LOCALITY AND CENTRALITY: THE VARIETY ZG

ANTOINE AMARILLI ⬤ [a] AND CHARLES PAPERMAN ⬤ [b]

[a] LTCI, Télécom Paris, Institut polytechnique de Paris, France
e-mail address: antoine.amarilli@telecom-paris.fr

[b] Univ. Lille, CNRS, INRIA, Centrale Lille, UMR 9189 CRIStAL, F-59000 Lille, France
e-mail address: charles.paperman@univ-lille.fr

ABSTRACT. We study the variety **ZG** of monoids where the elements that belong to a group are *central*, i.e., commute with all other elements. We show that **ZG** is *local*, that is, the semidirect product **ZG** ∗ **D** of **ZG** by definite semigroups is equal to **LZG**, the variety of semigroups where all local monoids are in **ZG**. Our main result is thus: **ZG** ∗ **D** = **LZG**. We prove this result using Straubing's delay theorem, by considering paths in the category of idempotents. In the process, we obtain the characterization **ZG** = **MNil** ∨ **Com**, and also characterize the **ZG** languages, i.e., the languages whose syntactic monoid is in **ZG**: they are precisely the languages that are finite unions of disjoint shuffles of singleton languages and regular commutative languages.

1. INTRODUCTION

In this paper, we study a variety of monoids called **ZG**. It is defined by requiring that the elements of the monoid that belong to a group are *central*, i.e., commute with all other elements of the monoid. The notation **ZG** thus stands for *Zentral Group*, inspired by the classical notion of centrality in group theory. We can also define **ZG** with the equation $x^{\omega+1}y = yx^{\omega+1}$ on all elements x and y, where ω is the idempotent power of the monoid.

The variety **ZG** has been introduced by Auinger [Aui00] as a subvariety of interest of the broader and better-known class **ZE** of semigroups where the idempotent elements are central. The study of **ZE** was initiated by Almeida and Azevedo [AA87]. Straubing studied in particular the variety **MNil** (called simply **V** in [Str82]) of regular languages generated by finite languages, and showed that it is exactly the variety of aperiodic monoids in **ZE**. From this, a systematic investigation of the subclasses of **ZE** was started by Almeida and pursued by Auinger: see [Alm96, page 211] and [Aui00, Aui02].

Our specific motivation to investigate **ZG** is our recently published study [AJP21] of the *dynamic membership problem for regular languages*. In this problem, introduced in [SFMS97], we study how to apply substitution update operations on an input word while maintaining the information of whether it belongs to a fixed regular language. In [AJP21], we show

that this can be performed in constant time per update for monoids in $\mathbf{ZG}$, and extend this to semigroups and languages. Further, $\mathbf{ZG}$ turns out to be a plausible boundary to characterize constant-time complexity for monoids, semigroups, and languages. However, the case of semigroups requires a study of the so-called *semidirect product* of $\mathbf{ZG}$ by definite ($\mathbf{D}$) semigroups, which we denote by $\mathbf{ZG} * \mathbf{D}$. Thus, some of our results on semigroups and languages in [AJP21] require an deeper understanding of $\mathbf{ZG}$ and $\mathbf{ZG} * \mathbf{D}$, which is the focus of this work.

The semidirect product operation on varieties used to define $\mathbf{ZG} * \mathbf{D}$ intuitively corresponds to composing finite automata via a kind of cascade operation. Its study is the subject of a large portion of semigroup theory, inspired by the classical study of semidirect products in group theory. There are also known results to understand the semidirect product specifically *by* $\mathbf{D}$. For instance, the *Derived category theorem* [Til87] studies it as a decisive step towards proving the decidability of membership to an arbitrary semidirect product, i.e., deciding if a given monoid belongs to the product. The product by $\mathbf{D}$ also arises naturally in several other contexts: the dotdepth hierarchies [Str85], the circuit complexity of regular languages [Str94], or the study of the successor relations in first-order logic [TW98, TW01, KL13].

Understanding this product with $\mathbf{D}$ is notoriously complicated. For instance, it requires specific dedicated work for some varieties like $\mathbf{J}$ or $\mathbf{Com}$ [Kna83b, Kna83a, TW85]. Also, this product does not preserve the decidability of membership, i.e., Auinger [Aui10] proved that there are varieties $\mathbf{V}$ such that membership in $\mathbf{V}$ is decidable, but the analogous problem for $\mathbf{V} * \mathbf{D}$ is undecidable. For the specific case of the varieties $\mathbf{ZG}$, $\mathbf{ZE}$, or even $\mathbf{MNil}$, we are not aware of prior results describing their semidirect product with $\mathbf{D}$.

Locality. Existing work has nevertheless identified some cases where the $* \mathbf{D}$ operator can be simplified to a much nicer *local operator*, that preserves the decidability of membership and is easier to understand. For any semigroup S, the *local monoids* of S are the subsemigroups of S of the form eSe with e an idempotent element of S. For a variety $\mathbf{V}$, we say that a semigroup belongs to $\mathbf{LV}$ if all its local monoids are in $\mathbf{V}$. It is not hard to notice that the variety $\mathbf{V} * \mathbf{D}$ is always a subvariety of $\mathbf{LV}$, i.e., that every monoid in $\mathbf{V} * \mathbf{D}$ must also be in $\mathbf{LV}$. In some cases, we can show a *locality result* stating that the other direction also holds, so that $\mathbf{V} * \mathbf{D} = \mathbf{LV}$. In those cases we say that the variety $\mathbf{V}$ is *local*. The locality of the variety of monoids $\mathbf{DA}$ [Alm96] is a famous result that has deep implications in logic and complexity [TW98, DP15, GMS17] and has inspired recent follow-up work [PS16]. Locality results are also known for other varieties, for instance the variety of *semi-lattice monoids* (monoids that are both idempotent and commutative) [McN74, BS73], any *sub-varieties of groups* [Str85, Theorem 10.2], or the $\mathcal{R}$-*trivial variety* [SJ73, Ste04, Str15]. Thus, a hope to understand the variety $\mathbf{ZG} * \mathbf{D}$ is to establish a locality result of this type for $\mathbf{ZG}$.

Contributions. This paper shows the locality of the variety $\mathbf{ZG}$. This is achieved by showing a slightly stronger statement: each variety $\mathbf{ZG}_p$ is local, where $\mathbf{ZG}_p$ is the variety of $\mathbf{ZG}$ monoids where each subgroup has a period dividing p:

Theorem 1.1. *For every $p > 0$, we have* $\mathbf{LZG}_p = \mathbf{ZG}_p * \mathbf{D}$.

As $\mathbf{ZG}$ is the union of the $\mathbf{ZG}_p$, this easily implies:

Corollary 1.2. *We have* $\mathbf{LZG} = \mathbf{ZG} * \mathbf{D}$.

Further, the variety **MNil** of aperiodic monoids in **ZE** (introduced in [Str82] and mentioned earlier) is exactly **ZG**$_1$. Thus:

Corollary 1.3. *We have* **LMNil** = **MNil** * **D**.

In the process of proving Theorem 1.1, we obtain a characterization of **ZG**$_p$-congruences, i.e., congruences $\sim$ on Σ^* where the quotient $\Sigma^*/\sim$ is a monoid of **ZG**$_p$. We show that they are always refined by a so-called n, p-*congruence*, which identifies the number of occurrences of the *frequent letters* (the ones occurring $> n$ times in the word) modulo p, and also identifies the exact subword formed by the *rare letters* (the ones occurring $\leq n$ times). Thanks to this (Theorem 3.4), we also obtain a characterization of the languages of **ZG**, i.e., the languages whose syntactic monoid is in **ZG**: they are exactly the finite unions of disjoint shuffles of singleton languages and commutative languages (Corollary 3.5). We also characterize **ZG** as a variety of monoids (Corollary 3.6): **ZG** = **MNil** $\vee$ **Com**, for **MNil** defined in [Str82] and **Com** the variety of commutative monoids.

Paper structure. We give preliminaries in Section 2 and formally define the variety **ZG**. We then give in Section 3 our characterizations of **ZG** via n, p-congruences. We then define in Section 4 the varieties **ZG** * **D**, **LZG**, **ZG**$_p$ * **D**, and **LZG**$_p$ used in Theorem 1.1 and Corollary 1.2, which we prove in the rest of the paper. We first introduce the framework of Straubing's delay theorem used for our proof in Section 5, and rephrase our result as a claim (Claim 5.7) on paths in the category of idempotents. We then study in Section 6 how to pick a sufficiently large value of n as a choice of our n, p-congruence, show in Section 7 two lemmas on paths in the category of idempotents, and finish the proof in Section 8 by two nested inductions. We conclude in Section 9.

2. Preliminaries

For a complete presentation of the basic concepts (automata, monoids, semigroups, groups, etc.) the reader can refer to the book of Pin [Pin86] or to the more recent lecture notes [Pin19]. Except for the free monoid, all semigroups, groups, and monoids that we consider are finite.

Semigroups, monoids, varieties. A *semigroup* S is a set equipped with an internal associative law (written multiplicatively). A *monoid* M is a semigroup with an *identity element* 1, i.e., an element with $1x = x1 = x$ for all $x \in M$: note that the identity element is necessarily unique. A *variety of semigroups* (resp., *variety of monoids*) is a class of semigroups (resp., monoids) closed under direct product, quotient, and subsemigroup (resp., submonoid).

For a semigroup S, we call $x \in S$ *idempotent* if $x^2 = x$. We call the *idempotent power* of $x \in S$ the unique idempotent element which is a power of x. (This means that x is idempotent iff it is its own idempotent power.) Now, the *idempotent power of S* is the least integer $\omega > 0$ such that for any element $x \in S$, the element x^ω is an idempotent power of x. We write $x^{\omega+k}$ for any $k \in \mathbb{Z}$ to mean $x^{\omega+k'}$ where $k' \in \{0, \ldots, \omega - 1\}$ is the remainder of k in the integer division by ω. For example, $x^{\omega-1}$ simply denotes $x^{2\omega-1}$, where we have $2\omega - 1 > 0$; in particular, as expected, we have $x^{\omega-1}x = x^\omega$.

Languages and congruences. We denote by Σ an alphabet and by Σ^* the set of all finite words on Σ. A *factor* of a word is a contiguous subword of that word. For $w \in \Sigma^*$, we denote by $|w|$ the length of w. For $u, v \in \Sigma^*$, we say that u is a *subword* of v if, letting $n := |u|$, there are $1 \leq i_1 < \cdots < i_n \leq |v|$ such that $u = v_{i_1} \cdots v_{i_n}$. For $w \in \Sigma^*$ and $a \in \Sigma$, we denote by $|w|_a$ the number of occurrences of a in w. A *language L* is a subset of Σ^*.

A *congruence* on a finite alphabet Σ is an equivalence relation on Σ^* which satisfies *compositionality*, i.e., it is compatible with the concatenation of words in the following sense: for any words x, y, z, and t of Σ^*, if $x \sim y$ and $z \sim t$, then $xz \sim yt$. It is a *finite index congruence* if it has a finite number of equivalence classes. For a given finite index congruence $\sim$, the quotient $\Sigma^*/{\sim}$ is a finite monoid, whose law corresponds to concatenation over Σ^* and whose identity element is the class of the empty word.

The *syntactic congruence* of a regular language L over an alphabet Σ is the congruence on Σ where $u, v \in \Sigma^*$ are equivalent if for all $s, t \in \Sigma^*$, we have $sut \in L$ iff $svt \in L$. As L is regular, this congruence is a finite index congruence. The *syntactic monoid* of L over Σ^* is then the quotient of Σ^* by the syntactic congruence for L. The *syntactic semigroup* of L is the quotient of Σ^+, the non-empty words over Σ, by the syntactic congruence.

A *variety of regular languages* is a class of regular languages which is closed under Boolean operations, left and right derivatives (also called left and right quotients), and inverse homomorphisms. By Eilenberg's theorem [Eil76], a variety $\mathbf{V}$ of monoids defines a *variety of regular languages*, namely, the languages whose syntactic monoid is in $\mathbf{V}$. Following standard practice, we abuse notation and identify varieties of monoids with varieties of languages following this correspondence, i.e., we write $\mathbf{V}$ both for the variety of monoids and the variety of languages.

Letting $\mathbf{V}$ be a variety of monoids, we say that a finite index congruence $\sim$ on Σ is a $\mathbf{V}$-*congruence* if the quotient $\Sigma^*/{\sim}$ is a monoid in $\mathbf{V}$. For a given $\mathbf{V}$-congruence $\sim$, the map $\eta : \Sigma^* \to \Sigma^*/{\sim}$, defined by associating each word with its equivalence class, is a surjective morphism to a monoid of $\mathbf{V}$. Hence, each equivalence class is a language of $\mathbf{V}$, since it is recognized by $\Sigma^*/{\sim}$.

The variety ZG. In this paper, we study the variety of monoids $\mathbf{ZG}$ defined (via Reiterman's theorem [Rei82]) by the equation: $x^{\omega+1}y = yx^{\omega+1}$. Intuitively, this says that the elements of the form $x^{\omega+1}$ are central, i.e., commute with all other elements. This clearly implies the same for elements of the form $x^{\omega+k}$ for any $k \in \mathbb{Z}$, as we will implicitly use throughout the paper:

Claim 2.1. For any monoid M in $\mathbf{ZG}$, for $x, y \in M$, and $k \in \mathbb{Z}$, we have: $x^{\omega+k}y = yx^{\omega+k}$.

Proof. Recall that $x^{\omega+k}$ denotes $x^{\omega+k'}$ where $k' \in \{0, \ldots, \omega-1\}$ is the remainder of k in the integer division by ω. Now, we can write $x^{\omega+k'}$ as $(x^{\omega+k'})^{\omega+1}$, because the latter is equal to $x^{(\omega+k')\times(\omega+1)}$ which is indeed equal to $x^{\omega+k'}$. Thus, by setting $x' := (x^{\omega+k'})$ and applying the equation of $\mathbf{ZG}$, we conclude. $\qquad\square$

Note that the elements of the form $x^{\omega+1}$ are precisely the *group elements*, namely, the elements of the monoid that are within a (possibly trivial) group that is a subsemigroup of the monoid. In particular, the neutral elements of these groups are not necessarily the neutral element of the monoid.

Claim 2.1 motivates the name $\mathbf{ZG}$, which stands for "Zentral Group": it follows the traditional notation $Z(\cdot)$ for central subgroups, and extends the variety $\mathbf{ZE}$ introduced

in [Alm94, p211] which only requires idempotents to be central. Thus, we have $\mathbf{ZG} \subsetneq \mathbf{ZE}$, and non-commutative groups are examples of monoids that are in $\mathbf{ZE}$ but not in $\mathbf{ZG}$.

In addition to the class $\mathbf{ZG}$, we will specifically study the subclasses $\mathbf{ZG}_p$ for $p > 0$ defined by imposing the equation $x^{\omega+p} = x^\omega$ in addition to the $\mathbf{ZG}$ equation. Intuitively, $\mathbf{ZG}_p$ is the variety of monoids (and associated regular languages) where group elements commute and where the *period* of all group elements divides p, where the *period* of a group element x is the smallest integer p' such that $x^{p'} = x$. Clearly, $\mathbf{ZG} = \bigcup_{p>0} \mathbf{ZG}_p$. Further, any finite language is in $\mathbf{ZG}_p$ for any $p > 0$.

The *period* of a semigroup is the least common multiple of all periods of group elements. Note that the period of a semigroup always divides ω, because the period of any group element is equal to its idempotent power, hence divides ω. Further, the period of a monoid in $\mathbf{ZG}_p$ always divides p.

3. Characterizations of **ZG**

In this section, we present our characterizations of $\mathbf{ZG}_p$ and of $\mathbf{ZG}$, which we will use to prove Theorem 1.1. We will show that $\mathbf{ZG}_p$ is intimately linked to a congruence on words called the n, p-congruence. Intuitively, two words are identified by this congruence if the subwords of the *rare letters* (occurring less than n times) are the same, and the numbers of occurrences of the *frequent letters* (occurring more than n times) are congruent modulo p. This is the standard technique of *stratification*, used also, e.g., in [TW85, Example 0.1] or in [Alm94, Section 10.8]. Formally:

Definition 3.1 (Rare and frequent letters, n, p-congruence). Fix an alphabet Σ and a word $w \in \Sigma^*$. Given a integer $n \in \mathbb{N}$ called the *threshold*, we call $a \in \Sigma$ *rare* in w if $|w|_a \leq n$, and *frequent* in w if $|w|_a > n$. The *rare alphabet* is the (possibly empty) set $\{a \in \Sigma \mid |w|_a \leq n\}$. We define the *rare subword* $w_{|\leq n}$ to be the subword of w obtained by keeping only the rare letters of w.

For $n > 0$ and for any integer $p > 0$ called the *period*, the n, p-congruence $\sim_{n,p}$ is defined by writing $u \sim_{n,p} v$ for $u, v \in \Sigma^*$ iff:

- The rare subwords are equal: $u_{|\leq n} = v_{|\leq n}$;
- The rare alphabets are the same: for all $a \in \Sigma$, we have $|u|_a > n$ iff $|v|_a > n$;
- The number of letter occurrences modulo p are the same: for all $a \in \Sigma$, we have that $|u|_a$ and $|v|_a$ are congruent modulo p. (Note that we already know this for rare letters using the previous conditions.)

We first remark that two n, p-equivalent words are also m, q-equivalent for any divisor q of p and any $m \leq n$:

Claim 3.2. For any alphabet Σ, for any $0 < m \leq n$ and $p, q > 0$ such that q divides p, the n, p-congruence refines the m, q-congruence.

Intuitively, it is less precise to look for exact subwords up to a lower threshold and with a modulo that divides the original modulo. Here is the formal proof:

Proof of Claim 3.2. We first assume that $p = q$. The claim is trivial for $m = n$, so we assume $n > m$. As $n > m$, if two words u and v have the same rare alphabet for n, then they have the same rare alphabet for m, because the number of occurrences of all rare letters for n is the same, so the same ones are also rare for m. Furthermore, if u and v have

the same rare subword for n, the restriction of this same rare subword to the rare letters for m yields the same rare subword for m. Last, the number of occurrences of the frequent letters modulo q is the same. Indeed, for the letters that were frequent for n, this is the case because their number of occurrences is congruent modulo $p = q$. For the letters that were not frequent for n, this is because their number of occurrences has to be the same because the rare subwords for n were the same.

We now assume that $q < p$. Let us now assume that $m = n$. Then the n, p-congruence refines the m, q-congruence because the rare alphabets and subwords must be equal (the threshold is unchanged), and the counts of the frequent letters modulo p determine these counts modulo q as q divides p.

To conclude in the general case, we know that the n, p-congruence refines the n, q-congruence which refines the m, q-congruence, so we conclude by transitivity. $\square$

What is more, observe that n, p-congruences are a particular case of $\mathbf{ZG}_p$-congruences:

Claim 3.3. For any alphabet Σ and $n > 0$ and $p > 0$, the n, p-congruence over Σ^* is a $\mathbf{ZG}_p$-congruence.

Proof. Let E be an equivalence class of the n, p-congruence, which we see as a language of Σ^*, and let us show that E is a language of $\mathbf{ZG}_p$. Intuitively, E defines a language where the number of occurrences of each letter modulo p is fixed, where the set of letters occurring $\leq n$ times (the rare letters) is fixed, and the subword that they achieve is also fixed. Formally, let $\Sigma = A \sqcup B$ the partition of Σ in rare and frequent letters for the class E, let u be the word over A^* associated to the class E, and let $\vec{k}$ be the $|B|$-tuple describing the modulo values for E.

We know that the singleton language $\{u\}$ is a language of $\mathbf{ZG}_p$, because it is finite. Hence, the language $U = B^* u_1 \cdots B^* u_n B^*$ is also in $\mathbf{ZG}_p$, because it is the inverse of $\{u\}$ by the morphism that erases the letters of B and is the identity on A. Similarly, the language C of words of B^* where the number of occurrences of each letter modulo p are as prescribed by $\vec{k}$ and where every letter occurs at least n times is a language of $\mathbf{ZG}_p$, because it is commutative and p is a multiple of the period of every group element. For the same reason, the language C' of words of Σ^* whose restriction to B are in C is also a language of $\mathbf{ZG}_p$, because it is the inverse image of C by the morphism that erases the letters of A and is the identity on B. Now, we remark that $E = C' \cap U$, so E is in $\mathbf{ZG}_p$, concluding the proof. $\square$

The goal of this section is to show the following result. Intuitively, it states that $\mathbf{ZG}_p$-congruences are always refined by a sufficiently large n, p-congruence. Formally:

Theorem 3.4. *For any $p > 0$, consider any $\mathbf{ZG}_p$-congruence $\sim$ over Σ^* and consider the quotient $M := \Sigma^*/\sim$. Let $n := (|M| + 1)$. Then the congruence $\sim$ is refined by the n, p-congruence on Σ.*

We first present some consequences of Theorem 3.4, and then prove Theorem 3.4.

3.1. **Consequences of Theorem 3.4.** The most important consequence of Theorem 3.4 is a characterization of the languages in $\mathbf{ZG}$, which is similar to the one obtained by Straubing in [Str82] for the variety $\mathbf{MNil}$. To define $\mathbf{MNil}$, first define the operation S^1 on a semigroup S: this is the monoid obtained from S by adding an identity element 1 to S if S does not have one. Now, define a *nilpotent semigroup* S to be a semigroup satisfying the equation $x^\omega y = y x^\omega = x^\omega$. The variety $\mathbf{MNil}$ is generated by monoids of the form S^1 for

S a nilpotent semigroup. Note that $\mathbf{MNil} \subseteq \mathbf{ZG}$, because the equation of $\mathbf{MNil}$ implies the equation of $\mathbf{ZG}$. It was shown in [Str82] that the languages of $\mathbf{MNil}$ are generated by *disjoint monomials*, that is, they are Boolean combinations of languages of the form $B^*a_1B^*a_2\cdots a_kB^*$ with $B \cap \{a_1,\ldots,a_k\} = \emptyset$.

Our analogous characterization for $\mathbf{ZG}$ is the following, obtained via Theorem 3.4:

Corollary 3.5. *A language is in* $\mathbf{ZG}$ *if and only if it can be expressed as a finite union of languages of the form* $B^*a_1B^*a_2\cdots a_kB^* \cap K$ *where* $B \cap \{a_1,\ldots,a_k\} = \emptyset$ *and* K *is a regular commutative language.*

Equivalently, we can say that every language of $\mathbf{ZG}$ is a finite union of *disjoint shuffles* of a singleton language (containing only one word) and of a regular commutative language, where the *disjoint shuffle* operator interleaves two languages (i.e., it describes the sets of words that can be achieved as interleavings of one word in each language) while requiring that the two languages are on disjoint alphabets. Intuitively, these characterizations are because the syntactic congruence of a $\mathbf{ZG}$ language is a $\mathbf{ZG}$-congruence, and Theorem 3.4 tells us that it is refined by an n,p-congruence, whose classes can be expressed as stated. Let us formally prove Corollary 3.5 using Theorem 3.4:

Proof of Corollary 3.5. One direction is easy: if a language L is of the prescribed form, then it is a Boolean combination of languages of $\mathbf{MNil}$ and regular commutative languages. These languages are in $\mathbf{ZG}$ and $\mathbf{ZG}$ is closed under Boolean operations, so indeed L is in $\mathbf{ZG}$.

For the converse direction, fix a language L in $\mathbf{ZG}$, and consider the syntactic congruence $\sim$ of L: it is a $\mathbf{ZG}$-congruence, more specifically a $\mathbf{ZG}_p$-congruence for some value $p > 0$. By Theorem 3.4, there exists $n \in \mathbb{N}$ such that $\sim$ is refined by a n,p-congruence $\sim'$. Now, by definition of the syntactic congruence, the set of words of Σ^* that are in L is a union of equivalence classes of $\sim$, hence of $\sim'$. This means that L can be expressed as the union of the languages corresponding to these classes.

Now, an equivalence class of the n,p-congruence $\sim'$ can be expressed as the shuffle of two languages: the singleton language containing the rare word defining the class, and the language that imposes that all frequent letters are indeed frequent (so the rare alphabet is as required) and that the modulo of their number of occurrences is as specified. The second language is commutative, and the disjointness of rare and frequent letters guarantees that the shuffle is indeed disjoint.

Thus, we have shown that L is a union of disjoint shuffles of a singleton language and a regular commutative language. The form stated in the corollary is equivalent, i.e., it is the shuffle of the singleton language $\{a_1\cdots a_k\}$ and of the commutative language obtained by restricting K to the subalphabet B. $\qquad\square$

Corollary 3.5 then implies a characterization of the variety of monoids $\mathbf{ZG}$. To define it, we use the *join* of two varieties $\mathbf{V}$ and $\mathbf{W}$, denoted by $\mathbf{V} \vee \mathbf{W}$, which is the variety of monoids generated by the monoids of $\mathbf{V}$ and those of $\mathbf{W}$. Alternatively, the join is the least variety containing both varieties. We then have:

Corollary 3.6. *The variety* $\mathbf{ZG}$ *is generated by commutative monoids and monoids of the form* S^1 *with* S *a nilpotent semigroup. In other words, we have:* $\mathbf{ZG} = \mathbf{MNil} \vee \mathbf{Com}$.

Proof. Clearly $\mathbf{ZG}$ contains both $\mathbf{Com}$ and $\mathbf{MNil}$. Furthermore, by Corollary 3.5, any language in $\mathbf{ZG}$ is a union of intersections of a language in $\mathbf{MNil}$ and a language in $\mathbf{Com}$. Hence, it is in the variety generated by these two varieties of languages, concluding the proof. $\qquad\square$

On a different note, we will also use Theorem 3.4 to show a technical result that will be useful later. It intuitively allows us to regroup and move arbitrary elements:

Corollary 3.7. *For any monoid M in $\mathbf{ZG}$, letting $n \geq |M| + 1$, for any element m of M and elements $m_1, \ldots, m_n$ of M, we have*

$$m \cdot m_1 \cdot m \cdot m_2 \cdot m \cdots m \cdot m_n \cdot m \cdot m_n \cdot m = m^{n+1} \cdot m_1 \cdots m_n.$$

Proof. We consider the free monoid M^*. Let $\eta : M^* \to M$ be the onto morphism defined by $\eta(m) := m$ for all $m \in M$. Let $\sim$ be the congruence that η induces over M^*, i.e., for $u, v \in M^*$, we have $u \sim v$ if $\eta(u) = \eta(v)$. Remark that, as M is in $\mathbf{ZG}$, more specifically in $\mathbf{ZG}_p$ for some $p > 0$, the congruence $\sim$ is a $\mathbf{ZG}_p$-congruence by definition. Hence, by Theorem 3.4, $\sim$ is refined by a n, p-congruence where $n = |M| + 1$. Now, consider the two words in the equation that we wish to show: they are words of M^*. As the letter m over is then a frequent letter, we know that the two words are indeed n, p-congruent, which concludes the proof. □

3.2. **Proof of Theorem 3.4.** Having spelled out the consequences of Theorem 3.4, we prove it in the rest of this section. It crucially relies on a general result about $\mathbf{ZG}$ that we will use in several proofs, and which is shown by elementary equation manipulations:

Lemma 3.8. *Let M be a monoid of $\mathbf{ZG}$, let ω be the idempotent power, and let $x, y \in M$. Then we have: $(xy)^\omega = x^\omega y^\omega$.*

Proof. We show this claim by showing that two equalities establishing that each side of the equation is equal to the same term, namely, $x^\omega y^\omega (xy)^\omega$. Let us first show the first equality:

$$(xy)^\omega = x^\omega y^\omega (xy)^\omega. \tag{3.1}$$

To show Equation 3.1, remember that we have $(xy)^\omega = xy(xy)^{\omega-1}$ from the definition of $(xy)^{\omega-1}$ in the preliminaries. Now, as $(xy)^{\omega-1}$ is central, the right-hand side is equal to $x(xy)^{\omega-1}y$. By injecting an $(xy)^\omega$ in the latter, we obtain:

$$(xy)^\omega = x(xy)^{\omega-1}(xy)^\omega y.$$

Applying this equality ω times gives:

$$(xy)^\omega = (x(xy)^{\omega-1})^\omega (xy)^\omega y^\omega.$$

Now, we can expand $(x(xy)^{\omega-1})^\omega$, commuting the $(xy)^{\omega-1}$ to regroup the x into x^ω and regroup the $(xy)^{\omega-1}$ into $((xy)^{\omega-1})^\omega$ which is equal to $(xy)^\omega$, so that the first factor of the right-hand side is equal to $x^\omega (xy)^\omega$. By commuting, we obtain Equation 3.1.

The second equality is:

$$x^\omega y^\omega = x^\omega y^\omega (xy)^\omega. \tag{3.2}$$

To show Equation 3.2, note that we have $x^\omega y^\omega = x^{\omega-1}xy y^{\omega-1}y$, so by the equation of $\mathbf{ZG}$ we get:

$$x^\omega y^\omega = x^{\omega-1}y^{\omega-1}xy.$$

Now, we have $x^{\omega-1}y^{\omega-1} = x^{\omega-1}x^\omega y^{\omega-1}y^\omega$, and by the equation of $\mathbf{ZG}$ we have:

$$x^{\omega-1}y^{\omega-1} = x^{\omega-1}y^{\omega-1}x^\omega y^\omega.$$

Inserting the second equality in the first, we have:

$$x^\omega y^\omega = x^{\omega-1}y^{\omega-1}x^\omega y^\omega xy.$$

Now, applying this equality ω times gives $x^\omega y^\omega = (x^{\omega-1} y^{\omega-1})^\omega x^\omega y^\omega (xy)^\omega$. As the first factor of the right-hand side is equal to $x^\omega y^\omega$, we get $x^\omega y^\omega = x^\omega y^\omega (xy)^\omega$. This establishes Equation 3.2.

From Equations 3.1 and 3.2, we immediately conclude the proof. $\qquad\square$

To continue with our proof of Theorem 3.4, thanks to Lemma 3.8, we can now show a kind of "normal form" for **ZG**-congruences, by arguing that any word can be rewritten to a word where frequent letters are moved to the beginning of the word, without breaking equivalence for the **ZG**-congruence. This relies on Lemma 3.8 and allows us to get to the notion of n-equivalence. Specifically:

Claim 3.9. Let $\sim$ be a **ZG**-congruence on Σ. Let $n := (|M| + 1)$ where M is the monoid associated to $\sim$. Then, for all $w \in \Sigma^*$, for every letter $a \in \Sigma$ which is frequent in w (i.e., $|w|_a > n$), writing w' the restriction of w to $\Sigma \setminus \{a\}$, and writing $w'' := a^{|w|_a} w'$, we have: $w \sim w''$.

Proof. Define $n := |M| + 1$ as in the claim statement, and let $\mu : \Sigma^* \to M = \Sigma^*/\sim$ be the morphism associated to $\sim$. Remark that by definition, for any words u, v, we have $u \sim v$ iff $\mu(u) = \mu(v)$.

Let us take an arbitrary w and $a \in \Sigma$ such that a is frequent in w. We can therefore write $w = w_1 a w_2 a \cdots w_m a w_{m+1}$ with $m = |w|_a > n > |M|$. Furthermore, letting $x_l = \mu(w_1 a w_2 a \cdots w_l a)$ for each $1 \le l \le m$, as $m > |M|$ we know by the pigeonhole principle that there exist $1 \le i < j \le m$ such that $x_i = x_j$. Furthermore, we have $x_j = x_i z \mu(a)$ where $z = \mu(w_{i+1} a \cdots w_j)$. By applying the equation ω times, we have that $x_i = x_i (z\mu(a))^\omega$.

Now, by Lemma 3.8, we have $(z\mu(a))^\omega = z^\omega \mu(a)^\omega$. This is equal to $z^\omega \mu(a)^\omega \mu(a)^\omega$, and by now applying Lemma 3.8 in reverse we conclude that $(z\mu(a))^\omega = (z\mu(a))^\omega \mu(a)^\omega$. Finally, we obtain $x_i = x_i(z\mu(a))^\omega = x_i(z\mu(a))^\omega \mu(a)^\omega = x_i \mu(a)^\omega$.

Now, the equation of **ZG** ensures that $\mu(a)^\omega$ is central, so we can commute it in $\mu(w)$ and absorb all occurrences of $\mu(a)$ in $\mu(w)$, then move it at the beginning, while keeping the same μ-image. Formally, from $x_i = x_i \mu(a)^\omega$, we have

$$\mu(w) = \mu(w_1)\mu(a) \cdots \mu(w_i)\mu(a)\mu(a)^\omega \mu(w_{i+1})\mu(a) \cdots \mu(w_m)\mu(a)\mu(w_{m+1}),$$

and we commute $\mu(a)^\omega$ to merge it with all $\mu(a)$ and then commute the resulting $\mu(a)^{\omega+|w|_a}$ to obtain $\mu(a)^{\omega+|w|_a} w'$ with w' as defined in the statement of the claim.

Now, remark that for any $i > |M|$, for any x in M, we have $x^{\omega+i} = x^i$. This is because, letting $k \le |M|$ be the idempotent power of x, we have $x^\omega = x^k$, hence $x^{\omega+i} = x^{k+i} = x^{i-k} x^k x^k = x^{i-k} x^k = x^i$. Note that $i - k > 0$ because $i > |M|$ and $k \le |M|$, so x^{i-k} is well-defined.

By applying this to $x = \mu(a)$ and $i = |w|_a > |M|$, we deduce that $\mu(w) = \mu(a)^{|w|_a} \mu(w')$. This establishes that $w \sim w''$ and concludes the proof. $\qquad\square$

We can now conclude the proof of Theorem 3.4:

Proof of Theorem 3.4. Let $\sim$ be a $\mathbf{ZG}_p$-congruence on Σ^*, M its associated monoid, and fix $n := |M| + 1$ as in the theorem statement. Let u and v be two n, p-congruent words of Σ^*, we need to prove that they are indeed $\sim$-equivalent. Let $\Sigma' = \{a_1, \ldots, a_r\}$ be the subset of letters in Σ that are frequent in u (hence in v, as they are n, p-congruent). By successive applications of Claim 3.9 for every frequent letter in Σ', starting with u, we know that $u \sim a_1^{|u|_{a_1}} \cdots a_r^{|u|_{a_r}} u_{|\le n}$. Likewise, we have $v \sim a_1^{|v|_{a_1}} \cdots a_r^{|v|_{a_r}} v_{|\le n}$. Now, for any $1 \le i \le r$,

the values $|u|_{a_i}$ and $|v|_{a_i}$ are greater than n which is $\geq |M|$, and they are congruent modulo p, which is a multiple of the period of a_i, so we have $a_i^{|u|_{a_i}} \sim a_i^{|v|_{a_i}}$. We also know by definition of the n, p-congruence that $u_{|\leq n} = v_{|\leq n}$. By compositionality of $\sim$, all of this establishes that $u \sim v$. Thus, the n, p-congruence indeed refines the $\sim$-congruence, concluding the proof. $\quad\square$

4. Defining $\mathbf{ZG} * \mathbf{D}$ and $\mathbf{LZG}$, and Result Statement

We have given our characterizations of $\mathbf{ZG}$ and presented some preliminary results. We now define $\mathbf{LZG}_p$ and $\mathbf{ZG}_p * \mathbf{D}$ and show that they are equal (Theorem 1.1), and deduce the same for $\mathbf{LZG}$ and $\mathbf{ZG} * \mathbf{D}$.

$\mathbf{ZG}_p * \mathbf{D}$ **and** $\mathbf{ZG} * \mathbf{D}$. We denote by $\mathbf{D}$ the variety of the *definite semigroups*, i.e., the semigroups satisfying the equation $yx^\omega = x^\omega$. For $p > 0$, the variety of semigroups $\mathbf{ZG}_p * \mathbf{D}$ is the variety generated by the *semidirect products* of monoids in $\mathbf{ZG}_p$ and semigroups in $\mathbf{D}$. We recall for completeness the definition of the semidirect product operator, even though we will not use it directly in this paper. Given two semigroups S and T, a *semigroup action* of S on T is defined by a map $\text{act} \colon S \times T \to T$ such that $\text{act}(s_1, \text{act}(s_2, t)) = \text{act}(s_1 s_2, t)$ and $\text{act}(s, t_1 t_2) = \text{act}(s, t_1)\text{act}(s, t_2)$. We then define the product $\circ_{\text{act}}$ on the set $T \times S$ as follows: for all s_1, s_2 in S and t_1, t_2 in T, we have: $(t_1, s_1) \circ_{\text{act}} (t_2, s_2) := (t_1 \text{act}(s_1, t_2), s_1 s_2)$. The set $T \times S$ equipped with the product $\circ_{\text{act}}$ is a semigroup called the *semidirect product* of S by T, denoted $T \circ_{\text{act}} S$. We then define $\mathbf{ZG} * \mathbf{D}$ as the variety of semidirect products of monoids in $\mathbf{ZG}$ and semigroups in $\mathbf{D}$.

Remark that the $*$ operation is equivalent to the *wreath product of varieties*. We further note for the expert reader that we could equivalently replace $\mathbf{D}$ by the variety $\mathbf{LI}$ of *locally trivial semigroups*. We refer to [Str85] for a detailed presentation on this subject.

$\mathbf{LZG}_p$ **and** $\mathbf{LZG}$. Last, we introduce the varieties $\mathbf{LZG}_p$ and $\mathbf{LZG}$. For $p > 0$, the variety $\mathbf{LZG}_p$ is the variety of semigroups S such that, for every idempotent e of S, the subsemigroup eSe of elements that can be written as ese for some $s \in S$ is in $\mathbf{ZG}_p$. Note that this subsemigroup is actually a monoid, called the *local monoid* of e, and that its identity is e. The variety $\mathbf{LZG}$ is defined analogously but for $\mathbf{ZG}$, and clearly $\mathbf{LZG} = \bigcup_{p>0} \mathbf{LZG}_p$. Remark that a semigroup is in $\mathbf{LZG}$ iff it satisfies the following equation: for any x, y and z in S, we have:
$$(z^\omega x z^\omega)^{\omega+1}(z^\omega y z^\omega) = (z^\omega y z^\omega)(z^\omega x z^\omega)^{\omega+1}.$$
This incidentally shows by Reiterman's theorem that $\mathbf{LZG}$ is indeed a variety of semigroups.

Main result. Our main result, stated in Theorem 1.1, is that, for any $p > 0$, the varieties $\mathbf{ZG}_p * \mathbf{D}$ and $\mathbf{LZG}_p$ are actually equal. In particular, for $p = 1$, we get that $\mathbf{MNil} * \mathbf{D} = \mathbf{LMNil}$.

Let us first remark that Theorem 1.1 implies Corollary 1.2 stated in the introduction, namely, that $\mathbf{LZG} = \mathbf{ZG} * \mathbf{D}$:

Proof of Corollary 1.2. Any semigroup in $\mathbf{LZG}$ is in $\mathbf{LZG}_p$ for some p, hence in $\mathbf{ZG}_p * \mathbf{D}$ by locality of $\mathbf{ZG}_p * \mathbf{D}$, hence in $\mathbf{ZG} * \mathbf{D}$. Conversely, by definition of $\mathbf{ZG} * \mathbf{D}$ being a variety, each semigroup S of $\mathbf{ZG} * \mathbf{D}$ is obtained by applying the quotient, subsemigroup, and product operators to semigroups $S_1, \ldots, S_k$ for some $k \in \mathbb{N}$, where each S_i is a semidirect product of a monoid in $\mathbf{ZG}_{p_i}$ for a certain $p_i > 0$ and of a semigroup in $\mathbf{D}$. Letting $p := \prod_{1 \leq i \leq k} p_i$,

all semigroups $S_1, \ldots, S_k$ are then in $\mathbf{ZG}_p * \mathbf{D}$, and applying the operators then witnesses that S also belongs to $\mathbf{ZG}_p * \mathbf{D}$. By locality of $\mathbf{ZG}_p * \mathbf{D}$, we know that S belongs to $\mathbf{LZG}_p$, hence to $\mathbf{LZG}$. $\qquad\square$

Thus, fixing the value $p > 0$ for the rest of this paper, the only remaining task is to prove Theorem 1.1 on $\mathbf{LZG}_p$ and $\mathbf{ZG}_p * \mathbf{D}$. Note that we can freely use all equational results shown about monoids of $\mathbf{ZG}$ (e.g., Lemma 3.8), as they also hold for monoids in $\mathbf{ZG}_p$.

To prove Theorem 1.1, we will first present the general framework of Straubing's delay theorem in the next section and show the easy inclusion $\mathbf{ZG}_p * \mathbf{D} \subseteq \mathbf{LZG}_p$, before moving on with the rest of the proof.

5. Straubing's Delay Theorem

To show our main result, we use Straubing's delay theorem from [Str85]. We first give some prerequisites to recall this result. To this end, let us first define a general notion of *finite category*:

Definition 5.1. A *finite category* on a finite set of *objects* O defines, for every pair $(o, o') \in O$ of objects, a finite set $\mathrm{C}_{o,o'}$ of *arrows*. An arrow in $\mathrm{C}_{o,o'}$ is said to be *going from o to o'*; we call o the *starting object* and o' the *ending object*.

The overall set of arrows $\bigcup_{o,o' \in O} \mathrm{C}_{o,o'}$ is equipped with a composition law: for any objects o, o', o'', for any arrows $a \in \mathrm{C}_{o,o'}$ and $b \in \mathrm{C}_{o',o''}$, the composition law gives us ab which must be an arrow of $\mathrm{C}_{o,o''}$. (Note that the objects o, o', and o'' are not necessarily pairwise distinct.) Further, this composition law must be associative. What is more, we require that for any object o, there exists an arrow in $\mathrm{C}_{o,o}$ which is the identity for all elements that it can be combined with (note that these arrows are in particular unique).

We now define the notion of *category of idempotents* of a semigroup:

Definition 5.2 (Category of idempotents)**.** Let S be a semigroup. The *category of idempotents S_E* of S is the finite category defined as follows:

- The objects of S are the idempotents of S.
- For any idempotents e_1 and e_2 and any element x of S such that $x \in e_1 S e_2$, we have an arrow labeled by x going from e_1 to e_2, which we will denote by (e_1, x, e_2). Formally, $\mathrm{C}_{e_1,e_2} = \{(e_1, x, e_2) \mid x \in e_1 S e_2\}$.

The composition law of the category is $(e_1, x, e_2)(e_2, y, e_3) = (e_1, xy, e_3)$. Note that it is clearly associative thanks to the associativity of the composition law on S. The identity element in $\mathrm{C}_{e,e}$ for an idempotent e is simply (e, e, e).

Let us now study S_E in more detail. We denote by $B := \{(e_1, x, e_2) \mid x \in e_1 S e_2\}$ the set of arrows of the category of idempotents.

A *path* of S_E is a nonempty word of B^* whose sequence of arrows is *valid*, i.e., the ending object of each arrow except the last one is equal to the starting object of the next arrow. Because S_E is a category, each path is equivalent to an element of the category, i.e., composing the arrows of the path according to the composition law of the category will give one arrow of the category, whose starting and ending objects will be the *starting object of the path* (i.e., that of the first arrow) and the *ending object of the path* (i.e., of the last arrow). Two paths are *coterminal* if they have the same starting and ending object. Two paths π_1 and π_2 are S_E-*equal* if they evaluate to the same category element, which we write

$\pi_1 \equiv \pi_2$. Note that if two paths are S_{E}-equal then they must be coterminal. A *loop* is a path whose starting and ending objects are the same.

A *congruence* on B^* is an equivalence relation $\sim$ over B^* that satisfies compositionality as previously defined. Note that the relation is also defined on words of B^* that are not valid, i.e., that do not correspond to paths; the compositionality requirement also applies to such words.

Definition 5.3 (Compatible congruence)**.** A congruence $\sim$ on B^* is said to be *compatible* with S_{E} if for any two coterminal paths π_1 and π_2 of S_{E} such that $\pi_1 \sim \pi_2$, we have $\pi_1 \equiv \pi_2$. In other words, $\sim$ is compatible with S_{E} iff, on words of B^* that are coterminal paths, it refines S_{E}-equality.

Recall the notion of a $\mathbf{ZG}_p$-congruence from Section 2. We are now ready to state Straubing's delay theorem, which was introduced in [Str85] and rephrased in [Til87]. The theorem applies to any variety, but we state it specifically for $\mathbf{ZG}_p$ for our purposes. The theorem gives us an alternative characterization of $\mathbf{ZG}_p * \mathbf{D}$:

Theorem 5.4 (Straubing's delay theorem, Theorem 5.2 of [Str85])**.** *A semigroup S is in $\mathbf{ZG}_p * \mathbf{D}$ iff, writing S_{E} the category of idempotents of S and B its set of arrows, there exists a $\mathbf{ZG}_p$-congruence on B^* which is compatible with S_{E}.*

Using our notion of n, p-congruence, via Claim 3.3 and Theorem 3.4, we rephrase it again:

Corollary 5.5. *A semigroup S is in $\mathbf{ZG}_p * \mathbf{D}$ iff, writing S_{E} and B as above, there exists an n, p-congruence on B^* which is compatible with S_{E}.*

Before moving on to the full proof of our main theorem (Theorem 1.1), we conclude the section by noticing that the Straubing delay theorem implies the easy direction of our result, namely, if L is in $\mathbf{ZG}_p * \mathbf{D}$ then L is in $\mathbf{LZG}_p$. This easy direction follows directly from [Til87], but let us provide a self-contained argument:

Claim 5.6. We have $\mathbf{ZG}_p * \mathbf{D} \subseteq \mathbf{LZG}_p$.

Proof. If S is in $\mathbf{ZG}_p * \mathbf{D}$, then by Theorem 5.4, there exists a $\mathbf{ZG}_p$-congruence $\sim$ compatible with S_{E}. Let us now show that S is in $\mathbf{LZG}_p$ by showing that, for any idempotent e, the local monoid eSe is in $\mathbf{ZG}_p$. Let e be an idempotent. By definition of S_{E}, the local monoid eSe is isomorphic to the subset of arrows of S_{E} going from e to e, with their composition law. Let us denote this subset by B_e. Define $\sim_e$ to be the specialization of the relation $\sim$ to B_e. Remark that $M = B_e^*/\sim_e$ is a submonoid of $B^*/\sim$, and is hence in $\mathbf{ZG}_p$ because $B^*/\sim$ is, and $\mathbf{ZG}_p$ is a variety. Remark that since all words in B_e^* are valid paths in S_{E}, the local monoid $N := eSe$ defines a congruence $\sim_2$ over B_e^* where two paths are equivalent if they evaluate to the same monoid element. We know that $\sim$, hence $\sim_e$, refines this congruence $\sim_2$. Hence, N is a quotient of M. Thus, eSe is a quotient of $B_e^*/\sim_e$, which is a submonoid of a monoid in $\mathbf{ZG}_p$, concluding the proof. $\qquad\square$

In the rest of this paper, we show the much harder direction, i.e., if S is in $\mathbf{LZG}_p$ then S is in $\mathbf{ZG}_p * \mathbf{D}$. To prove this, using Corollary 5.5, it suffices to show:

Claim 5.7. Let S be a semigroup of $\mathbf{LZG}_p$, write S_{E} its category of idempotents and B the set of arrows of S_{E}. There exists an integer $n > 0$ such that the n, p-congruence on B^* is compatible with S_{E}.

This result then implies, by our rephrasing of Straubing's result (Corollary 5.5), that S is in $\mathbf{ZG}_p * \mathbf{D}$. So in the rest of this paper we prove Claim 5.7. The proof is structured in three sections. First, in Section 6, we carefully choose the threshold n of the n, p-congruence to be "large enough" so that we can enforce a gap between the number of occurrences of the *rare* and *frequent* arrows according to the n, p-congruence, also ensuring a strong connectedness property on the category of idempotents. Second, in Section 7, we show auxiliary results about paths in the category of idempotents, using the **ZG** equation and the properties of the threshold. Third, in Section 8, we conclude the proof, doing first an outer induction on the total number of rare arrow occurrences, and then an inner induction on the number of frequent arrows, using a so-called *ear decomposition* of the category.

6. Choosing the Threshold of the Congruence

In this section, we explain how to choose the threshold n to prove Claim 5.7. Our threshold will enforce a large enough "gap" between the total number of occurrences of the rare letters and the number of occurrences of the frequent letters. This is called a *distant* threshold and will be useful in the rest of the argument.

This section is split into two subsections. In the first subsection, we formally define the notion of a distant rare-frequent threshold, and show how such a threshold can be used to find factors with no rare letters and a large number of copies of a frequent letter. We further explain that we can indeed find a sufficiently distant rare-frequent threshold. This is a generic result on words and alphabets that does not depend on **ZG** or on the category of idempotents. In the second subsection, we instantiate this result for paths in the category of idempotents, and explain how a distant rare-frequent threshold incidentally ensures a strong connectedness property on the category. This last subsection focuses on the category of idempotents, but is also generic in the sense that it does not use the **ZG** equation. Hence, all of the present section applies to arbitrary semigroups.

6.1. **Finding sufficiently distant thresholds.** We now give our formal definition of a *distant* rare-frequent threshold:

Definition 6.1. For Σ an alphabet, $u \in \Sigma^*$, and $m > 0$, we say that an integer $n > 0$ is an *m-distant rare-frequent threshold* for u if, letting $\Sigma_r := \{a \in \Sigma \mid |u|_a \leq n\}$ be the rare alphabet of u for n, then the *total* number of occurrences of rare letters in u is less than a proportion $1/m$ of the threshold n, minus one. Formally: $\sum_{a \in \Sigma_r} |u|_a \leq \frac{n}{m} - 1$.

Note that if $n > 0$ is an m-distant rare-frequent threshold for u, then it is in particular an m'-distant rare-frequent threshold for any $0 < m' \leq m$.

If a word has a distant rare-frequent threshold, then any frequent letter can be found in sufficiently many occurrences in some factor containing no rare letter. This will be useful in pumping arguments, and is the motivation for the definition. Formally:

Lemma 6.2. *For any alphabet Σ and $m > 0$, if a word $u \in \Sigma^*$ has an m-distant rare-frequent threshold, then for any frequent letter a, the word u has a factor containing no rare letter and containing at least $m + 1$ occurrences of a.*

Proof. As a is a frequent letter, by definition, its number $|u|_a$ of occurrences in u is such that $|u|_a > n$. Let $r := \sum_{a \in \Sigma_r} |u|_a$ be the total number of occurrences of the rare letters. The definition of an m-distant rare-frequent threshold ensures that $r \leq \frac{n}{m} - 1$. Thus, $r + 1 \leq \frac{n}{m}$.

Multiplying by m, we obtain $(r+1)m \leq n$. Thus, $|u|_a > (r+1)m$. As there are r rare letters in total, there are $r+1$ subwords between them containing no rare letter, so the inequality implies that one of them contains $> m$ occurrences of a, i.e., $\geq m+1$ occurrences of a. This concludes the proof. □

We now intuitively state the existence, for any $m > 0$, of an m-distant rare-frequent threshold that can be used for any word u. The precise claim is more complicated to phrase, as we cannot pick one n which can serve as a m-distant rare-frequent threshold for any word u. Indeed, for any choice of n, there will always be words u where the number of rare letter occurrences happens to be close to n. However, we can pick a large enough n such that, given any word u, we can pick some $0 < n' \leq n$ as an m-distant rare-frequent threshold for u. We will state the claim more generally about picking a threshold for a pair of words u_1 and u_2.

Here is the formal claim. We repeat that this result is a general claim about words, which is not specific to **ZG**.

Lemma 6.3. *For any alphabet Σ and $m > 0$, there exists an integer $n > 0$ ensuring the following: for any words $u_1, u_2 \in \Sigma^*$, there exists an integer n' with $0 < n' \leq n$ which is an m-distant rare-frequent threshold for u_1 and for u_2.*

Let us now prove Lemma 6.3 in the rest of this subsection. We first show an abstract result capturing the essence of the underlying pigeonhole principle argument:

Claim 6.4. *For any $d > 0$ and $m' > 0$, there exists an integer $n \geq m'$ ensuring the following: for any d-tuple T of integers, there exists an integer n' with $m' \leq n' \leq n$ such that $\sum_{i \in F} T_i \leq \frac{n'}{m'}$, where $F = \{i \mid T_i \leq n'\}$.*

Intuitively, the value m' will be computed from m to ensure the "minus one" gap in Definition 6.1, the "dimension" d will be the cardinality of the alphabet Σ (multiplied by 2 as we consider two words), the integer n is the threshold that we will choose, and n' is the value that we wish to obtain. Let us prove Claim 6.4:

Proof of Claim 6.4. Let us take $n := (m'd)^{d+2}$, which ensures $n \geq m'$. The candidate values of n' that we will consider are intuitively the following: $m'd, (m'd)^2, \ldots, (m'd)^{d+2}$. Now take any d-tuple T. For any $1 \leq i \leq d+2$, let $R_i = \{i' \mid T_{i'} \leq (m'd)^i\}$ be the coordinates where T has a value $\leq (m'd)^i$. By definition, we have $\emptyset \subseteq R_1 \subseteq \cdots \subseteq R_{d+2} \subseteq \{1, \ldots, d\}$. Applying the pigeonhole principle on the cardinalities of these sets, there are $1 \leq i < j \leq d+2$ such that $R_i = R_j$.

Let us set $n' := (m'd)^j$. By construction, we have $n' \geq m'$. Now, consider the sum $\sum_{i' \in R_j} T_{i'}$. As $R_j = R_i$, we know that for every $i' \in R_j$, we have $T_{i'} \leq (m'd)^i$. Thus, the sum is at most d times this value because T is a d-tuple. Formally, $\sum_{i' \in R_j} T_{i'}$ is at most $d(m'd)^i$, hence it is $\leq \frac{(m'd)^{i+1}}{m'}$, so it is $\leq \frac{n'}{m'}$ by definition of $n' = (m'd)^j$ because $i < j$. This concludes the proof. □

With Claim 6.4, it is now easy to show Lemma 6.3:

Proof of Lemma 6.3. Let n be the value given by Claim 6.4 when taking $d := 2|\Sigma|$ and $m' := (m+1)m$.

Let us consider any pair of words $u_1, u_2 \in \Sigma^*$. Let T be the d-tuple of the letter occurrences of u_1, followed by those of u_2. The statement of Claim 6.4 ensures that there

exists an integer n' with $n' \geq m' > m > 0$ such that, when using n' as the rare-frequent threshold, the total number of rare letters in u_1 plus in u_2 is $\leq \frac{n'}{m'}$.

Now, let us show that $\frac{n'}{m'} \leq \frac{n'}{m} - 1$. To this end, as $m' = m(m+1)$, first note that $\frac{n'}{m'} \leq \frac{n'}{m+1}$. To show now that $\frac{n'}{m+1} \leq \frac{n'}{m} - 1$, let us evaluate $\frac{n'}{m} - 1 - \frac{n'}{m+1}$ and check that it is nonnegative. This difference evaluates to $\frac{n'-m(m+1)}{m(m+1)}$. Now, we have $n' \geq m'$, so $n' \geq m(m+1)$, which is ≥ 0. So indeed $\frac{n'}{m'} \leq \frac{n'}{m} - 1$.

Combining the conclusions of the two previous paragraphs, we obtain that the total number of rare letters u_1 plus in u_2 is $\leq \frac{n'}{m} - 1$. So the same is true of the rare letters in u_1, and of the rare letters in u_2. By contrast, the frequent letters in u_1 occur $> n'$ times by definition, and the same is true of the frequent letters in u_2. Hence, by Definition 6.1, the value n' is an m-distant rare-frequent threshold for u_1 and for u_2, concluding the proof. □

Thus, we have shown that, for any alphabet size and desired distance m, we can pick a value n ensuring that for any two words we can use some $0 < n' \leq n$ as an m-distant rare-frequent threshold for them.

6.2. **Choice of threshold and resulting properties.** In the rest of this section, we instantiate the generic argument of the previous section to the choice of an arbitrary semigroup S and its category of idempotents S_E: we claim the existence of a suitably distant rare-frequent threshold, and observe that it additionally ensures the strong connectedness of the category. However, the argument is still generic in the sense that it applies to an arbitrary semigroup S, even if it is not in **ZG**.

Let us first rephrase Lemma 6.3 to this setting by picking as alphabet the arrows B of the category of idempotents S_E of S and by choosing $m := |S|$. We immediately obtain:

Claim 6.5. Let S be a semigroup and let S_E be its category of idempotents. There exists an integer $n > 0$ ensuring the following: for any paths u_1, u_2 in S_E, there is $0 < n' \leq n$ which is a $|S|$-distant rare-frequent threshold for u_1 and for u_2.

The rephrasing of Lemma 6.2 is:

Claim 6.6. If a path u in S_E has an $|S|$-distant rare-frequent threshold, then for any frequent arrow a, the path u has a factor containing no rare arrow and containing at least $m + 1$ occurrences of a.

We now close the section by observing that having an m-distant rare-frequent threshold for a path w, indeed having simply an 1-distant rare-frequent threshold for w, enforces a strong connectedness property on the category. Specifically, the set of frequent arrows in this path for this threshold must form a so-called *union of strongly connected components (SCCs)*:

Definition 6.7. Given S_E and a subset B' of its set of arrows B, we say that B' *is a union of SCCs* if, letting G be the directed graph on the objects of S_E formed of the arrows of B', then all connected components of G are strongly connected.

Claim 6.8. Fix S and S_E and B, let $w \in B^*$ be a path of S_E, and let $n' > 0$ be a 1-distant rare-frequent threshold of w. Then the set of frequent arrows of w for n' is a union of SCCs.

Proof. Consider G the directed graph of Definition 6.7. Let us assume by way of contradiction that G has a connected component which is not strongly connected. This means that there

exists an edge (u, v) of G such that there is no path from u to v in G. Consider any frequent arrow a in S_{E} achieving the edge (u, v) of G.

As a is frequent in w, we know that a occurs strictly more than n' times in w, hence w contains at least n' *return paths*, i.e., paths from the ending object v of a back to the starting object u of a. As there is no path from v to u in G, each one of these paths must contain an arrow of B which is rare in w.

Hence, the total number of rare arrows in w is at least n'. But the 1-distant rare-frequent threshold condition imposes that the total number of rare arrow occurrences in w is $\leq n' - 1$. We have thus reached a contradiction.   □

7. The Loop Insertion and Prefix Substitution Lemmas

We now show two auxiliary results on the category of idempotents of **LZG** semigroups, to be used in the sequel. The first result is the *loop insertion lemma*: it allows us to insert any loop of frequent arrows to the power ω without affecting equivalence. The second is the *prefix substitution lemma*: it allows us to replace a prefix of frequent arrows by another, without affecting equivalence, up to inserting a loop later in the path.

The results shown in this section hold for semigroups in **LZG**, hence for those in **LZG**$_p$ for any $p > 0$. Thus, let us fix a semigroup S in **LZG**. Recall that S_{E} denotes the category of idempotents of S, and denote by B the set of arrows of S_{E}.

Remember that, by definition of the category of idempotents S_{E}, for any idempotent e of S, the set of loops with starting and ending object e under the product law of the category forms a monoid which is isomorphic to the local monoid eSe of S. As we know that S is in **LZG**, all its local monoids are in **ZG**. Hence the **ZG** equation immediately applies to loops in the category of idempotents, which we will often use in the results of this section:

Claim 7.1. Let x and y be two coterminal loops of S_{E}, let $k \in \mathbb{Z}$, and let ω be an idempotent power of S. We have: $x^{\omega+k} y \equiv y x^{\omega+k}$.

7.1. **Loop insertion lemma.** The loop insertion lemma allows us, when we have a sufficiently distant rare-frequent threshold n', to insert any arbitrary loop raised to the power ω without changing the category element to which a path evaluates. We also show that this change does not affect n', ω-equivalence: this will suffice to preserve n', p-equivalence (where p is the period of S), which will be useful later (we come back to this at the beginning of Section 8). Formally:

Lemma 7.2 (Loop insertion lemma). *Let π be a path, and assume that n' is an $|S|$-distant rare-frequent threshold for π. Let $\pi = rt$ be a decomposition of π (with r or t possibly empty), let o be the object between r and t (i.e., the ending object of r, or the starting object of t if r is empty), and let π' be a loop on o that only uses frequent arrows. Then $\pi \equiv r(\pi')^\omega t$. (Note that these two paths are also n', ω-equivalent by construction.)*

In the rest of this subsection, we show Lemma 7.2. We first rephrase the claim to the following auxiliary result:

Claim 7.3. Let π be a path, and assume that n' is an $|S|$-distant rare-frequent threshold for π. Let $\pi = rt$ be a decomposition of π and o be the object between r and t. Let X be the set of elements of the local monoid on o that can be achieved as some loop q_x of frequent

arrows raised to the power ω: formally, X is the set of elements $x \in S$ such that there is a loop q_x of frequent arrows such that q_x^ω evaluates to (o, x, o), noting that this implies that x is idempotent. Then letting $q := \prod_{x \in X} q_x^\omega$, we have that $\pi \equiv rqt$. (Note that these two paths are also n', ω-equivalent by construction.)

Intuitively, in this claim, X stands for the set of elements of S that we can achieve as a loop on o raised at the power ω and using frequent arrows only. The claim states that inserting loops of this form at o to achieve all such elements x will preserve equivalence.

We first explain why Claim 7.3 implies Lemma 7.2. Indeed, when taking $p = rt$, letting o be the object between r and t, and taking $r(\pi')^\omega t$, the sets X defined in Claim 7.3 after the occurrence of r in $p = rt$ and $r(\pi')^\omega t$ will be the same for both paths (because X only depends on o). Thus, Claim 7.3 implies that there is a loop q such that $rt \equiv rqt$ and $r(\pi')^\omega t \equiv rq(\pi')^\omega t$. Now, as $(\pi')^\omega$ must correspond to an arrow of the form (o, x, o) for $x \in X$, it must be the same idempotent as one of the idempotents achieved by one of the loops in the definition of q, and as the local monoid is in **ZG** these idempotents commute and $q \equiv q(\pi')^\omega$. Hence, we have $rqt \equiv rq(\pi')^\omega t$. We know that $rqt \equiv rt$, and $rq(\pi')^\omega t \equiv r(\pi')^\omega t$. Thus we obtain $rt \equiv r(\pi')^\omega t$. Thus, Lemma 7.2 is proved once we have shown Claim 7.3.

Hence, all that remains is to show Claim 7.3. We will do so by establishing a number of claims.

We first show that, for any frequent arrow x, we can insert some loop of the form xu where u is a return path using only frequent arrows, while preserving equivalence. This uses the notion of distant rare-frequent threshold; specifically, this is where we perform the pumping made possible by Claim 6.6.

Claim 7.4. Let $\pi = rt$ be a path with an $|S|$-distant rare-frequent threshold, let o be the object between r and t, and let x be any frequent arrow starting at o. Then we have $\pi \equiv r(xu)^\omega t$ for some return path u using only frequent arrows.

Proof. As x is a frequent arrow and π has an $|S|$-distant threshold, we know by Claim 6.6 that π contains a factor ρ that contains only frequent arrows and contains $k > |S|$ occurrences of x. This provides a decomposition of ρ in the form: $\rho = \rho_1 x \rho_2 x \cdots \rho_k x s$.

By the pigeonhole principle, there exists $i < j$ such that $\rho_1 x \cdots \rho_i x \equiv \rho_1 x \cdots \rho_j x$. Hence, iterating, we obtain:

$$\rho_1 x \cdots \rho_j x \equiv \rho_1 x \cdots \rho_i x (\rho_{i+1} x \cdots \rho_j x)^\omega.$$

Moving the ω, we get:

$$\rho_1 x \cdots \rho_j x \equiv \rho_1 x \cdots \rho_{i-1} x \rho_i (x \rho_{i+1} x \cdots \rho_j)^\omega x.$$

This proves that ρ and $h(xu)^\omega g$ achieve the same category element when taking $u := \rho_{i+1} x \cdots \rho_j$, $h := \rho_1 x \cdots \rho_{i-1} x \rho_i$ and $g := x \rho_{j+1} x \cdots \rho_k x s$. By compositionality, π and $h'(xu)^\omega g'$ achieve the same category element, where h' is the part of π preceding ρ followed by h, and g' is g followed by the part of π following ρ.

Now, recall that we must show the result for our decomposition $\pi = rt$, where the object o between r and t is the starting object of the arrow x and the ending object of h. Either h' is a prefix of r, or vice-versa. Assume first that we are in the first case, so $h' = rw$ for some path w. Then, w and $(xu)^\omega$ belong to the local monoid of o which is in **ZG**. Since idempotents commute with all elements, we have $w(xu)^\omega \equiv (xu)^\omega w$ establishing that $rw(xu)^\omega g' \equiv r(xu)^\omega wg' \equiv r(xu)^\omega t$ since $wg' = t$. The other case is symmetrical. This concludes the proof of Claim 7.4. □

We then prove a generalization of the previous claim, going from a single frequent arrow to an arbitrary path of frequent arrows:

Claim 7.5. Let $\pi = rt$ be a path with an $|S|$-distant rare-frequent threshold. Let o be the object between r and t. Let h be any path starting at o which only uses frequent arrows. Then we have $\pi \equiv r(hg)^\omega t$ for some return path g using only frequent arrows.

Proof. We show the claim by induction on the length of h. The base case of the induction, with h of length 0, is trivial with g also having length 0.

For the inductive claim, write $h = h'a$. Intuitively, we will insert a loop starting with the path h', then insert a loop starting with the arrow a within that loop, and then recombine.

Formally, by induction hypothesis, there exists a g' using only frequent arrows such that:

$$\pi \equiv r(h'g')^\omega t.$$

Furthermore, by applying Claim 7.4 to the decomposition $r' = rh'$ and $t' = g'(h'g')^{\omega-1}t$ and with the frequent arrow a we get a return path u using only frequent arrows such that:

$$r(h'g')^\omega t \equiv rh'(au)^\omega g'(h'g')^{\omega-1}t$$

So, iterating the ω power, and combining with the preceding equation, we get:

$$\pi \equiv rh'((au)^\omega)^\omega g'(h'g')^{\omega-1}t.$$

Now, by applying $\omega - 1$ times Claim 7.1 to each $(au)^\omega$ except the first and to each loop going from after this $(au)^\omega$ to the position between an occurrence of h' and g', we get that:

$$\pi \equiv rh'(au)^\omega g'(h'(au)^\omega g')^{\omega-1}t.$$

Note the right-hand side is equal to: $r(h'(au)^\omega g')^\omega t$. So we have shown:

$$\pi \equiv r(h'(au)^\omega g')^\omega t.$$

So this establishes the inductive claim by taking $g := u(au)^{\omega-1}g'$. $\qquad\square$

We are interested in the specialization of this result when the path of frequent arrows to insert is a loop. In this case, the return path is also a loop. Formally, the specialization is the following:

Corollary 7.6. *Let $\pi = rt$ be a path with an $|S|$-distant rare-frequent threshold n', let o be the object between r and t, and let q be a loop on o using only frequent arrows. We have that $rt \equiv rq^\omega (q')^\omega t$ for some loop q' on o using only frequent arrows (note that the two are also n', ω-equivalent).*

Proof. We use Claim 7.5 with $h := q$. This gives us the existence of a return path g using only frequent arrows, which is then also a loop on o, such that $rt \equiv r(qg)^\omega t$. Now, applying Lemma 3.8 to the local monoid on object o, we know that this evaluates to the same category element as $rq^\omega g^\omega t$. Taking $q' := g$ concludes the proof of Corollary 7.6. $\qquad\square$

The only step left is to argue that Corollary 7.6 implies our rephrasing (Claim 7.3) of the loop insertion lemma (Lemma 7.2). To do this, let $\pi = rt$ be the path, let o be the object between r and t, and let X be the set of idempotents definable from frequent arrows. Let us write $X = \{x_1, \ldots, x_k\}$ with $k = |X|$. For each $1 \le i \le k$, take a loop q_i on o consisting only of frequent arrows that achieves x_i, i.e., which ensures that q_i^ω evaluates to (o, x_i, o). Now, we apply Corollary 7.6 k times with q being each of the loops $q_1, \ldots, q_k$. We get the

following, where $q'_1, \ldots, q'_k$ are the loops q' on o of frequent arrows obtained by the statement of Corollary 7.6:

$$rt \equiv rq_1^\omega (q'_1)^\omega \cdots q_k^\omega (q'_k)^\omega t$$

Note that the left-hand-side and right-hand-side are also n', ω-equivalent.

Now, since the q'_i are loops on o consisting of frequent arrows, each $(q'_i)^\omega$ evaluates to (o, x, o) for some $x \in X$. As the elements of X are in the local monoid of o which is in **ZG**, commuting the loops using Claim 7.1, we can combine each $(q'_i)^\omega$ with some q_j^ω such that $(q'_i)^\omega \equiv q_j^\omega$, hence $(q'_i)^\omega q_j^\omega = q_j^\omega$. Thus, we get that rt is n', ω-equivalent to, and evaluates to the same category element as, the path:

$$r \prod_{x \in X} q_x^\omega t.$$

This concludes the proof of Claim 7.3, and thus establishes our desired result, Lemma 7.2.

7.2. **Prefix substitution lemma.**
The prefix substitution lemma allows us to change any prefix of frequent arrows of a path, up to inserting a loop of frequent arrows elsewhere:

Lemma 7.7. *Let $\pi = xry$ be a path, and assume that n' is an $|S|$-distant rare-frequent threshold for π. Let x' be a path coterminal with x. Assume that every arrow in x and in x' is frequent. Assume that some object o in the SCC of frequent arrows of the starting object of r occurs again in y, say as the intermediate object of $y = y_1 y_2$. Then there exists $y' = y_1 y'' y_2$ for some loop y'' consisting only of frequent arrows such that $\pi \equiv x'ry'$ and such that π and $x'ry'$ are n', ω-equivalent.*

This lemma uses Claim 6.8 to argue that frequent arrows are a union of SCCs. Its proof relies on the loop insertion lemma (Lemma 7.2), but with extra technical work using the **ZG** equation.

To prove Lemma 7.7, we will first show that frequent loops can be "recombined" without changing the category image, simply by equation manipulation:

Claim 7.8. *For x, x' two coterminal paths in S_E and y, y' coterminal paths in S_E such that xy and $x'y'$ are valid loops, we have: $(xy)^\omega (x'y')^\omega \equiv (xy')^\omega (x'y)^\omega (xy)^\omega (x'y')^\omega$.*

Proof. Let us first show that:

$$(xy)^\omega (x'y')^\omega \equiv x'y(xy)^{\omega-1}(x'y')^{\omega-1}xy' \tag{7.1}$$

To show Equation 7.1, first rewrite $(xy)^\omega$ as $x(yx)^{\omega-1}y$ and likewise for $(y'x')^\omega$, to get:

$$(xy)^\omega (x'y')^\omega \equiv x(yx)^{\omega-1}yx'(y'x')^{\omega-1}y'$$

Then, we use Claim 7.1 to move $(yx)^{\omega-1}$, so the above evaluates to the same category element as:

$$xyx'(yx)^{\omega-1}(y'x')^{\omega-1}y'$$

We again rewrite $(yx)^{\omega-1}$ to $y(xy)^{\omega-2}x$, yielding:

$$xyx'y(xy)^{\omega-2}x(y'x')^{\omega-1}y'$$

We again use Claim 7.1 to move $(xy)^{\omega-2}$, merge it with the prefix xy, and move it back to its place, yielding:

$$x'y(xy)^{\omega-1}x(y'x')^{\omega-1}y'$$

We rewrite $(y'x')^{\omega-1}$ to $y'(x'y')^{\omega-2}x'$, yielding:

$$x'y(xy)^{\omega-1}xy'(x'y')^{\omega-2}x'y'$$

Again by Claim 7.1, we can merge $(x'y')^{\omega-1}$ with $x'y'$ and move it to finally get:

$$x'y(xy)^{\omega-1}(x'y')^{\omega-1}xy'$$

Thus, the left-hand side of Equation 7.1 evaluates to the same category element as the right-hand side, and we have shown Equation 7.1.

Now, we have $(xy)^{\omega}(x'y')^{\omega} \equiv (xy)^{2\omega}(x'y')^{2\omega}$, so using Claim 7.1 again, Equation 7.1 gives:

$$(xy)^{\omega}(x'y')^{\omega} \equiv x'y(xy)^{\omega}(x'y')^{\omega}(xy)^{\omega-1}(x'y')^{\omega-1}xy'$$

We can now use Equation 7.1 to replace $(xy)^{\omega}(x'y')^{\omega}$ by the right-hand side of Equation 7.1 and use Claim 7.1 to commute, yielding:

$$(xy)^{\omega}(x'y')^{\omega} \equiv (x'y)^2(xy)^{\omega-2}(x'y')^{\omega-2}(xy')^2$$

By definition we have $(xy)^{\omega-2} \equiv (xy)^{\omega-2}(xy)^{\omega}$ and $(x'y')^{\omega-2} \equiv (x'y')^{\omega}(x'y')^{\omega-2}$. Injecting these in the equation above, we get:

$$(xy)^{\omega}(x'y')^{\omega} \equiv (x'y)^2(xy)^{\omega-2}\underbrace{(xy)^{\omega}(x'y')^{\omega}}_{(\star)}(x'y')^{\omega-2}(xy')^2$$

Note that $(\star)$ is now equal to the left-hand-side of the equation. Substituting ω times the right-hand-side into $(\star)$, we obtain:

$$(xy)^{\omega}(x'y')^{\omega} \equiv (x'y)^{\omega}(xy)^{\omega}(x'y')^{\omega}(xy')^{\omega}$$

As these elements commute (thanks to Claim 7.1), we have shown the desired equality. $\square$

We will now extend this result to show that we can change the initial part of a path, even if it is not a loop, provided that there is a coterminal path under an ω-power with which we can swap it.

Claim 7.9. For x, x' two coterminal paths in S_{E} and y, y' coterminal paths in S_{E} such that xy and $x'y'$ are valid loops, and for any path t coterminal with y, the following equation holds: $xt(xy)^{\omega}(x'y')^{\omega} \equiv x't(xy)^{\omega}xy'(x'y')^{\omega-1}$.

We establish this again by equation manipulation.

Proof of Claim 7.9. We apply Claim 7.8 to show the following equality about the left-hand side:

$$xt(xy)^{\omega}(x'y')^{\omega} \equiv xt(xy')^{\omega}(x'y)^{\omega}(xy)^{\omega}(x'y')^{\omega}$$

By commutation of $(x'y)^{\omega}$ thanks to Claim 7.1, the right-hand side evaluates to the same category element as:

$$(x'y)^{\omega}xt(xy')^{\omega}(x'y')^{\omega}(xy)^{\omega}$$

By expanding $(x'y)^{\omega} = x'(yx')^{\omega-1}y$, we get:

$$x'(yx')^{\omega-1}yxt(xy')^{\omega}(x'y')^{\omega}(xy)^{\omega}$$

By commutation of $(xy)^{\omega}$ and expanding it to $x(yx)^{\omega-1}y$, we get:

$$x'(yx')^{\omega-1}yx(yx)^{\omega-1}yxt(xy')^{\omega}(x'y')^{\omega}$$

Combining $(yx)^{\omega-1}$ with what precedes and follows, we get:

$$x'(yx')^{\omega-1}(yx)^{\omega+1}t(xy')^{\omega}(x'y')^{\omega}$$

By expanding $(x'y')^\omega = x'(y'x')^{\omega-1}y'$, and commuting $(yx')^{\omega-1}$ and $(yx)^{\omega+1}$, we get:

$$x't(xy')^\omega x'(yx')^{\omega-1}(yx)^{\omega+1}(y'x')^{\omega-1}y'$$

Now, we have $x'(yx')^{\omega-1} = (x'y)^{\omega-1}x'$, so we get:

$$x't(xy')^\omega(x'y)^{\omega-1}x'(yx)^{\omega+1}(y'x')^{\omega-1}y'$$

Commuting $(y'x')^{\omega-1}$ and doing a similar transformation, we get:

$$x't(xy')^\omega(x'y)^{\omega-1}(x'y')^{\omega-1}x'(yx)^{\omega+1}y'$$

Now, expanding $(yx)^{\omega+1}$, we get:

$$x't(xy')^\omega(x'y)^{\omega-1}(x'y')^{\omega-1}x'y(xy)^\omega xy'$$

Commuting $(x'y)^{\omega-1}$ and merging it with $x'y$, we get:

$$x't(xy')^\omega(x'y')^{\omega-1}(x'y)^\omega(xy)^\omega xy'$$

Note that $(x'y')^{\omega-1} \equiv (x'y')^\omega(x'y')^{\omega-1}$, so applying commutation we get:

$$x't(xy')^\omega(x'y')^\omega(x'y)^\omega(xy)^\omega(x'y')^{\omega-1}xy'$$

Now, applying Claim 7.8 in reverse (using commutation again), we can obtain:

$$x't(x'y')^\omega(xy)^\omega(x'y')^{\omega-1}xy'$$

Then commuting $(x'y')^\omega$ and merging it yields:

$$x't(xy)^\omega(x'y')^{\omega-1}xy'$$

A final commutation of $(x'y')^{\omega-1}$ yields the desired right-hand side, and we have preserved equivalence in the category of idempotents, establishing the result. □

With Claim 7.9 in hand, and using Lemma 7.2, we can now prove Lemma 7.7:

Proof of Lemma 7.7. As n' is an $|S|$-distant rare-frequent threshold, it is in particular a 1-distant rare-frequent threshold, so we know by Claim 6.8 that the frequent arrows occurring in π are a union of SCCs. Thus, there is a return path s for x' (i.e., $x's$ is a loop, hence xs also is) where s only consists of frequent arrows.

By our hypothesis on the starting object of r, we can decompose $y = y_1y_2$ such that the ending object of y_1 and starting object of y_2 is an object o which is in the SCC of frequent arrows of the starting object o' of r. Let ρ_1 be any path of frequent arrows from o to o', and ρ_2 be any path of frequent arrows from o' to o. Now, take π' to be the loop $\rho_1 s(xs)^\omega(x's)^\omega x\rho_2$: note that all arrows of π' are frequent. Hence, by Lemma 7.2, xry evaluates to the same category element as, and is n', ω-equivalent to,

$$xry_1(\pi')^\omega y_2 = xry_1(\rho_1 s(xs)^\omega(x's)^\omega x\rho_2)^\omega y_2$$

By unfolding the power ω, we get the following:

$$xry_1(\pi')^\omega y_2 = xry_1\rho_1 s(xs)^\omega(x's)^\omega x\rho_2(\rho_1 s(xs)^\omega(x's)^\omega x\rho_2)^{\omega-1}y_2 = xry_1\rho_1 s(xs)^\omega(x's)^\omega xz$$

where we write $z := \rho_2(s(xs)^\omega(x's)^\omega x)^{\omega-1}y_2$ for convenience. We can therefore apply Claim 7.9 to obtain that:

$$(x(ry_1\rho_1 s))(xs)^\omega(x's)^\omega xz \equiv (x'(ry_1\rho_1 s))(xs)^\omega(xs)(x's)^{\omega-1}xz.$$

What is more, these two paths are clearly n', ω-equivalent, as they only differ in terms of frequent arrows (all arrows in x and x' being frequent) and the number of these arrows modulo ω is unchanged by the transformation. This path is of the form given in the

statement, taking $y' := y_1\rho_1 s(xs)^\omega(xs)(x's)^{\omega-1}xz$ from which we can extract the right y''. This concludes the proof. $\square$

8. Concluding the Proof of the Main Result (Theorem 1.1)

We are now ready to prove the second direction of Theorem 1.1, namely Claim 5.7. Let us fix the semigroup S in $\mathbf{LZG}_p$, write S_E its category of idempotents, and write B for the set of arrows of S_E. Let p' be the period of S: we know that p' divides the idempotent power ω of S. Further, p' divides p: this is because, for any group element x of S, we have that x belongs to the local monoid $x^\omega S x^\omega$, hence it belongs to $\mathbf{ZG}_p$ and its period divides p, thus p' is a multiple of the periods of all group elements, hence it divides p.

Let n be given by Claim 6.5. Our goal is to show that the n, p-congruence on B^* is compatible with S_E. We will in fact show the same for the n, p'-congruence, which is coarser than the n, p-congruence because p' divides p (using Claim 3.2), so suffices to establish the result.

To do so, we will show that two coterminal paths that are n, p'-equivalent must evaluate to the same category element, by two nested inductions. We first explain the outer induction, which is on the number of rare arrows in the paths, for an $|S|$-distant rare-frequent threshold $n' \leq n$ chosen from the two paths that we consider (via Claim 6.5). The threshold n' is not fixed from the beginning but it depends on the two paths considered, which is why we will not fix it in the inductive claim. However, we will only apply the claim to one value of n' chosen from the initial paths, i.e., it will not change during the induction.

Formally, we show by induction on the integer r the following:

Claim 8.1 (Outer inductive claim on r). For any two paths π_1 and π_2, for any $|S|$-distant rare-frequent threshold n' for π_1 and π_2, if π_1 and π_2 are n', p'-equivalent and contain r rare arrows each, then we have $\pi_1 \equiv \pi_2$.

Once we have established this, we can conclude the proof of Claim 5.7. To do so, take any two coterminal paths π_1 and π_2 that are n, p'-equivalent. By Claim 6.5, we can pick a threshold n' (depending on u_1, u_2, and n) such that $0 < n' \leq n$ which is an $|S|$-distant rare-frequent threshold for π_1 and π_2. By Claim 3.2, as π_1 and π_2 are n, p'-equivalent, we know that they are also n', p'-equivalent. Now that we have fixed the threshold n', following Definition 3.1, we call an arrow of B *rare* in u_1 and u_2 if it occurs $\leq n'$ times in each, and *frequent* otherwise. Let r_0 be the number of rare arrows in u_1 and u_2: this number is the same for both, because they are n', p'-equivalent. We now apply the outer inductive claim (Claim 8.1) on π_1, π_2, n', and $r = r_0$, to conclude that $\pi_1 \equiv \pi_2$, which concludes the proof of Claim 5.7.

It remains to show Claim 8.1, which we do in the rest of the section.

8.1. Outer base case (B): all arrows in π_1 and π_2 are frequent. We now show Claim 8.1. The base case of the outer induction is for paths that contain no rare arrows:

Claim 8.2. For any two paths π_1 and π_2, for any $|S|$-distant rare-frequent threshold n' for π_1 and π_2, if π_1 and π_2 are n', p'-equivalent and contain no rare arrows, then we have $\pi_1 \equiv \pi_2$.

We prove this result by an induction over the number of different arrows that occur in the paths π_1 and π_2, noting that thanks to n', p'-equivalence an arrow occurs in π_1 iff it occurs in π_2. Specifically, we show by induction on the integer f the following:

Claim 8.3 (Inner inductive claim on f). For any two paths π_1 and π_2, for any $|S|$-distant rare-frequent threshold n' for π_1 and π_2, if π_1 and π_2 are n', p'-equivalent and contain no rare arrows and there are $\leq f$ different frequent arrows that occur, then we have $\pi_1 \equiv \pi_2$.

Once we have established this, we can conclude the proof of the base case of the outer induction, Claim 8.2. To do so, take any two coterminal paths π_1 and π_2 and an $|S|$-distant rare-frequent threshold n' such that π_1 and π_2 are n', p'-equivalent, let f_0 be the number of frequent arrows that occur in π_1 and in π_2, and conclude using Claim 8.3 with π_1, π_2, n', and $f = f_0$.

We now explain the proof of the inner inductive claim (Claim 8.3), before moving on to the inductive case of the outer induction. The base case of Claim 8.3 with $f = 0$ is trivial as π_1 and π_2 are then empty. For the induction step, assume that the claim holds for any π_1, π_2, and n' such that there are $\leq f$ different frequent arrows that occur. Fix π_1, π_2, and n' where there are $\leq f + 1$ frequent arrows that occur. Consider the multigraph G of all arrows of S_{E} that occur in π_1 and π_2: G has $\leq f + 1$ edges. If it has $\leq f$ edges we immediately conclude by induction hypothesis, so assume it has exactly $f + 1$ edges. Recall that G is strongly connected: indeed, as all arrows are frequent, thanks to the existence of the $|S|$-distant rare-frequent threshold n' (which is in particular 1-distant), we know by Claim 6.8 that G is a union of SCCs. Further, π_1 (or π_2) is a path where all edges of G occur, so it witnesses that G is connected, and G is therefore strongly connected. The induction case is shown using a so-called *ear decomposition* result on strongly connected multigraphs:

Lemma 8.4 [BJG08]. *Let G be a strongly connected nonempty directed multigraph. We have:*

- *G is a simple cycle; or*
- *G contains a simple cycle $v_1 \to \cdots \to v_n \to v_1$ with $n \geq 1$, where all vertices $v_1, \ldots, v_n$ are pairwise distinct, such that all intermediate vertices $v_2, \ldots, v_{n-1}$ only occur in the edges of the cycle, and such that the removal of the cycle leaves the graph strongly connected (note that the case $n = 1$ corresponds to the removal of a self-loop); or*
- *G contains a simple path $v_1 \to \cdots \to v_n$ with $n \geq 2$ where all vertices are pairwise distinct, such that all intermediate vertices $v_2, \ldots, v_{n-1}$ only occur in the edges of the path, and such that the removal of the path leaves the graph strongly connected (note that the case $n = 2$ corresponds to the removal of a single edge).*

This is a known result [BJG08], but we give a self-contained proof in Appendix A for the reader's convenience.

Thanks to this result, we can distinguish three cases in the inner induction step: if G is non-empty, it must be a simple cycle, contain a removable simple cycle, or contain a removable simple path. We first give a high-level view of the argument in each case:

Proof sketch. The first case (B.1) is when G is a simple cycle. In this case n', p'-equivalence ensures that the cycle is taken by π_1 and π_2 some number of times with the same remainder modulo p', so they evaluate to the same element because p' is a multiple of the period.

The second case (B.2) is when G contains a removable simple cycle. This time, we argue as in the previous case that the number of occurrences of the cycle must have the same remainder, and we can use Corollary 3.7 to merge all the occurrences together. However, to

eliminate them, we need to use Lemma 7.7, to modify π_1 and π_2 to have the same prefix (up to and including the cycle occurrences), while preserving equivalence. This allows us to consider the rest of the paths (which contains no occurrence of the cycle), apply the induction hypothesis to them, and conclude by compositionality. A technicality is that we must ensure that removing the common prefix does not make some arrows insufficiently frequent relative to the distant rare-frequent threshold. We avoid this using Lemma 7.2 to insert sufficiently many copies of a suitable loop.

The third case (B.3) is when G contains a removable simple path τ. The reasoning is similar, but we also use Lemma 7.2 to insert a loop involving a return path for τ and a path that is parallel to τ (i.e., does not share any arrows with it). The return path in this loop can then be combined with τ to form a loop, which we handle like in the previous case. $\square$

We now give the detailed argument for each case.

Case B.1: G is a simple cycle. If G is a simple cycle, then distinguish the starting object of π_1 (hence, of π_2) as o, and let α be the cycle from o to itself, and τ the path from o to the common ending object of π_1 and π_2. We have: $\pi_1 = \alpha^{n_1}\tau$ and $\pi_2 = \alpha^{n_2}\tau$ with n_1 and n_2 being $\geq n'-1$ and having the same remainder r modulo p'. We use the loop insertion lemma (Lemma 7.2) to insert α^ω: the lemma tells us that π_1 and π_2 respectively evaluate to the same category element as $\pi_1' = \alpha^{\omega+n_1}$ and $\pi_2' = \alpha^{\omega+n_2}$. Further, π_1 and π_1', and π_2 and π_2', are n', p'-equivalent.

As α is a loop on the idempotent o, we know that there exists an element $m \in oSo$ such that $\alpha \equiv (o, m, o)$. Hence, $\pi_1' = \alpha^{\omega+n_1}\tau \equiv (o, m^{\omega+n_1}, o)\tau$ (resp. $\pi_2' = \alpha^{\omega+n_2}\tau \equiv (o, m^{\omega+n_2}, o)\tau$). Now, we have $m^{\omega+n_1} = m^{\omega+k_1 \cdot p'+r}$ and $m^{\omega+n_2} = m^{\omega+k_2 \cdot p'+r}$ where k_1 and k_2 are the respective quotients of n_1 and n_2 in the Euclidean division by p' and where r is the common remainder modulo p'. By definition of the period p', we have $m^{\omega+n_1} = m^{\omega+r}$ and $m^{\omega+n_2} = m^{\omega+r}$. We conclude that $m^{\omega+n_1} = m^{\omega+n_2}$, so that $\pi_1' \equiv \pi_2'$, and $\pi_1 \equiv \pi_2$. This concludes case B.1.

Case B.2: G has a simple cycle. Recall that, in this case, we know that G has a simple cycle whose intermediate objects have no other incident edges and such that the removal of the simple cycle leaves the graph strongly connected. Let α be the simple cycle, starting from the only object o of the cycle having other incident edges. We can then decompose π_1 and π_2 to isolate the occurrences of the simple cycle (which must be taken in its entirety), i.e.:

$$\pi_1 = x_1\alpha x_2\alpha x_3 \cdots x_{t-1}\alpha x_t$$
$$\pi_2 = y_1\alpha y_2\alpha y_3 \cdots y_{t'-1}\alpha y_{t'}$$

This ensures that the edges of α do not occur elsewhere than in the α factors, except possibly in x_1, y_1 and in x_t, $y_{t'}$ if the paths π_1 and/or π_2 start and/or end in the simple cycle. This being said, in that case, we know that the prefixes of π_1 and π_2 containing this incomplete subset of the cycle must be equal (same sequence of arrows), and likewise for their suffixes. For this reason, it suffices to show the claim that π_1 and π_2 evaluate to the same category element under the assumption that both their starting and ending objects are not intermediate vertices of the cycle. The claim then extends to the general case, by adding the common prefixes and suffixes to the two paths that satisfy the condition, using compositionality of the congruence. Thus, in the rest of the proof for this case, we assume that the edges of α only occur in the α factors.

We will now argue that, to show that $\pi_1 \equiv \pi_2$, it suffices to show the same of two n', p'-equivalent coterminal paths from which all occurrences of the edges of the cycle have been removed and where all other edges still occur sufficiently many times. As this deals with paths where the underlying multigraph contains fewer edges, the induction hypothesis will conclude.

To do this, by Lemma 7.7, as x_1 and y_1 are coterminal and consist only of frequent arrows, and as the starting object of α occurs again in both paths, the path π_1 is n', ω-equivalent, and evaluates to the same category element as, some path:

$$\pi'_1 = y_1 \alpha x'_2 \alpha x'_3 \cdots x'_{t''-1} \alpha x'_{t''}$$

The above is also n', p'-equivalent to π_1, because p' divides ω. Thus, up to replacing π_1 by π'_1, we can assume that $x_1 = y_1$.

Now, furthermore, $x_2, \ldots, x_{t-1}$ (resp. $y_2, \ldots, y_{t'-1}$) and α are coterminal cycles over the object o (which by definition corresponds to an idempotent of S). Hence, $\alpha x_2 \alpha x_3 \cdots x_{t-1} \alpha \equiv (o, mm_2 mm_3 \cdots m_{t-1} m, o)$ where $\alpha \equiv (o, m, o)$, $x_i \equiv (o, m_i, o)$ for $2 \leq i \leq t-1$ and where m and all m_i's are in oSo, which is by hypothesis a monoid in **ZG**. Now, as the arrows of α are frequent, each one of them must occur $> n'$ times, so α (which contains exactly one occurrence of each of these arrows) must occur $> n'$ times, and as n' is a $|S|$-distant rare-frequent threshold we clearly have by Definition 6.1 that $n' \geq |S|$, so α occurs $\geq |S| + 1$ times. Hence, by Corollary 3.7, we know that $mm_2 mm_3 \cdots m_{t-1} m = m^{t-1} m_2 m_3 \cdots m_{t-1}$, because $|S| + 1 \geq |oSo| + 1$. By applying the same reasoning to π_2, it suffices to show that the two following paths evaluate to the same category element, where $x_1 = y_1$:

$$x_1 \alpha^{t-1} x_2 x_3 \cdots x_{t-1} x_t$$
$$y_1 \alpha^{t'-1} y_2 y_3 \cdots y_{t'-1} y_{t'}$$

Now, because these two paths are n', p'-equivalent, we know that $t-1$ and $t'-1$ have the same remainder modulo p'. By the same reasoning as in case B.1, they evaluate to the same category element as α^r, where r is the remainder. So it suffices to show that the two following paths evaluate to the same category element, with $x_1 = y_1$:

$$x_1 \alpha^r x_2 x_3 \cdots x_{t-1} x_t$$
$$y_1 \alpha^r y_2 y_3 \cdots y_{t'-1} y_{t'}$$

We now intend to use the induction hypothesis, but for this, we need to ensure that n' is still an $|S|$-distant rare-frequent threshold on the paths to which we apply it. Specifically, we need to ensure that the edges not in α still occur sufficiently many times. To this end, let β be any loop on o that visits all edges of G except the ones in α: this is doable because G is still strongly connected after the removal of α. Up to exponentiating β to some power $\beta^{n''\omega}$, we can assume that β traverses each edge sufficiently many times to satisfy the lower bound imposed by the requirement of n' being an $|S|$-distant rare-frequent threshold. By Lemma 7.2, it suffices to show that the following paths evaluate to the same category element:

$$\pi'_1 = x_1 \alpha^r \beta^{n''\omega} x_2 x_3 \cdots x_{t-1} x_t$$
$$\pi'_2 = y_1 \alpha^r \beta^{n''\omega} y_2 y_3 \cdots y_{t'-1} y_{t'}$$

Note that n', p'-equivalence is preserved because n'' is a multiple of ω, hence of p'. Now, observe that both paths start by $x_1 \alpha = y_1 \alpha$, and the arrows of α do not occur in the rest of the paths. Consider the paths $\beta^{n''\omega} x_2 x_3 \ldots x_t$ and $\beta^{n''\omega} y_2 y_3 \ldots y_{t'}$. They are paths that

are coterminal, n', p'-equivalent because π_1 and π_2 were, where the frequent letters that are used are a strict subset of the ones used in π_1 and π_2, and where all other frequent letters occur sufficiently many times for n' to still be an $|S|$-distant rare-frequent threshold (as guaranteed by $\beta^{n''\omega}$). Thus, by induction hypothesis of the inner induction, we know that these two paths evaluate to the same category element, so that π_1' and π_2' also do. This concludes case B.2.

Case B.3: G **has a simple path.** Recall that, in this case, we know that G has a simple path where the starting and ending objects of intermediate arrows have no other incident edges, and such that the removal of the simple path leaves the graph strongly connected. We denote the path by τ and denote by $x \neq y$ its starting and ending objects. Since the removal of the path does not affect strong connectedness of the graph, there is a simple path from x to y sharing no edges with τ, which we denote by κ. Furthermore, there is a simple path from y to x sharing no edges with τ (this is because all intermediate objects of τ only occur in the edges of τ), which we denote by ρ.

Like in the previous case, up to removing common prefixes and suffixes, it suffices to consider the case where π_1 and π_2 do not start or end in the intermediate vertices of τ. For that reason, isolating all occurrences of τ also isolates all occurrences of the edges of τ, and we can write:

$$\pi_1 = x_1 \tau x_2 \tau x_3 \cdots x_{t-1} \tau x_t$$
$$\pi_2 = y_1 \tau y_2 \tau y_3 \cdots y_{t'-1} \tau y_{t'}$$

where the x_i and y_i do not use the edges of τ. Like in the previous case, by Lemma 7.7, we can assume that $x_1 = y_1$.

By Lemma 7.2, we insert a loop $(\rho\kappa)^\omega$ after every occurrence of τ without changing the category element and still respecting the n', ω-congruence, hence the n', p'-congruence because p' divides ω. By expanding $(\rho\kappa)^\omega = \rho\kappa(\rho\kappa)^{\omega-1}$, it suffices to show that the following paths evaluate to the same category element, with $x_1 = y_1$:

$$\pi_1' = x_1 \tau\rho\kappa(\rho\kappa)^{\omega-1} \cdots x_{t-1} \tau\rho\kappa(\rho\kappa)^{\omega-1} x_t$$
$$\pi_2' = y_1 \tau\rho\kappa(\rho\kappa)^{\omega-1} \cdots y_{t'-1} \tau\rho\kappa(\rho\kappa)^{\omega-1} y_{t'}$$

We can now regroup the occurrences of $\tau\rho$, which are loops such that some edges (namely, the edges of τ) only occur in these factors. This means that we can conclude as in case 2 for the cycle $\tau\rho$, as this cycle contains some edges that only occur there; we can choose β at the end of the proof to be a loop on x visiting all edges of G except those of τ, which is again possible because G is still strongly connected even after the removal of τ.

This establishes case 3 and concludes the induction step of the proof, establishing Claim 8.3.

We have thus proved by induction that π_1 and π_2 evaluate to the same category element, in the base case (Claim 8.2) of the outer induction (Claim 8.1) where all edges of π_1 and π_2 are frequent.

8.2. **Outer induction step (I): some arrows are rare.** Let us now show the induction step for the outer induction (Claim 8.1), namely, the induction on the number of occurrences of rare arrows. We assume the claim of the outer induction for $r \in \mathbb{N}$. Consider two paths π_1 and π_2 and an $|S|$-distant rare-frequent threshold n' such that π_1 and π_2 are n', p'-equivalent and such that they contain $r + 1$ rare letters. Let us partition them as $\pi_1 = q_1 a s_1$ and

$\pi_2 = q_2 a s_2$ where q_1 and q_2 all consist of frequent arrows, and a is the first rare arrow of π_1 and π_2 (note that n', p'-equivalence implies that the first rare arrow is the same in both paths). In this case, q_1 and q_2 are two coterminal paths consisting only of frequent arrows (or they are empty), and s_1 and s_2 are two coterminal paths (possibly empty) with r rare letter occurrences.

Remember that, as n' is an $|S|$-distant rare-frequent threshold for π_1 and π_2, then we know that the frequent arrows of π_1 form a union of SCCs (Claim 6.8); note that, thanks to n', p'-equivalence, the same is true of π_2 with the same SCCs. Consider the SCC C of frequent arrows that contains the starting object of a. There are two cases, depending on whether some object of C occurs again in s_1 or not. Note that some object of C occurs again in s_1 iff the same is true of s_2, because which frequent arrow components occur again is entirely determined by the ending objects of the rare arrows of s_1 and s_2, which are identical thanks to n', p'-equivalence.

Case I.1: C occurs again after a. In this case, we are in a situation where we can apply Lemma 7.7, because q_1 and q_2 only consist of frequent arrows and some object of the SCC C of the starting object of a occurs again in s_1. The lemma tells us that there is a path:

$$\pi_1' = q_2 a s_1'$$

which evaluates to the same category element as π_1 and is n', ω-equivalent to it, hence n', p'-equivalent. Hence, by compositionality, it suffices to show that s_1' and s_2 evaluate to the same category element.

To apply the induction hypothesis, we simply need to ensure that n' is still a $|S|$-distant rare-frequent threshold for s_1' and s_2. To do this, we need to ensure that the arrows that are frequent in π_1 and π_2 are still frequent there, and still satisfy the $|S|$-distant condition. Fortunately, we can simply ensure this by inserting a loop using Lemma 7.2. Formally, write $s_1' = r_1 t_1$ where the intermediate object is the object of the SCC C that occurred in s_1 (the existence of such a decomposition is a consequence of the statement of Lemma 7.7), and write $s_2 = r_2 t_2$ in the same way (which we already discussed must be possible with s_2). Let β be an arbitrary loop of frequent arrows where all arrows of C occur: this is possible because C is strongly connected. We know by Lemma 7.2 that $s_1' = r_1 t_1$ and $r_1 \beta^\omega t_1$ are both n', ω-equivalent, hence n', p'-equivalent, and evaluate to the same category element: this is also true with $w_1 := r_1 \beta^{n''\omega} t_1$ for a sufficiently large n'' such that every frequent arrow of C occurs as many times as it did in π_1. Likewise, $s_2 = r_2 t_2$ and $w_2 := r_2 \beta^{n''\omega} t_2$ are both n', p'-equivalent and evaluate to the same category element. So it suffices to consider w_1 and w_2.

Let us apply the induction hypothesis to them. They are two coterminal paths, and they are n', p'-equivalent because $w_1 \sim_{n', p'} \pi_1' \sim_{n', p'} \pi_1$ and $w_2 \sim_{n', p'} \pi_2$ and by hypothesis $\pi_1 \sim_{n', p'} \pi_2$. What is more, the arrows that were rare in π_1 and π_2 are still rare for them, and they have r occurrences in total: this was true by construction of s_1 and s_2 and is true of s_1' because $\pi_1 = q_1 a s_1 \sim_{n', p'} q_2 a s_1'$ and all arrows of q_2 are frequent so the rare subwords of s_1 and s_1' are the same. The arrows that were frequent in π_1 and π_2 are still frequent in s_1 and s_2 and occur at least as many times as they did in π_1 and π_2 respectively: we have guaranteed this for the arrows of C by inserting $\beta^{n''\omega}$, and this is clear for the arrows outside of C as all their occurrences in π_1 and π_2 were in s_1 and s_2 respectively, and s_1' has at least as many occurrences of every letter as s_1 does (this is a consequence of the statement of

Lemma 7.7). This ensures that $s_1' \sim_{n',p'} s_2$, and that n' is still an $|S|$-distant rare-frequent threshold for them.

Hence, by the induction hypothesis, we have $s_1' \equiv s_2$, so that by compositionality we have $\pi_1 \equiv \pi_2$.

Case I.2: C does not occur again after a. In this situation, we cannot apply Lemma 7.7. However, intuitively, the arrows visited in q_1 and in q_2 must be disjoint from those visited in the rest of the paths, so we can independently reason on q_1 and q_2, and on s_1 and s_2.

Formally, we first claim that $q_1 \equiv q_2$ by the base case (Claim 8.2) of the outer induction (Claim 8.1). Indeed, first note that they are two coterminal paths. Now, there are two cases: every arrow x which is frequent in π_1 and π_2 is either in the SCC C of the starting object of a or not. In the first case, all the occurrences of x in π_1 must be in q_1, as any occurrence of x in s_1 would witness that we are in Case I.1; and likewise all its occurrences in π_2 must be in q_2. In the second case, all its occurrences in π_1 must be in s_1 and all its occurrences in π_2 must be in s_2, for the same reason. Thus, q_1 and q_2 contain no letter which was rare in π_1 and π_2, some of the frequent letters of π_1 and π_2 (those of the other SCCs) do not occur there at all, and the others occur there with the same number of occurrences. Thus indeed $q_1 \sim_{n',p'} q_2$, they contain no rare arrows, and n' is still an $|S|$-distant rare-frequent threshold for them. Thus, the base case of the outer induction concludes that they evaluate to the same category element.

We now claim that $s_1 \equiv s_2$ by the induction case of the outer induction (Claim 8.1). Indeed, they are again two coterminal paths. What is more, by the previous reasoning, the arrows that are frequent in π_1 and π_2 either occur only in s_1 and s_2 or do not occur there at all. Thus, s_1 and s_2 contain r rare arrows (for the arrows that were already rare in π_1 and π_2), and the frequent arrows either occur in s_1 and s_2 with the same number of occurrences as in π_1 and π_2 or not at all. This implies that n' is still an $|S|$-distant rare-frequent threshold for s_1 and s_2. Thus, we have $s_1 \sim_{n',p'} s_2$ and the induction case of the outer induction establishes that $s_1 \equiv s_2$.

Thus, by compositionality, we know that π_1 and π_2 evaluate to the same category element. We have concluded both cases of the outer induction proof and shown Claim 8.1.

8.3. **Concluding the proof.** We have proven Claim 8.1, and explained afterwards how to use it to show that π_1 and π_2 evaluate to the same category element. This implies that n', p'-equivalence for our choice of n', hence also n, p-equivalence, is compatible with S_E. Thus, by Corollary 5.5 we know that L is in $\mathbf{ZG} * \mathbf{D}_p$. Hence, $L \in \mathbf{LZG}_p$ implies that $L \in \mathbf{ZG} * \mathbf{D}_p$, so we have shown Claim 5.7. Together with Claim 5.6, it establishes the locality result $\mathbf{LZG}_p = \mathbf{ZG} * \mathbf{D}_p$, and we have shown Theorem 1.1.

9. CONCLUSION

In this paper, we have given a characterization of the languages of $\mathbf{ZG}$, and proved that the variety $\mathbf{ZG}$ is local. More specifically, we have shown this for all the varieties $\mathbf{ZG}_p$ for $p > 0$, in particular $\mathbf{MNil} = \mathbf{ZG}_1 = \mathbf{ZG} \cap \mathbf{A}$.

A natural question for further study is whether the variety $\mathbf{ZE}$ is also local. This question seems more complicated. Indeed, as proved by Almeida [Alm94], we have $\mathbf{ZE} = \mathbf{G} \vee \mathbf{Com}$, that is, $\mathbf{ZE}$ is the variety of monoids generated by the variety $\mathbf{G}$ of groups and the variety $\mathbf{Com}$ of commutative languages. Now, $\mathbf{G}$ is a local variety [TW85, Example 1.3], while

Com is not [TW85, Example 1.4]. Further, as we have shown that **LZG** = **ZG** ∗ **D** and **Com** is a subset of **ZG**, the counter-example languages (e.g., $e^*af^*be^*cf^*$) to the locality of **Com** (i.e., that are in **LCom** but not in **Com** ∗ **D**) cannot be counter-examples to the locality of **ZE** (because they are in **LCom**, hence **LZG**, hence **ZG** ∗ **D**, hence **ZE** ∗ **D**). This being said, if **ZE** is indeed local, a proof would probably require different techniques from ours, given that we do not see how our techniques could be used even to reprove the locality of **G**.

We hope that extending our approach to a study of locality for centrally defined varieties in general could lead to such general results on the interplay of join operations and of the locality or non-locality for arbitrary varieties, in the spirit of the results shown in [CE13] for various Mal'cev products.

References

[AA87] Jorge Almeida and Assis Azevedo. Implicit operations on certain classes of semigroups. In *Semigroups and their Applications*. Springer, 1987.

[AJP21] Antoine Amarilli, Louis Jachiet, and Charles Paperman. Dynamic membership for regular languages. In *ICALP*, 2021.

[Alm94] Jorge Almeida. *Finite semigroups and universal algebra*. Series in Algebra: Volume 3. World Scientific, 1994.

[Alm96] Jorge Almeida. A syntactical proof of locality of **DA**. *Int. J. Algebra Comput.*, 6(2), 1996. `doi:10.1142/S021819679600009X`.

[Aui00] Karl Auinger. Semigroups with central idempotents. In *Algorithmic problems in groups and semigroups*. Springer, 2000.

[Aui02] Karl Auinger. Join decompositions of pseudovarieties involving semigroups with commuting idempotents. *Journal of Pure and Applied Algebra*, 170(2), 2002. `doi:10.1016/S0022-4049(01)00081-0`.

[Aui10] Karl Auinger. On the decidability of membership in the global of a monoid pseudovariety. *Int. J. Algebra Comput.*, 20(2), 2010. `doi:10.1142/S0218196710005571`.

[BJG08] Jørgen Bang-Jensen and Gregory Z Gutin. *Digraphs: theory, algorithms and applications*. Springer Science & Business Media, 2008. Second edition.

[BS73] J.A. Brzozowski and Imre Simon. Characterizations of locally testable events. *Discrete Mathematics*, 4(3), 1973. `doi:10.1016/s0012-365x(73)80005-6`.

[CE13] Alfredo Costa and Ana Escada. Some operators that preserve the locality of a pseudovariety of semigroups. *International Journal of Algebra and Computation*, 23(03):583–610, 2013.

[DP15] Luc Dartois and Charles Paperman. Alternation hierarchies of first order logic with regular predicates. In *FCT*, 2015.

[Eil76] Samuel Eilenberg. *Automata, languages, and machines. Vol. B*. Academic Press, New York-London, 1976. With two chapters by Bret Tilson, Pure and Applied Mathematics, Vol. 59.

[GMS17] Nathan Grosshans, Pierre McKenzie, and Luc Segoufin. The power of programs over monoids in **DA**. In *MFCS*, 2017.

[KL13] Manfred Kufleitner and Alexander Lauser. Quantifier alternation in two-variable first-order logic with successor is decidable. In *STACS*, 2013.

[Kna83a] Robert Knast. A semigroup characterization of dot-depth one languages. *RAIRO Theor. Informatics Appl.*, 17(4), 1983. `doi:10.1051/ita/1983170403211`.

[Kna83b] Robert Knast. Some theorems on graph congruences. *RAIRO Theor. Informatics Appl.*, 17(4), 1983. `doi:10.1051/ita/1983170403311`.

[McN74] Robert McNaughton. Algebraic decision procedures for local testability. *Mathematical Systems Theory*, 8(1), 1974. `doi:10.1007/bf01761708`.

[Pin86] Jean-Éric Pin. *Varieties of formal languages*. Foundations of Computer Science. Plenum Publishing Corp., New York, 1986. With a preface by M.-P. Schützenberger, Translated from the French by A. Howie. `doi:10.1007/978-1-4613-2215-3`.

[Pin19]   Jean-Éric Pin. Mathematical foundations of automata theory. `https://www.irif.fr/~jep/PDF/` `MPRI/MPRI.pdf`, 2019.

[PS16]   Thomas Place and Luc Segoufin. Decidable characterization of $\mathrm{FO}^2(<,+1)$ and locality of **DA**, 2016. `arXiv:arXiv:1606.03217`.

[Rei82]   Jan Reiterman. The Birkhoff theorem for finite algebras. *Algebra universalis*, 14, 1982.

[SFMS97]   Gudmund Skovbjerg Frandsen, Peter Bro Miltersen, and Sven Skyum. Dynamic word problems. *JACM*, 44(2), 1997.

[SJ73]   Price Stiffler Jr. Chapter 1. Extension of the fundamental theorem of finite semigroups. *Advances in Mathematics*, 11(2), 1973. `doi:10.1016/0001-8708(73)90007-8`.

[Ste04]   Benjamin Steinberg. A modern approach to some results of Stiffler. In *Semigroups and Languages*. World Scientific, 2004. `doi:10.1142/9789812702616_0013`.

[Str82]   Howard Straubing. The variety generated by finite nilpotent monoids. *Semigroup Forum*, 24(1), 1982. `doi:10.1007/bf02572753`.

[Str85]   Howard Straubing. Finite semigroup varieties of the form **V*D**. *Journal of Pure and Applied Algebra*, 36, 1985. `doi:10.1016/0022-4049(85)90062-3`.

[Str94]   Howard Straubing. *Finite automata, formal logic, and circuit complexity*. Birkhäuser Boston Inc., 1994.

[Str15]   Howard Straubing. A new proof of the locality of **R**. *International Journal of Algebra and Computation*, 25(01&02), 2015. `doi:10.1142/s0218196715400111`.

[Til87]   Bret Tilson. Categories as algebra: An essential ingredient in the theory of monoids. *J. Pure Appl. Algebra*, 48(1-2), 1987. `doi:10.1016/0022-4049(87)90108-3`.

[TW85]   Denis Thérien and Alex Weiss. Graph congruences and wreath products. *J. Pure Appl. Algebra*, 36(2), 1985. `doi:10.1016/0022-4049(85)90071-4`.

[TW98]   Denis Thérien and Thomas Wilke. Over words, two variables are as powerful as one quantifier alternation. In *STOC*, 1998.

[TW01]   Denis Thérien and Thomas Wilke. Temporal logic and semidirect products: An effective characterization of the until hierarchy. *SIAM J. Comput.*, 31(3), 2001. `doi:10.1137/S0097539797322772`.

APPENDIX A. SELF-CONTAINED PROOF OF LEMMA 8.4

Lemma 8.4 [BJG08]. *Let G be a strongly connected nonempty directed multigraph. We have:*

- *G is a simple cycle; or*
- *G contains a simple cycle $v_1 \to \cdots \to v_n \to v_1$ with $n \geq 1$, where all vertices $v_1, \ldots, v_n$ are pairwise distinct, such that all intermediate vertices $v_2, \ldots, v_{n-1}$ only occur in the edges of the cycle, and such that the removal of the cycle leaves the graph strongly connected (note that the case $n = 1$ corresponds to the removal of a self-loop); or*
- *G contains a simple path $v_1 \to \cdots \to v_n$ with $n \geq 2$ where all vertices are pairwise distinct, such that all intermediate vertices $v_2, \ldots, v_{n-1}$ only occur in the edges of the path, and such that the removal of the path leaves the graph strongly connected (note that the case $n = 2$ corresponds to the removal of a single edge).*

We repeat here that the result is standard. The proof given below is only for the reader's convenience, and follows [BJG08].

Proof. This result is showed using the notion of an ear decomposition of a directed multigraph. Specifically, following Theorem 5.3.2 of [BJG08], for any nonempty strongly connected multigraph G, we can build a copy of it (called G') by the following sequence of steps, with the invariant that G' remains strongly connected:

- First, take some arbitrary simple cycle in G and copy it to G';
- Second, while there are some vertices of G that have not been copied to G', then pick some vertex v of G that was not copied, such that there is an edge (v', v) in G with v' a vertex that was copied. Now take some shortest path (hence a simple path) $v \to \cdots \to v''$ from v to the subset of the vertices of G that had been copied to G'. This path ends at a vertex v'' which may or may not be equal to v'. If $v'' \neq v'$, then we have a simple path $v' \to v \to \cdots \to v''$, which we copy to G'; otherwise we have a simple cycle, which we copy to G'. Note that, in both cases, all intermediate vertices in the simple path or simple cycle that we copy only occur in the edges of the path or cycle (as they had not been previously copied to G'). Further, G' clearly remains strongly connected after this addition.
- Third, once all vertices of G have been copied to G', take each edge of G that has not been copied to G' (including all self-loops), and copy it to G' (as a simple path of length 1). These additions preserve the strong connectedness of G'.

At the end of this process, G' is a copy of G.

Now, to show the result, take the graph G, consider how we can construct it according to the above process, and distinguish three cases:

- If the process stopped at the end of the first step, then G is a simple cycle (case 1 of the statement).
- If the process stopped after performing a copy in the second step, then considering the last simple path or simple cycle that we added, then it satisfies the conditions and its *removal* from G gives a graph which is still strongly connected (case 2 or case 3 of the statement).
- If the process stopped after performing a copy in the third step, then considering the last edge that we added, then it is a simple path of length 1 and its *removal* from G gives a graph which is still strongly connected (case 3 of the statement).

This concludes the proof.   □