of words is a equivalence relation.*

**Proof.** Reflexivity is trivial. Symmetry follows from the fact that insertion and deletion of subwords are each others inverse operations. Thus, provided a sequence of operations  $s_1, \dots, s_n$  that converts  $w_1$  to  $w_2$ , the sequence  $s_n^{-1}, \dots, s_1^{-1}$  converts  $w_2$  to  $w_1$ . Transitivity is also easy since these operations can be concatenated.  $\square$

The product of words extends in a natural way to a product on the equivalence classes of words via

$$[w_1] \cdot [w_2] := [w_1 \cdot w_2].$$

This is well-defined since  $w_1 w_2$  and  $w'_1 w_2$  are equivalent if  $w_1$  and  $w'_1$  are. Associativity is inherited from the product of words. Thus, the set of equivalence classes of words on a certain alphabet  $A$  without any more than the trivial relations together with the given multiplication form a group. This group is called the *free group* on  $A$ , denoted by  $\langle A \rangle$ .

## 2 Basic Notions of Graph Theory

The graph notion in the context of free groups is somewhat different from the notion the usually seen in (finite) combinatorics, therefore these are completely defined and explained below.

**Definition 2.1** (Graph). A *graph* is a couple  $G = (V, E)$ , where  $V$  is a set of *vertices* and  $E$  a set of *edges*. Each edge  $e_j \in E$  is a set  $\{e_j^{+1}, e_j^{-1}\}$ , containing two *directed edges*, one in positive and one in negative direction. The set of directed edges, the union of all element in  $E$  is denoted as  $\mathcal{E}$ . There are two incidence functions  $\text{initial}(), \text{final}() : \mathcal{E} \rightarrow V$ , which mark an *initial point* and an *end point* of every directed edge, with the property that  $\text{initial}(e_j^{-1}) = \text{final}(e_j^{+1})$  and  $\text{initial}(e_j^{+1}) = \text{final}(e_j^{-1})$ .

Initial and final point are referred to collectively as *endpoints*. Every edge is said to be *incident* with its endpoints and any edge  $e_j$  without a superscript is understood to be directed positively. Also  $e_j^{-1}$  and  $e_j^{+1}$  are called each others inverses, with  $(e_j^{-1})^{-1} := e_j^{+1}$ .

A *subgraph* of a graph  $G = (V, E)$  is a graph  $G' = (V', E')$ , with  $V' \subset V$ ,  $E' \subset E$  and where the incidence functions the ones inherited from  $G$ .

One could criticise this definition of a graph with the sets of directed edges as undirected edges as more complex then necessary, but this definition makes it easy to switch between the directed and undirected version of the same graph without losing formal accuracy.

Notice that the sets of vertices and edges are often written as  $\{e_1, e_2, \dots\}$  or  $\{v_1, v_2, \dots\}$  although they are not assumed to be countable sets (but they can be). This style is chosen for easy reading and referring and to avoid the need of some index set every time.

**Definition 2.2** (Path, path reducing, path equivalence). A path  $p$  of length  $r$  in a graph  $G$  a finite sequence  $d_1, \dots, d_r$  of directed edges with the property

$$\text{initial}(d_i) = \text{final}(d_{i-1}), \quad \forall i \in \{2, \dots, r-1\}.$$

One says that  $p$  is a path from  $\text{initial}(d_1)$  to  $\text{final}(d_r)$ . It is called *closed* if  $\text{initial}(d_1) = \text{final}(d_r)$ . One writes  $p = d_1, \dots, d_r$  and also  $p_3 = pp'$  if  $p_3 = d_1, \dots, d_r, d'_1, \dots, d'_l$  and  $p' = d'_1, \dots, d'_l$ .

A path  $p$  is called *reduced* if no two successive directed edges are each others inverses. Such a subpath  $e_j^{-1}e_j^{+1}$  or  $e_j^{+1}e_j^{-1}$  is called *spur*. A single vertex is also admitted as a path (which is trivially reduced). The *inverse* of the path  $p = d_1, \dots, d_r$  is the path  $p^{-1} = d_r^{-1}, \dots, d_1^{-1}$ .

Two paths  $p_1$  and  $p_2$  are called *equivalent* if one can be derived from the other by a finite number of insertions or removals of spurs between successive directed edges, at the beginning or at the end.

**Lemma 2.3.** *Path equivalency is an equivalence relation.*

**Proof.** There is nothing to show for reflexivity. Since insertion and deletion of a spur at a certain point are each others inverse operations, symmetry follows directly. Transitivity is also easy since we deal with finite sequences of operations which can be performed successively.  $\square$

The lemma and its proof are not surprising in any sense, but its consequences are interesting: reducing a path  $p$  iteratively by removing stepwise an arbitrary spur, one will get to an end, i.e. a reduced path which contains no spurs. Transitivity and symmetry guarantee that this reduced path is unique. Therefore, it makes sense to talk about *the* reduced form of  $p$ , which is a canonical representative for the equivalence class of  $p$ .

**Definition 2.4** (Connected graph, tree). A graph  $G = (V, E)$  is called *connected* if for any two vertices  $x, y \in V$  there is a path from  $x$  to  $y$ . A *tree* is a connected graph with no other reduced closed paths than single vertices.

One can define trees in a number of different ways from which two are stated in the following lemma:

**Lemma 2.5.** *The following are equivalent for a graph  $G$ :*

- (i)  $G$  is a tree.
- (ii) Any two vertices of  $G$  are connected by a unique reduced path.
- (iii)  $G$  is a tree and any two paths between two fixed vertices in  $G$  are equivalent.

**Proof.** (i) $\Rightarrow$ (ii): Assume there are vertices  $x, y$  with at least two different reduced paths from  $x$  to  $y$ ,  $p_1 = a_1, \dots, a_r$  and  $p_2 = b_1, \dots, b_l$ . One can assume one of these paths to be of minimal length, for example  $p_1$ , and that there is no other pair of vertices with a non-unique reduced path of smaller length. The path  $p_1 p_2^{-1}$  is closed, and therefore cannot be reduced since the only closed reduced paths in a tree are single vertices. But  $p_1$  and  $p_2^{-1}$  are themselves reduced, thus the subpath  $a_r b_l$  is a spur. Removing this spur leads to a pair of vertices with a non-unique path of shorter length than before, a contradiction.

(ii) $\Rightarrow$ (iii): Any path  $p$  is equivalent to its reduced form. Thus, if two paths have the same reduced form, they are equivalent.

(iii) $\Rightarrow$ (i): A reduced closed path  $p = d_1, \dots, d_r$  different from a single vertex cannot be equivalent to the single vertex  $x := \text{initial}(d_1)$  as a path since both are reduced. Thus, such a path does not exist.  $\square$

In arbitrary graphs reduced paths between two fixed vertices need not to be unique. Therefore, it sometimes makes sense to reduce the view to a subgraph where this uniqueness holds. This leads to the concept of a spanning tree:

**Definition 2.6.** A *spanning tree* of a graph  $G$  is a subgraph that is a tree with the same vertex set as  $G$ .

**Lemma 2.7.** *Every connected graph contains a spanning tree.*

**Proof.** The proof is easy for finite graphs since one can construct a subtree starting with a single edge iteratively. One can expand this to the countable case. More interesting is the uncountable part. The proof given here is taken from [Ser80].

The set of subgraphs of  $G$  that are trees is a directed set and therefore Zorns lemma can be used: there is a maximal subtree  $T$  in  $G$ . Assume  $T$  would not contain all vertices, then there is an edge  $e$  in  $G$  with initial point in  $T$  and final point outside  $T$  because  $G$  is connected. Adding this edge and its final point to  $T$  gives another tree  $T'$  where  $T$  is an unequal larger subtree. This is a contradiction to the assumption that  $T$  is maximal.  $\square$

### 3 The Cayley Graph

**Definition 3.1.** Given a free group  $F$  on the generator set  $A$ , the *Cayley graph*  $G$  is defined as  $CG(F) := (V, E)$  where the vertex set is  $V := G$  and the edge set is

$$E := \{[g, h]_x = \{[g, h]_x, [h, g]_{x^{-1}}\} \mid g, h \in F, x \in A, g \cdot x = h\}.$$

The incidence relation here is  $\text{initial}([g, h]_x) = g$ ,  $\text{final}([g, h]_x) = h$  and  $([g, h]_x)^{-1} = ([h, g]_{x^{-1}})$ .

Since  $F$  is a group, every element  $x$  has an inverse  $x^{-1}$ . Thus, the defined structure indeed fits the definition 2.1.

**Lemma 3.2.** *Suppose  $F = \langle a_1, \dots, a_n \rangle$  a free group. Then  $CG(F)$  is a tree.*

**Proof.** Assume there is a nontrivial reduced closed path starting and ending in  $x$ . Thus, there is a series of multiplications indicated by this path such that  $x = x \cdot x_1 \cdots x_r$ . This implies  $\varepsilon = x_1 \cdots x_r$ . The path was nontrivial, hence  $r \neq 0$ . Since there are no relations between the generators besides the trivial ones, this product contains two successive factors  $x_j, x_{j+1}$  with  $x_j^{-1} = x_{j+1}$ . But these factors are represented in the path by two successive edges  $[u, v]_{x_j}, [v, w]_{x_{j+1}}$ , and from  $x_j^{-1} = x_{j+1}$  we have  $w = u$ . Thus, these edges are the two directions of the same edge. This is a contradiction since the path was assumed to be reduced.  $\square$

From this proof one can see that paths in the Cayley graph of a free group  $F$  emanating from the vertex  $\varepsilon$  correspond to elements in  $F$  in a natural, bijective way: there is a unique reduced path to every vertex, thus a unique reduced word corresponding to this vertex and thus representing this vertex as element in  $F$ . Conversely, given a reduced word in  $F$ , the series of generators or their inverses indicates a unique reduced path to a certain vertex, which again corresponds to the given word. In particular, spurs correspond to the relations  $aa^{-1}$  or  $a^{-1}a$ , thus the notion of reduced words in a free group and reduced paths in a graph coincide.

## 4 The Fundamental Group of a Graph

**Definition 4.1.** Let  $G$  be a graph and  $v$  a fixed vertex in  $G$ . The *fundamental group* of  $G$  in  $v$  is defined as

$$\pi_1(G, v) := \{[p] \mid p \text{ is a closed path starting in } v\},$$

where  $[p]$  is the equivalence class of the path  $p$ . The multiplication on  $\pi_1(G, v)$  is defined as

$$[p_1] \cdot [p_2] := [p_1 \cdot p_2].$$

**Lemma 4.2.** *The fundamental group of a graph  $G$  is a group. When  $G$  is connected it is independent from the choice of  $v$ .*

**Proof.** As for the product of equivalence classes of words, the concatenation of paths is independent from insertion or deletion of spurs, thus the product of equivalence classes of paths is well-defined. The multiplication is obviously associative and we have (just as for words)  $[p]^{-1} = [p^{-1}]$ . Also obvious is the identity element, namely the equivalence class of the trivial path  $v$ .

Now assume  $G$  is connected. Fix another vertex  $w$  in  $G$ , then there is a path  $p$  from  $v$  to  $w$ . Now define the map  $\varphi : \pi_1(G, v) \rightarrow \pi_1(G, w)$ ,  $[q] \mapsto [p^{-1}][q][p]$ . We will show that  $\varphi$  is an isomorphism. We have a path  $p$  from  $v$  to  $w$ ,  $p^{-1}$  is the same path in the opposite direction, thus the concatenation  $p^{-1}qp$  is well-defined and a closed path on  $w$ , hence  $\varphi$  is well-defined. It also is an homomorphism since

$$\begin{aligned} \varphi([q_1][q_2]) &= [p^{-1}][q_1][q_2][p] \\ &= [p^{-1}][q_1][p][p^{-1}][q_2][p] \\ &= \varphi([q_1]) \cdot \varphi([q_2]). \end{aligned}$$

The map  $\varphi$  is surjective, because for every closed path  $q$  on  $w$ ,  $\varphi([p][q][p^{-1}]) = [q]$ , and injective since from  $[p^{-1}][q_1][p] = [p^{-1}][q_2][p]$ ,  $[q_1] = [q_2]$  immediately follows.  $\square$

For the fundamental group of connected graphs the notation  $v$  is often omitted since it makes no difference for the group structure.

The following theorem gives the fundamental connection between graphs and free groups. Since the Nielsen-Schreier theorem 6.3 is proved by the construction of certain graphs, it is in fact the reason why the topological proof is possible at all.

**Theorem 4.3.** *The fundamental group of a connected graph  $G$  is free.*

**Proof.** Fix a vertex  $v$ . We will construct a canonical representative for the equivalence class of any path. Consider a spanning tree  $T$  of  $G$ . Then every vertex  $w_i$  has a unique reduced *approach path*  $p_i$  in  $T$ . For each edge  $e_i$  with

initial( $e_i$ ) =  $w_j$  and final( $e_i$ ) =  $w_k$ , define the path  $a_i := p_j e_i p_k^{-1}$  in  $G$  which is a closed path on  $v$ . Now any closed path on  $v$  has a canonical equivalent by the row of  $a_i$ s corresponding to the row of edges. This is, because the approach paths of succeeding edges cancel and the approach path  $w_i$  for  $p_i = v$  is the path containing just the single vertex  $v$ , thus 1 in the fundamental group.

Furthermore, any path  $a_i$  corresponding to an edge  $e_i$  in  $T$  is a closed path in the tree  $T$ , thus equivalent to 1. All these paths  $a_i$  are therefore unnecessary: every closed path in  $G$  on  $v$  can equivalently be written as product of  $a_i$ s, where  $e_i$  is not in  $T$ . In other words, the set  $A = \{[a_i] \mid e_j \text{ is not in } T\}$  generates the fundamental group of  $G$ .

If it can be shown that these generators are free, i.e. that no relations between the generators but the trivial ones hold, the theorem is proved. Assume there is some nontrivial reduced relation, that is  $[a_1] \cdots [a_n] = 1$ , where no subsequent factors are each others inverse. We have:

$$\begin{aligned} 1 &\sim a_1 \cdots a_n \\ &= p_{i_1} e_1 p_{i_2}^{-1} p_{i_2} \cdots p_{i_n} e_n p_{i_{n+1}}^{-1} \\ &\sim p_{i_1} e_1 \cdots e_n p_{i_{n+1}}^{-1} \end{aligned}$$

By construction,  $p_{i_1}$  and  $p_{i_{n+1}}^{-1}$  are paths in  $T$  while the edges  $e_i$  are not in  $T$ . Thus, there cannot be any possible cancellation between these paths and the edges  $e_i$ . The subpath  $e_1, \dots, e_n$  must therefore contain spurs, i.e. two successive  $e_j$ s are each others inverses. But then, the corresponding  $a_j$ s would be each others inverses, a contradiction to the assumption that the product is reduced. This shows that there are no relations among the  $[a_j]$ s but the trivial ones, which proves the theorem.  $\square$

It is important to notice that the maps  $\pi_1(\cdot)$  and  $CG(\cdot)$  are *not* each others inverses. This is easy to see: the fundamental group  $G$  of a connected graph  $T$  is free, thus the Cayley graph of  $G$  is a tree by lemma 3.2. But if  $T$  is no tree, this Cayley graph and  $T$  cannot be the same graph.<sup>1</sup>

The good news is, given a free group  $F$ , a graph with fundamental group isomorphic to  $F$  is easy to construct.

**Lemma 4.4.** *Assume  $F$  is a free group on the generators  $a_i$ . Define the bouquet of circles as the graph  $B = (V, E)$ , where  $V = \{v\}$  and  $E = \{a_1, a_2, \dots\}$ . All edges in  $E$  have initial and final point  $v$ . Then  $\pi_1(B, v)$  is isomorphic to  $F$ .*

**Proof.** Every word in  $F$  can be read as path in  $B$  and conversely. A word is reduced in  $F$  if and only if the corresponding path is reduced. But since there is only one vertex in  $B$ , every path in  $B$  is a closed path on  $v$ .  $\square$

---

<sup>1</sup>To be more exactly: they cannot be isomorphic. But since graph isomorphism is not declared, I avoid this notion here.

## 5 Graph Coverings

**Definition 5.1** (Covering Graph). Consider two graphs  $\tilde{G} = (\tilde{V}, \tilde{E})$ ,  $G = (V, E)$  with  $\tilde{V} \neq \emptyset$  and a map  $\varphi : \tilde{V} \cup \tilde{E} \rightarrow V \cup E$ , where the sets of edges and vertices are assumed to be disjoint. Suppose the following properties hold:

- (i)  $\varphi$  preserves endpoints, i.e. for every directed edge  $\tilde{e} \in \tilde{E}$  we have  $\varphi(\text{initial}(\tilde{e})) = \text{initial}(\varphi(\tilde{e}))$  and  $\varphi(\text{final}(\tilde{e})) = \text{final}(\varphi(\tilde{e}))$ ,
- (ii)  $\varphi(\tilde{e}^{-1}) = \varphi(\tilde{e})^{-1}$ ,
- (iii) If  $\tilde{v}$  is a vertex in  $\tilde{G}$  and  $\{e_1, e_2, \dots\}$  is the set of directed edges in  $G$  with initial point  $\varphi(\tilde{v})$ , then  $\varphi$  maps the set  $\{\tilde{e}_1, \tilde{e}_2, \dots\}$  of edges in  $\tilde{G}$  with initial point  $\tilde{v}$  bijective on  $\{e_1, e_2, \dots\}$ .

Then  $\tilde{G}$  is said to *cover*  $G$  where  $\varphi$  is called the *covering map* or *projection*.

The third condition can be understood as a ‘local homeomorphism’ condition, because it says that ‘neighbourhoods’ of corresponding edges look alike. Altogether, these conditions give a strict bound on the structure of  $\tilde{G}$  which is stated in lemma 5.3. The notation of the covering graph  $\tilde{G}$  and the covered graph  $G$  will be used throughout the following sections. An element in  $\tilde{G}$  mapped to the element  $a$  in  $G$  will always be denoted as  $\tilde{a}$ . Different elements in the inverse image of  $a$  will be distinguished by upper indexes  $\tilde{a}^{(1)}, \tilde{a}^{(2)}, \dots$  if necessary.

**Example 5.2.** Consider the graph  $G$  with one single vertex and two edges  $e_1, e_2$  and the covering graph  $\tilde{G}$  as in figure 1. One can see that the map  $\varphi : \tilde{G} \rightarrow G$  can be fully described by a certain labelling of the graph  $\tilde{G}$  given some labelling of  $G$ . In the case of figure 1 this is indeed a covering map: since all vertexes are mapped to one single vertex, condition (i) holds trivially. Condition (ii) is complied since the edges are only given in one direction implying the inverse edge with convenient properties. Condition (iii) holds since every vertex in  $\tilde{G}$  is incident with four directed edges,  $e_1$  and  $e_2$  each outgoing and incoming. The same is true for the single vertex in  $G$ . Thus the local homeomorphism condition holds which can also be seen quite literally in figure 2.<sup>2</sup>

**Lemma 5.3.** *Given a graph  $\tilde{G}$  covering the connected graph  $G$  the following properties hold:*

- (i) *Given a path  $p = d_1, \dots, d_n$  in  $G$ , there is a unique path in  $\tilde{G}$  which is mapped on  $p$  by  $\varphi$  for every vertex  $\tilde{v}$  in  $\varphi^{-1}(\text{initial}(d_1))$ .*
- (ii) *Every path in  $\tilde{G}$  corresponding to a spur in  $G$  is a spur in  $\tilde{G}$  itself.*
- (iii) *The cardinality of  $\varphi^{-1}(\{v\})$ , called the sheet number, is a constant, i.e. independent from the choice of  $v$ .*

<sup>2</sup>Notice that, different from what the second figure might suggest, the labelling of the edges is of course important in this condition.

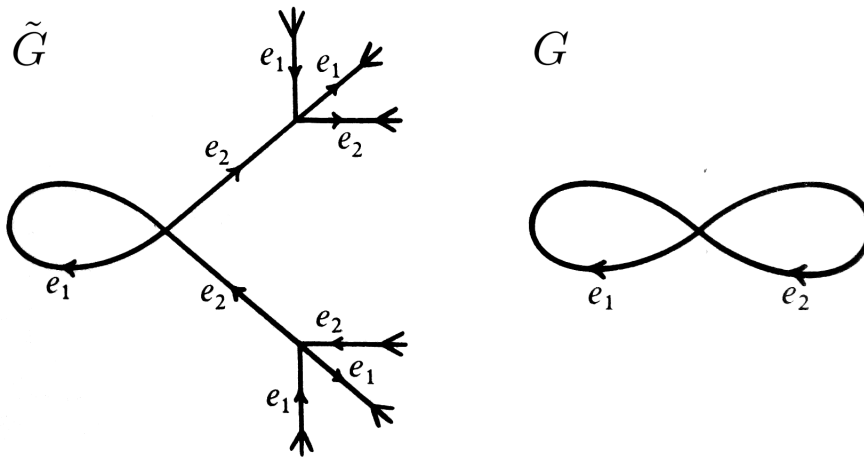


Figure 1: An example for a graph covering.

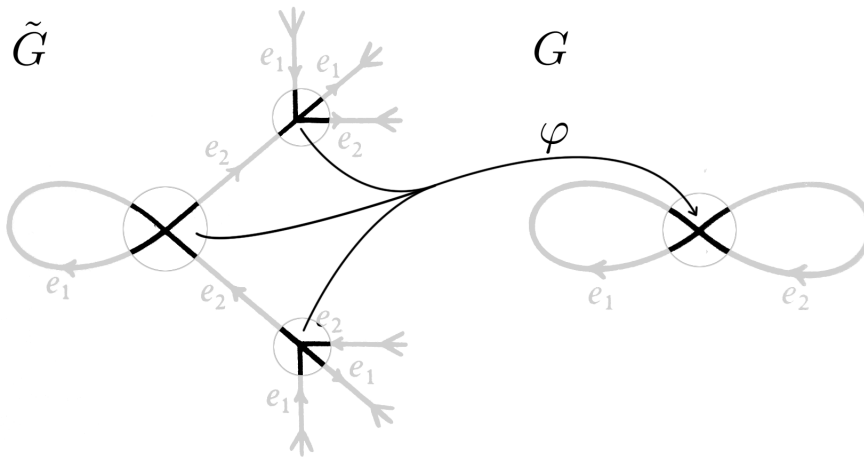


Figure 2: Neighbourhoods of corresponding vertices look alike.

**Proof.** To proof the lemma, one first has to see that  $\varphi$  is surjective. Assume it is not, then there is a vertex  $v$  or an edge  $e$  not in the image of  $\varphi$ .

The case of an edge can be easily reduced to the vertex case. Consider  $v = \text{initial}(e)$ . If  $v$  is in the image of  $\varphi$  while  $e$  isn't, this contradicts condition (iii) of definition 5.1. Thus  $v$  is not in the image of  $\varphi$ .

The case of a vertex works as follows. Consider  $v$  not in the image of  $\varphi$ . Since  $\tilde{V}$  is not empty, there is a vertex  $\tilde{w}$  in  $\tilde{G}$  with  $\varphi(\tilde{w}) = w$ .  $G$  is connected, so there is a path  $e_1, \dots, e_n$  from  $w$  to  $v$ . Condition (iii) of the definition gives an edge  $\tilde{e}_1$  that is mapped on  $e_1$  and condition (i) guarantees that the initial and final points are mapped on each other respectively. Doing the same with the  $\text{final}(e_1)$  and iterating this process leads to the corresponding edges  $\tilde{e}_1, \dots, \tilde{e}_n$  with  $\varphi(\tilde{e}_i) = e_i$  for every  $i$ . Endpoints are preserved, therefore we have  $\varphi(\text{final}(\tilde{e}_n)) = v$ , a contradiction. Thus  $\varphi$  is surjective.

- (i) The above construction can be repeated for every element in the inverse image of  $\text{initial}(d_1)$ . Any of these paths is unique, because of condition (iii) in the definition: for an edge in  $G$  and a vertex in  $\tilde{v}$  there can only be exactly one corresponding edge in  $\tilde{G}$  with initial point  $\tilde{v}$  since  $\varphi$  gives a bijective relation.
- (ii) Assume a path  $p = e_1, e_2$  is a spur. That means  $e_2 = e_1^{-1}$ . For every edge  $\tilde{e}_1$ , the completion of the corresponding path is unique and conditions (i) and (ii) in the definition make this path a spur as well.
- (iii) Given two vertices  $v$  and  $w$  in  $G$ , there is a path  $p$  from  $v$  to  $w$  since  $G$  is connected. Every vertex  $\tilde{v}^{(i)}$  in the inverse image of  $v$  corresponds to a unique path  $\tilde{p}^{(i)}$  with endpoint  $\tilde{w}^{(i)}$ . This induces a bijective map between the inverse images of  $v$  and  $w$ .

This proves the lemma. □

This properties immediately lead to the following:

**Theorem 5.4.** *For a covering  $\tilde{G}$  of a connected graph  $G$ ,  $\pi_1(\tilde{G})$  is isomorphic to a subgroup of  $\pi_1(G)$ .*

**Proof.** Item (ii) of the above lemma implies that equivalent paths in  $G$  are covered by equivalent paths in  $\tilde{G}$  and vice versa. Elements in  $\pi_1(G)$  are equivalence classes of loops  $p$  on a certain vertex  $v$ . But item (i) then tells us, once having fixed an element  $\tilde{v}$  that is mapped to  $v$ , that the corresponding path in  $\tilde{G}$  is unique. Thus, there is a bijective map  $\varphi_*$  between the elements of  $\pi_1(G)$  and the equivalence classes  $[\tilde{p}]$  of paths  $\tilde{p}$  that emanate from some fixed vertex  $\tilde{v}$ . These paths aren't necessarily closed. But one can restrict the map to closed paths in  $\tilde{G}$  on  $\tilde{v}$ , which gives an injective map

$$\varphi_* : \pi_1(\tilde{G}, \tilde{v}) \longrightarrow \pi_1(G, v).$$

Properties (i) and (ii) of the definition of coverings imply that products are sent to products and inverses to inverses, thus  $\varphi_*$  in fact is a monomorphism, which proves the theorem.  $\square$

Since  $\varphi_*$  is injective, it is convenient not to distinguish between  $\pi_1(\tilde{G}, \tilde{v})$  and the image  $\varphi_*(\pi_1(\tilde{G}, \tilde{v}))$ .

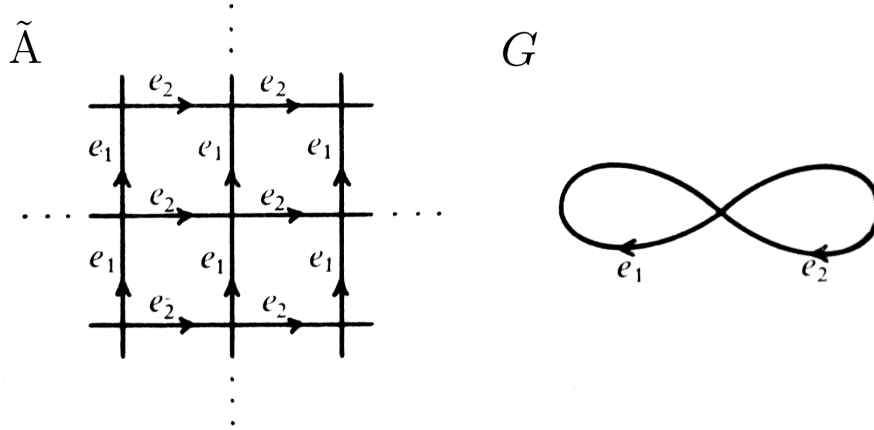


Figure 3: The *universal abelian cover* for  $G$

**Example 5.5.** Reconsidering example 5.2, we want to find out which groups are represented by  $G$  and  $\tilde{G}$ . The fundamental group of  $G$  is generated by exactly two free generators  $e_1$  and  $e_2$  since the unique vertex is a spanning tree. Thus we have  $\pi_1(G) = \langle e_1, e_2 \rangle$ .

For the fundamental group of  $\tilde{G}$ , fix the unique vertex where the outgoing edge mapped to  $e_1$  is a loop (see figure 1). Since the rest of the graph is a tree, the only possible reduced closed paths in  $\tilde{G}$  are paths of the form  $e_1^n$ ,  $n \in \mathbb{Z}$ . Thus the fundamental group of  $\tilde{G}$  is the cyclic group generated by  $e_1$ , which indeed is a subgroup of  $\langle e_1, e_2 \rangle$ .

As a more complex example consider the covering graph  $\tilde{A}$  given in figure 3. Fixing a vertex in  $\tilde{A}$ , a path emanating from this vertex is closed if and only if the sums of the exponents of  $e_1$  and  $e_2$  each equal zero. As a word in the fundamental group of  $\tilde{A}$ , each such path would equal the neutral element if the generators  $e_1$  and  $e_2$  would commute, which is expressed by the relator  $e_1 e_2 e_1^{-1} e_2^{-1}$ . Thus the quotient group of  $\pi_1(\tilde{A})$  modulo the *commutator subgroup*  $\langle e_1 e_2 e_1^{-1} e_2^{-1} \rangle$  is trivial which is equivalent to  $\pi_1(\tilde{A}) \cong \langle e_1 e_2 e_1^{-1} e_2^{-1} \rangle$ . This cover with the commutator subgroup as fundamental group is called the *universal abelian cover* for  $G$ .

The subgroup induced by the covering  $\tilde{G}$  is the fundamental group of  $\tilde{G}$  and hence free. If it is possible to construct a covering graph such that its fun-

damental group isomorphic a given subgroup, the Nielsen-Schreier theorem would be proved.

## 6 Subgroups of Free Groups

In the construction of  $\varphi_*$  above, an element from the inverse image of  $v$  is chosen. It is important to understand, how the other possible choices correspond to the right cosets of  $\pi_1(G, v)$  modulo  $\pi_1(\tilde{G}, \tilde{v}^{(0)})$ . Closed paths in  $G$  which lift to non-closed paths in  $\tilde{G}$  end in some vertex  $\tilde{v}^{(i)} \neq \tilde{v}$ , and thus can be classified by their endpoints in  $\tilde{G}$ . But this is in fact a right coset decomposition of  $\pi_1(G)$  modulo  $\pi_1(\tilde{G})$ .

**Lemma 6.1.** *Consider the following equivalence class on  $\pi_1(G, v)$ :*

$$[p] \sim [p'] :\Leftrightarrow \tilde{p} \text{ and } \tilde{p}' \text{ end in the same point } \tilde{v}^{(j)}$$

*The paths  $\tilde{p}$  and  $\tilde{p}'$  are understood to emanate from a fixed point  $\tilde{v}^{(0)}$ . There is a bijective map between the right coset decomposition of  $\pi_1(G, v)$  modulo  $\pi_1(\tilde{G}, \tilde{v}^{(0)})$  and  $\pi_1(G, v) / \sim$ .*

*As a consequence, the sheet number  $\tilde{G}$  covering  $G$  is equal to the index of  $\pi_1(\tilde{G}, \tilde{v}^{(0)})$  in  $\pi_1(G, v)$ .*

**Proof.** Two elements  $[p], [p'] \in \pi_1(G, v)$  are in the same right coset if and only if  $[g][p] = [p']$  for some  $[g] \in \pi_1(\tilde{G}, \tilde{v}^{(0)})$ . This implies  $[\tilde{g}][\tilde{p}] = [\tilde{p}']$ , where  $\tilde{g}$  is a path from  $\tilde{v}^{(0)}$  to  $\tilde{v}^{(i)}$ , thus the paths  $[\tilde{p}] = [\tilde{p}']$  must have the same initial point  $\tilde{v}^{(0)}$  and end point  $\tilde{v}^{(i)}$ . By definition, this means  $[p] \sim [p']$ .

Now consider two equivalent elements  $[p], [p'] \in \pi_1(G, v)$ . These lift to paths  $\tilde{p}, \tilde{p}'$  in  $\tilde{G}$  from  $\tilde{v}^{(0)}$  to  $\tilde{v}^{(i)}$ . Since  $\varphi(\tilde{v}^{(0)}) = \varphi(\tilde{v}^{(i)}) = v$ , the projected paths  $p$  and  $p'$  are loops on  $v$ , thus elements of  $\pi_1(G, v)$ . The path  $\tilde{p}'\tilde{p}^{-1}$  is closed in  $\tilde{G}$ , thus element of  $\pi_1(\tilde{G}, \tilde{v}^{(0)})$ . But this implies  $[p'] [p]^{-1} \in \varphi(\pi_1(\tilde{G}, \tilde{v}^{(0)}))$ , thus  $[p]$  and  $[p']$  are in the same right coset of  $\pi_1(G, v)$  modulo  $\pi_1(\tilde{G}, \tilde{v}^{(0)})$ .  $\square$

We're now ready to proof the basic result from which the Nielsen-Schreier theorem directly will follow.

**Lemma 6.2.** *Given a free group  $F$ ,  $F$  can be realized as the fundamental group of the bouquet of circles  $G$ . For every subgroup  $H$  of  $F$ , there is a graph  $\tilde{G}$  that covers  $G$  with fundamental group isomorphic to  $H$ .*

**Proof.** We will construct such a graph  $\tilde{G} = (\tilde{V}, \tilde{E})$  covering  $G$  as follows.

According to the previous lemma, there has to be a vertex in  $\tilde{V}$  that is mapped to the single vertex  $v$  in  $G$  for every right coset of  $H$  in  $F$ . If  $\{p_0, p_1, \dots\}$  is a set of representatives for these cosets, then define  $\tilde{V} := \{\tilde{v}_0, \tilde{v}_1, \dots\}$  as the vertex set of  $\tilde{G}$ , where every vertex  $v_i$  corresponds to the coset  $H[p_i]$ .

As  $\tilde{G}$  is intended to be a covering, it is subject to the ‘local homeomorphism’ condition 5.1(iii). Since  $G$  has just one vertex, every edge of  $G$  starts in  $v$ . Consider the set of directed edges

$$\tilde{\mathcal{E}} := \{\tilde{e}_i^{(j)} \mid e_i \in \mathcal{E}, \tilde{v}_j \in \tilde{V}\},$$

where each of the edges  $\tilde{e}_i^{(j)}$  has initial point  $\tilde{v}_j$ . But from this, these edges are fully determined: the final point of such an edge  $\tilde{e}_i^{(j)}$  must be the vertex corresponding to the coset  $H[p_j e_i]$ , because of the previous lemma.

The fundamental group of the constructed graph is isomorphic to  $H$ . A path  $\tilde{p}$  in  $\tilde{G}$  emanating from  $\tilde{v}^{(0)}$  and covering the path  $p$  ends at some vertex  $\tilde{v}^{(j)}$  corresponding to the the right coset  $H[p]$ . This is,  $\tilde{p}$  is closed if and only if  $H[p] = H$  which is equivalent to  $[p] \in H$ . That means  $\pi_1(\tilde{G})$  and  $H$  correspond bijectively, while the bijection is in fact a isomorphism according to the proof of theorem 5.4.  $\square$

Our goal is an immediate consequence.

**Theorem 6.3** (Nielsen-Schreier). *Any subgroup of a free group is free.*

**Proof.** Given a free group  $F$  and a subgroup  $H$ , one can realize  $F$  as the fundamental group of a bouquet of circles. A subgroup can be realized by the fundamental group of some covering graph since lemma 6.2, and therefore is free.  $\square$

## 7 Sketch of the Algebraic Proof

To proof the Nielsen-Schreier theorem algebraically, we will need the following results from combinatorial group theory. The proofs can be found in [Kna06].

For the whole section let  $F$  be a free group on the generator set  $A$ , define  $A' := A \cup A^{-1}$ , let  $H$  be a subgroup of  $F$  and  $C$  a set of representatives for the cosets of  $H$  in  $F$ . Furthermore, define the map  $\rho : F \rightarrow C$ , where  $g \in F$  is mapped to the coset representative of  $Hg$ . This function has the property that  $\rho(hg) = \rho(g)$  for every  $h \in H$  since  $Hhg = Hg$ .

**Lemma 7.1.** *There is a set  $C$  of representatives for the cosets  $Hg \subset F$  where every  $g \in C$  is a reduced word  $g = b_1 b_2 \cdots b_n$  in  $F$  with the property that also  $b_1 b_2 \cdots b_{n-1} \in C$ . Such a set of representatives is called a Schreier set.*

Notice that from the conditions it follows that the empty word which is the identity in  $F$  is an element of  $C$  and therefore the representative for  $H$  as a coset.

**Lemma 7.2.** *The elements of the form  $ga\rho(ga)^{-1}$  with  $g \in C$  and  $a \in A$  form a set of generators for  $H$ .*

*Furthermore the element  $g' = \rho(ga)$  has the properties that  $g = \rho(g'a^{-1})$  and  $ga^{-1}\rho(ga^{-1})^{-1} \sim (g'a\rho(g'a)^{-1})^{-1}$ .*

**Algebraic proof of the Nielsen-Schreier theorem.** We will show that the nontrivial generators from lemma 7.2 are free.

Consider an element  $u = ga\rho(ga)^{-1}$ . This element is already reduced or equivalent to 1, which can be seen as follows. Define  $g' := \rho(ga)$ . Since  $g$  and  $g'$  are elements of  $C$  they are reduced, thus the only possible place for a reduction in  $u$  is at  $a$ . That is, the last factor of  $g$  is  $a^{-1}$  or the last factor of  $g'$  is  $a$ . If the last factor of  $g$  is  $a^{-1}$ , then  $ga$  is equivalent to a prefix of  $g$  which must be contained in the Schreier set, thus  $\rho(ga) \sim ga$  and therefore  $u = ga\rho(ga)^{-1} \sim 1$ . The same works for the other case: if the last factor of  $g'$  is  $a$ , then  $g'a^{-1}$  is equivalent to a prefix of  $g'$  and therefore  $\rho(g'a^{-1}) \sim g'a^{-1}$ . Lemma 7.2 gives  $u^{-1} = (ga\rho(ga)^{-1})^{-1} \sim g'a^{-1}\rho(g'a^{-1})^{-1} \sim 1$ .

As a consequence, elements of the form  $u = ga\rho(ga)^{-1}$  and  $v = fb\rho(fb)^{-1}$  where  $f, g \in C$  and  $a, b \in A$ , are different if and only if they are not equal to one and  $(g, a) \neq (f, b)$ . Assuming  $u$  and  $v$  were equal, the words  $g$  and  $f$  must have the same length. If not, one of them is shorter, for example  $g$ . But this means  $ga$  is a prefix of  $f$  and therefore in  $C$  since  $C$  is a Schreier set. But this means  $\rho(ga) = ga$ , from which  $u = 1$  would follow. Thus they do have the same length, that means  $u$  and  $v$  are the same reduced word, thus symbolically the identical, from which follows  $(g, a) = (f, b)$ .

It remains to show that a product  $u_1 \cdots u_n$  is not equivalent to 1, when the elements  $u_i$  are of the form as  $u$  and  $v$  above and  $u_i \neq u_{i+1}^{-1}$  for  $i \in \{1, \dots, n-1\}$ . We will show this first for products of length two.

Assume  $u$  and  $v$  have the form as above,  $u^{-1} \neq v$  and both are not equivalent to 1. The element  $u$  is reduced and therefore determines its (reduced) subwords  $g$  and  $a$ , the same is true for  $v$ . Thus it makes sense to refer to  $a$  and  $b$  as the *significant factors* of  $u$  and  $v$  respectively. Define  $g' := \rho(ga)$  and  $f' := \rho(fb)$ . To show the non-equivalency of the product  $u \cdot v = gag'^{-1} \cdot fbf'^{-1}$  to 1 we will show that the significant factors cannot vanish due to cancellation. Assume they do, then we have one of the following cases:

- (i)  $a$  cancels because the last factor of  $g'$  is  $a$ . Then  $g'a^{-1}$  is equivalent to a prefix of  $g'$ , thus  $g'b^{-1} = \rho(g'b^{-1})$ , from which  $u \sim 1$  follows,
- (ii)  $b$  cancels because the last factor of  $f$  is  $b^{-1}$ . As in (i), then  $fb$  is equivalent to a prefix of  $f$  and it follows  $v \sim 1$ ,
- (iii)  $a$  and  $b$  cancel because  $g'^{-1}f \sim 1$  and  $ab \sim 1$ . Then  $f = g'$  and  $b = a^{-1}$ , which with lemma 7.2 implies  $u = v^{-1}$ .

All these cases are contradicted by the assumptions, thus the significant factors do not cancel, which means that the product cannot be equivalent to 1. This automatically expands to finite products  $u_1 \cdots u_n$ , because this is true

for every pair of successive factors. This shows that the nontrivial generators for  $H$  are free, and therefore form a free generating set for  $H$ .  $\square$

## 8 Consequences

From the previous sections there are some easy consequences. Consider the free group  $F$  to be generated by a finite set of generators  $A = \{a_1, \dots, a_r\}$ . The number  $r$  is called the *rank* of the free group  $F$ . The following result is referred to as the *Schreier index formula*.

**Lemma 8.1.** *Suppose  $F$  has rank  $r$  and the subgroup  $H \leq G$  has index  $i$  in  $G$ . Then the rank  $s$  of  $H$  is given by*

$$s = i \cdot r - i + 1.$$

**Proof.** As  $F$  has rank  $r$  it is realized as the fundamental group of a bouquet of circles  $G$  with  $r$  edges, one for each generator of  $F$ . The subgroup  $H$  is realized by the fundamental group of the graph  $\tilde{G}$  which is a covering for  $G$ . There are exactly  $i$  vertexes in  $\tilde{G}$  since  $i$  this is the sheet number of  $\tilde{G}$  and  $G$  has just one unique vertex. Accordingly,  $\tilde{G}$  has exactly  $i \cdot r$  edges due to construction in the proof of lemma 6.2. A tree with  $i$  vertexes has exactly  $i - 1$  edges, thus for every spanning tree  $T$  of  $\tilde{G}$  there are  $i \cdot r - (i - 1) = i \cdot r - i + 1$  edges in  $\tilde{G}$  that are not in  $T$ . But this is the number of generators for  $H$ , according to the proof of theorem 4.3.  $\square$

Another form in which the formula is often written is

$$i = \frac{(s - 1)}{r - 1}$$

since it reveals the boundary of this formula on the possible ranks of subgroups of free groups with finite index:  $r - 1$  must be divisor of  $s - 1$ . This says of course nothing about subgroups with infinite index, and counterexamples for this boundary with subgroups of infinite index are easily found. For example contains the group generated by three elements  $F = \langle a, b, c \rangle$  the subgroup of infinite index generated by two elements  $H = \langle a, b \rangle$ , while 2 is no divisor of 1. Thus, the contraposition of the lemma can also useful: a subgroup of rank  $s$  where  $r - 1$  is no divisor of  $s - 1$  must be of infinite index.

Also interesting is the special case of normal subgroups.

**Lemma 8.2.** *Let  $F$  be a free group realized as the fundamental group of a bouquet of circles  $G$ , and let  $N$  be a normal subgroup. The covering graph  $\tilde{G}$  which realizes the subgroup  $N$  is the Cayley graph of  $F/N$ .*

**Proof.** Define  $C := CG(F/N)$ . There is a vertex in  $\tilde{G}$  for every (right) coset of  $N$  in  $F$ . But these are exactly the elements of  $F/N$  and thus the vertices of  $C$ . Fix a vertex  $v$  in  $C$  respectively  $\tilde{G}$ , representing the coset  $Nv$ . For every generator  $a$  of  $F$ , there is an edge starting in  $v$  end ending in the vertex  $v \cdot a$ , which represents the coset  $Nva$ . These are all edges of  $\tilde{G}$ . But the same is true for  $C$ : there is exactly one edge in  $C$  for every possible multiplication of the element  $Nv$  by some element  $Na \in F/N$ , which ends in the vertex corresponding to the product,  $Na \cdot Na' = N \cdot aa'$ . These are all edges of  $C$ . Therefore the graphs  $C$  and  $\tilde{G}$  are the same.  $\square$

The presented lemmas are just a small cutout of what can be done starting with the Nielsen-Schreier theorem. The interested reader will find more in [Sti80].

## 9 Connection to Topology

All theory given in this article makes use of topological terms, although no topological space is declared and all treated terms are newly defined. It is possible to build up the same theory from a strict topological point of view. I chose not to do so, because pure algebraically most calculations are literally the same, while the combinatorial terms are sometimes easier to understand. An outline of how the strict topological viewpoint can be treated is give in this section. Such a viewpoint is for example given in [Ser80].

For some graph  $G = (V, E)$ , consider the discrete topology on  $V$  and  $\mathcal{E}$  (this is, every subset is an open set) and the following equivalence relation on the disjoint union  $T := V \cup (\mathcal{E} \times [0, 1])$ :

$$\begin{aligned} v_1 \sim v_2 &\Leftrightarrow v_1 = v_2 \\ (e_1, s) \sim (e_2, t) &\Leftrightarrow (e_1, s) = (e_2, t) \text{ or} \\ &e_1 = e_2^{-1} \text{ and } s = 1 - t \\ v_1 \sim (e_1, s) &\Leftrightarrow s = 0 \text{ and } v_1 = \text{initial}(e_1) \text{ or} \\ &s = 1 \text{ and } v_1 = \text{final}(e_1) \end{aligned}$$

The topological quotient space  $R = T / \sim$  is called the *realization* of the graph  $G$ . What happens here, is that every directed edge is replaced by a standard interval  $[0, 1]$  while the endpoints of these intervals are identified according to the incidence relations in  $G$ . Also, the intervals belonging to two directions of one edge are identified appropriately. Paths in  $G$  directly translate to paths in  $R$  in the topological sense since every edge  $e \in \mathcal{E}$  can be identified with the path  $p(t) := (e, t)$ . Thus if  $G$  is connected, so is  $R$  (it is even path-connected). The opposite is also true: every path in  $R$  with endpoints  $(e_1, s), (e_2, t)$  can be ‘stretched’ to a path from  $(e_1, 0)$  to  $(e_2, 1)$ , and thus is homotopic to a

sequence of paths in the form of  $p$  above. This is a consequence of the simple connectedness of  $[0, 1]$ .

It is now clear that equivalent paths in  $G$  translate to homotopic paths in  $R$ , thus the fundamental group defined on graphs is isomorphic to the fundamental group of its realization.

The terms of a coverings also coincide. Suppose  $\tilde{G}$  is a covering for  $G$  then the realization  $\tilde{R}$  of  $\tilde{G}$  is a covering space for the realization  $R$  of  $G$ . Notice that the map  $\varphi : \tilde{G} \rightarrow G$  induces a map  $\psi : \tilde{R} \rightarrow R$ , by carrying over the map of vertices and define

$$\psi(e, s) := (\varphi(e), s). \tag{1}$$

One can proof that this is well-defined considering the equivalence relation. Now the following needs to be shown: for every point  $x$  in  $R$  there is a open neighbourhood  $U$  for  $x$  where the inverse image of  $U$  under  $\psi$  is a disjoint union of some open subsets  $U_\alpha \subset \tilde{R}$  where each  $U_\alpha$  is homeomorphic to  $U$ . This can be sketched as follows:

Suppose first that  $x \in R$  is no vertex in  $G$ . Then  $x = (e, t)$  is a point ‘somewhere on an edge’, that is  $e$  an edge in  $G$  and  $t \in ]0, 1[$ . The open edge  $\tilde{e} := \{(e, s) \in \tilde{R} \mid s \in ]0, 1[ \}$  then is a open neighbourhood of  $x$ , whereas the inverse image of  $\tilde{e}$  is a union of open edges in  $\tilde{R}$ , which are distinct or equal by definition (1). It clear that  $\psi$  is a homeomorphism on each of these edges in  $\tilde{R}$ .

Let  $x$  be a vertex in  $G$ . A half-open edge in  $R$  with initial point  $x$  is the set  $\dot{e} := \{(e, s) \in R \mid s \in [0, \frac{1}{2}[, \text{initial}(e) = x\} \subset R$ . The union  $U_x$  of all the half-open edges with initial point  $x$  is open (because  $[0, \frac{1}{2}[$  is open in  $[0, 1]$  and  $V, \mathcal{E}$  have discrete topology) and therefore a neighbourhood for  $x$ . As in the last paragraph the inverse image of this neighbourhood under  $\psi$  is the union of neighbourhoods  $U_{\tilde{x}^{(i)}}$  for every element  $\tilde{x}^{(i)}$  in the inverse image of  $\{x\}$ . These neighbourhoods are indeed homeomorphic to  $U_x$  from condition (iii) of definition 5.1. (This is in fact where the name ‘local homeomorphism’ condition comes from.) They are also disjoint by construction since  $[0, \frac{1}{2}[\cap] \frac{1}{2}, 1] = \emptyset$ . This shows that  $\tilde{R}$  is a covering space for  $R$ .

It is now clear, that lemma 5.3 is corresponding to the result from topology known as the path-lifting-lemma. Theorem 5.4 states functoriality, and the fact that different choices of  $\tilde{v}^{(0)}$  in the inverse image of  $v$  correspond to right cosets of the subgroup corresponds to homotopy-lifting.

As a result, the whole article could have been written in a strict topological way, which was avoided for reasons mentioned above. Thus the proof for the Nielsen-Schreier theorem given in section 6 does not only use ‘topological methods’, but is factual a proof by topology, although not explicitly visible.

## References

- [B<sup>+</sup>10] Tom De Medts Jan De Beule et al., 2010. Lecture Notes on the course Computational group theory, University of Ghent.
- [Kna06] Anthony W. Knapp. *Basic Algebra*. Birkhäuser, 2006.
- [Ser80] Jean-Pierre Serre. *Trees*. Springer, 1980.
- [Sti80] John Stillwell. *Classical Topology and Combinatorial Group Theory*. Springer, 1980.