involve

a journal of mathematics

msp

2014 vol. 7, no. 1

involve

msp.org/involve

See inside back cover or msp.org/involve for submission instructions. The subscription price for 2014 is US $120/year for the electronic version, and $165/year (+$35, if shipping outside the US) for print and electronic. Subscriptions, requests for back issues from the last three years and changes of subscribers address should be sent to MSP.

Involve (ISSN 1944-4184 electronic, 1944-4176 printed) at Mathematical Sciences Publishers, 798 Evans Hall #3840, c/o University of California, Berkeley, CA 94720-3840, is published continuously online. Periodical rate postage paid at Berkeley, CA 94704, and additional mailing offices.

Involve peer review and production are managed by EditFLOW® from Mathematical Sciences Publishers.

PUBLISHED BY

mathematical sciences publishers

nonprofit scientific publishing

http://msp.org/

INVOLVE 7:1 (2014)
dx.doi.org/10.2140/involve.2014.7.1

Seriation algorithms for determining the evolution of *The Star Husband Tale*

Crista Arangala, J. Todd Lee and Cheryl Borden

(Communicated by Kenneth S. Berenhaut)

We give an introduction to seriation techniques and apply such techniques to the North American folklore tale known as the *Star Husband Tale*. In particular, a spectral algorithm with imposed clustering is applied, with significant results that support the algorithm's effectiveness.

1. Introduction

In the field of archeology, many researchers investigate whether objects can be chronologically ordered based strictly on their physical characteristics, in a process known as seriation. Typically, one tries to arrange artifacts from numerous sites in sequential order.

A variety of seriation techniques are in common use; historical reviews can be found in [Lyman et al. 1998; O'Brien and Lyman 2002]. One common technique, known as frequency seriation, is based on the relative frequency of artifact types. Other seriation techniques, such as occurrence and phyletic seriation, are based on similar characteristics between artifacts. The idea behind them is that one can use the presence or absence of certain characteristics, or attributes, in particular digs in order to order the artifacts chronologically. These seriation techniques, introduced in [Petrie 1899], are based on a binary incidence matrix (called the Petrie matrix) and attempt to minimize dissimilarities between digs by ordering them appropriately. (Another class of seriation techniques, which will not concern us, involves what's called phylogenetic trees. See [Buneman 1971; Huson and Bryant 2006] for information.)

Despite their popularity in archeology, seriation techniques have not been widely used in studying the geographical spread of stories and folklore. In this paper, we use a dissimilarity approach to seriate a well-known North American Indian folk tale, the *Star Husband Tale* (see Section 2). We do this both by brute force and by using an elegant spectral algorithm from [Atkins et al. 1999]. We show that,

MSC2000: 15-04, 15A18.
Keywords: seriation, linear algebra, folk tales.

based strictly on the content dissimilarities among versions, one can track the tale's progression in a way that matches the geographical proximity of the tribes telling these tales.

Organization of the paper. Section 2 describes the tale and mentions some prior studies. A basic seriation technique is described in Section 3 and results from its application to a small subset of the data are analyzed in Section 5. To be able to study a larger data set, we discuss in Section 4 a spectral algorithm from [Atkins et al. 1999]. As shown in Section 6, this algorithm is able to order and cluster successfully a larger data set consisting of eighty-six versions of our tale.

2. The *Star Husband Tale*

The basic form of the *Star Husband Tale* [Young 1978] tells of two girls who are sleeping out in the open during the night. While outside, they see two stars and each girl makes a wish to be married to a star. When they awake, both have been transported to the heavens and are married to the stars as they wished. One of the star husbands is a young man and the other is an older man. Heedless of a warning they've received, the girls at one point start digging in the heavens and make a hole through which they can see their old homes below. Overcome with homesickness, they eventually lower themselves down to earth using a rope.

Dundes [1965] discusses various narrative elements peculiar to eighty-six versions of the *Star Husband Tale* coming from 44 tribes throughout North America. These tribes are grouped into nine geographical zones: Eskimo, Mackenzie, North Pacific, California, Plateau, Plains, Southeast, Southwest, and Woodlands [Carroll 1979]. Thompson chose those characteristics that occur most frequently as the principal tale elements. These principal tale elements are then collated into archetypes and subarchetypes [Rich 1971]. A sample of the principal tale elements followed by their archetypes is presented in Table 1.

Table 2 gives a sample of the records (tribes) and the traits that are present (or absent) in their versions of the tale.

In total, 86 versions of the *Star Husband Tale* and a total of 135 traits (archetypes and subarchetypes) are included in the study.

3. Seriation

In this section we express the seriation problem mathematically: how to list a set of objects so as to minimize the sum of the dissimilarities between consecutive objects. This is the traveling salesman problem: our objects (versions of the tale) correspond to cities, and the measure of dissimilarities corresponds to distances. However we can make a simplifying assumption (which often holds only approximately) that makes the problem easier than the general traveling salesman problem.

Trait A: Number of women		Trait B: Introductory action	
A1	One	B1	Trait not present
A2	Two	B2	Wish for star husband
A3	Two at first, then one	B3	Pursuit of porcupine
A4	More than two	B4	Miscellaneous
Trait D: Method of ascent		**Trait H: Taboo broken in Upper World**	
D1	Not indicated	H1	No taboo broken
D2	Stretching tree	H2	Digging or disturbing ground
D3	Translation during sleep	H3	Moving a large rock
D4	Carried through the air	H4	Looking somewhere
D5	Carried in a basket	H5	Shooting a meadow lark
D6	Carried by whirlwind	H6	Making noise before an animal sings
D7	Carried by a feather		

Table 1. Sample of traits and their archetypes. Some archetypes for traits B, D, and H are further subdivided (H1, H1a...).

Tale 1	Eskimo	Smith Sound	A1	B3a	D3	H2
Tale 2	Eskimo	Kodiak 1	A3	B1	D3,4	H2
Tale 3	California	Patwin	A1	B3a	D6	H1
Tale 4	California	Washo 1	A2	B1	D2	H1a
Tale 5	North Pacific	Snuqualmi 1	A2	B1	D2	H1
Tale 6	North Pacific	Snuqualmi 2	A2	B1	D2	H1
Tale 7	Plains	Sarsi	A3	B1	D3	H1a
Tale 8	Plains	Blackfoot 1	A3	B1	D3,7	H1a

Table 2. Classification of a sampling of tales with respect to the traits listed in Table 1. More than one archetype or subarchetype can be present for a given trait in a tribe's version of the tale.

The first step is to express the information in an *incidence matrix*. For the data in Table 2, this matrix is

$$A = \begin{array}{r|ccccccccccccc|}
 & \text{A1} & \text{A2} & \text{A3} & \text{B1} & \text{B3a} & \text{D2} & \text{D3} & \text{D4} & \text{D6} & \text{D7} & \text{H1} & \text{H1a} & \text{H2} \\
\text{Tale 1} & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\
\text{Tale 2} & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 1 \\
\text{Tale 3} & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\
\text{Tale 4} & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
\text{Tale 5} & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
\text{Tale 6} & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
\text{Tale 7} & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\
\text{Tale 8} & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 0
\end{array}.$$

Each row corresponds to a tale, and each column to a trait that the tale may or may not possess. The measure of dissimilarity between tales is the number of places where the corresponding rows differ.

If the rows were written in the order in which the tale evolved, it would be natural to expect 1s to cluster together in each column: traits may be introduced or removed as the tale is handed down, but a given trait is unlikely to jump in and out of existence several times. We formalize this with the following concept:

Definition. A binary matrix A is called a *Petrie matrix*, or *P-matrix*, if each column contains a sequence of 0s, followed by a sequence of 1s, followed again by a sequence of 0s. (In each column, it is possible for any of these sequences to have length zero.)

Thus a P-matrix is characterized by the absence of *embedded* 0s (that is, 0s that have 1s above and below them in the same column). Equivalently, all 1s in a column are consecutive.

In our application, saying that the incidence matrix is a P-matrix means that once a trait is present in the tale it may remain in the tale throughout its progression; but if the trait then disappears from the tale, it will not reappear in later renditions.

Definition. A matrix A is called *pre-P*, or pre-Petrie, if there is a row-permutation matrix Σ such that ΣA is Petrie. (From a permutation σ we obtain a permutation matrix Σ by setting $\Sigma(i, \sigma(i)) = 1$ for each row index i, and setting other entries equal to 0.)

The *consecutive ones problem* consists in rearranging the rows of an incidence matrix so it becomes a P-matrix — corresponding, in our case, to sorting the tales into a temporal order consistent with the changes in traits. If the matrix is pre-P the problem is solvable (by definition!) and an appropriate permutation matrix can be found quickly by any of several efficient algorithms. However, in applications, it is often the case that the incidence matrix is not pre-P, just “almost” so. In that case, the problem becomes more complex and a solution is not guaranteed to exist [Dundes 1965]; nonetheless one can look for a permutation that brings the incidence matrix into a form as close as possible to a P-matrix.

Dissimilarity. Our first approach is brute force: we test all possible permutations and choose one that gives a result closest to a P-matrix. Because of exponential growth, we are limited with this approach to small data sets; but at least we can avoid having to compare the rows of A itself each time. Instead, we introduce the *similarity matrix* $S = AA^T$, whose rows and columns both correspond to tales. Its name is due to the fact that the off-diagonal entries of S express how many 1s two rows of A have in common — that is, how many traits two tales share. (The diagonal entries in S show the number of traits possessed by each tale. It is easy to see that S is symmetric and nonnegative.)

For instance, taking again the data in Table 2, the similarity matrix is

$$S = AA^T = \begin{pmatrix} 4 & 2 & 2 & 0 & 0 & 0 & 1 & 1 \\ 2 & 5 & 0 & 1 & 1 & 1 & 3 & 3 \\ 2 & 0 & 4 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 4 & 3 & 3 & 2 & 2 \\ 0 & 1 & 1 & 3 & 4 & 4 & 1 & 1 \\ 0 & 1 & 1 & 3 & 4 & 4 & 1 & 1 \\ 1 & 3 & 0 & 2 & 1 & 1 & 4 & 4 \\ 1 & 3 & 0 & 2 & 1 & 1 & 4 & 5 \end{pmatrix}.$$

Even more convenient to use is the *dissimilarity matrix* D, defined by setting $D(i,j) = n - S(i,j)$ for $1 \leq i, j \leq m$. Here n is the number of columns of A (possible traits) and m the number of rows (tales). For our running example, with $m = 8$ and $n = 13$, the dissimilarity matrix is

$$D = \begin{pmatrix} 9 & 11 & 11 & 13 & 13 & 13 & 12 & 12 \\ 11 & 8 & 13 & 12 & 12 & 12 & 10 & 10 \\ 11 & 13 & 9 & 13 & 12 & 12 & 13 & 13 \\ 13 & 12 & 13 & 9 & 10 & 10 & 11 & 11 \\ 13 & 12 & 12 & 10 & 9 & 9 & 12 & 12 \\ 13 & 12 & 12 & 10 & 9 & 9 & 12 & 12 \\ 12 & 10 & 13 & 11 & 12 & 12 & 9 & 9 \\ 12 & 10 & 13 & 11 & 12 & 12 & 9 & 8 \end{pmatrix}.$$

Note that the entry $D(i, i+1)$ gives the number of changes (dissimilarities) from the i-th row to the next, so the quantity that concerns us is the sum $\sum_{i=1}^{m-1} D(i, i+1)$, which we call the *total dissimilarity of* A. For our running example this number is $11+13+13+10+9+12+9 = 77$.

Once we apply a permutation σ to the rows of A we are looking at the quantity

$$L(\sigma) := \sum_{i=1}^{m-1} D(\sigma(i), \sigma(i+1)).$$

We call $L(\sigma)$ the *total dissimilarity of A permuted by* σ. We can also view it as the sum of the $(i, i+1)$ entries of the conjugate matrix $\Sigma D \Sigma^{-1}$, where Σ is the permutation matrix. Indeed, under the action of σ, the incidence matrix A becomes ΣA, so S becomes $\Sigma A(\Sigma A)^T = \Sigma A A^T \Sigma^T = \Sigma S \Sigma^{-1}$, since for a permutation matrix, transposing is the same as inverting. Similarly, Σ turns D into $\Sigma D \Sigma^{-1}$.

Our goal now is to *minimize* the total dissimilarity over all permutations. When a dissimilarity-minimizing permutation is applied to A, the result will be as close to a P-matrix as possible, and the following criterion says whether or not it is in fact a P-matrix:

Theorem 1 [Shuchat 1984]. *Let A be an $m \times n$ incidence matrix, with similarity matrix $S = AA^T$ and dissimilarity matrix D. For any row permutation Σ, the total dissimilarity $L(\Sigma)$ satisfies*

$$L(\Sigma) \geq \operatorname{trace}(D) = mn - \operatorname{trace}(S).$$

Further, A is a pre-P matrix if and only if equality is attained for some Σ, in which case ΣA is a Petrie matrix.

This is the translation of [Shuchat 1984, Theorem 1] to our setup, which differs from Shuchat's in that his matrices include a dummy row.

For our running example one dissimilarity-minimizing permutation is $\sigma = 78213654$ (this means the entries $\Sigma(1,7)$, $\Sigma(2,8)$, …, $\Sigma(8,4)$ equal 1). Upon its application the dissimilarity matrix becomes

$$\Sigma D \Sigma^{-1} = \begin{pmatrix} 8 & 9 & 12 & 12 & 11 & 13 & 10 & 12 \\ 9 & 9 & 12 & 12 & 11 & 13 & 10 & 12 \\ 12 & 12 & 9 & 9 & 10 & 12 & 12 & 13 \\ 12 & 12 & 9 & 9 & 10 & 12 & 12 & 13 \\ 11 & 11 & 10 & 10 & 9 & 13 & 12 & 13 \\ 13 & 13 & 12 & 12 & 13 & 9 & 13 & 11 \\ 10 & 10 & 12 & 12 & 12 & 13 & 8 & 11 \\ 12 & 12 & 13 & 13 & 13 & 11 & 11 & 9 \end{pmatrix},$$

and $L(\Sigma) = 72$ is the minimum total dissimilarity. Since this is strictly more than $mn - \operatorname{tr}(S) = 70$, we conclude that A is not a pre-P matrix.

The permutation 78213654 encodes a possible reconstruction of the evolution of *Star Husband Tale* based on dissimilarities between traits. It suggests this scenario:[1] The earliest tale is 7, from the Plains Sarsi tribe. From the Plains tribes the tale went to the Eskimo tribes, followed by California and Northern Pacific tribes. Note that even with just a few traits included in these tales, this method tends to appropriately group the Plains tribes together, holding true to their geographic location. One can see the location of the tribes in Figure 1.

We used Mathematica version 8 to generate all permutations that minimize the total dissimilarity; with just 8 tribes and 13 traits this took approximately 248 seconds of CPU time. (Mathematica 8 has a built-in function that finds one shortest tour in the traveling salesman problem; this could be applied to the dissimilarity matrix to find a minimizing permutation.)

[1] By construction, the reverse permutation, 45631287, is also minimizing, so the reverse order would be equally possible: Tale 7 the most recent, etc. Many other minimizing permutations exist.

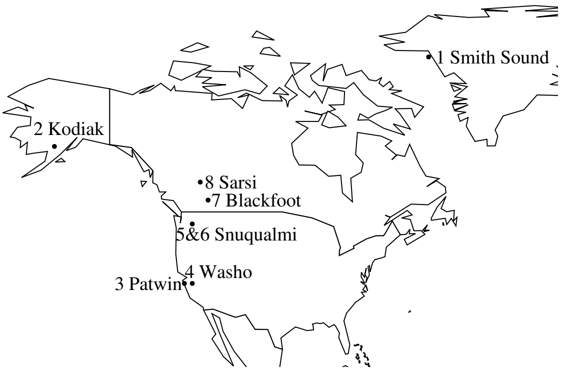


Figure 1. Locations of tribes from Table 2.

The consecutive ones problem. We return to the formulation given on page 4, where we mentioned that a P-matrix is one having no embedded 0s. Equivalently, for a P-matrix the distance between the first and last 1s in each column is 1 less than the number of 1s in that column (the distance is $k-1$ when there are k consecutive 1s). Equivalently, for a P-matrix the sum of these distances over all columns is simply the total number of 1s minus the number of columns. And obviously, for *any* binary matrix the sum must be at least as great as this difference.

Since the total number of 1s in A is the sum of diagonal elements of $S = AA^T$, we have proved the following:

Theorem 2 [Shuchat 1984]. *Let A be an $m \times n$ incidence matrix, and for each column j, let $r_j(A)$ be the difference between the row index of the last* 1 *in column j and that of the first* 1 *in the same column. Given a row-permutation matrix Σ, define the* 1*-content*

$$R(\Sigma) = \sum_{j=1}^{n} r_j(\Sigma A),$$

where of course $r_j(\Sigma A)$ is the corresponding difference for the permuted matrix ΣA. Then

$$R(\Sigma) \geq \operatorname{tr}(S) - n.$$

The matrix A is a pre-P matrix if and only if equality is attained for some Σ, in which case ΣA is a Petrie matrix.

For the permutation matrix A of page 3, we have $R(\text{identity}) = \sum_{j=1}^{n} r_j(A) = 35$. The minimum 1-content $R(\Sigma)$ using the 8 tribes and 13 tales from Table 2 turns out to be 26, again showing that the incidence matrix A is not a pre-P matrix. An example of a permutation that can be used to obtain this minimum 1-content is 54876132.

These give the chronological order of evolution of *Star Husband Tale* based on the number of embedded 0s within tales. For example, the minimum dissimilarity permutation 54876132 orders the progression of the tale starting with Tale 5, which came from the North Pacific Snuqualmi 1 tribe. Based on this analysis, of this small data set, the tale may have bounced between North Pacific tribes and Plains tribes before moving to the Eskimo and California tribes. Mathematica version 8 generated all permutations that minimize the number of embedded 0s for the matrix A in 61 seconds of CPU time. This is four times faster than with the dissimilarity method (page 6); but a similar computation for the full problem — 86 tribes with 135 traits — is impossible using the brute-force approach. In the next section, we discuss an algorithm that is computationally more practical for larger data sets. A comparison of all three algorithms is presented in Section 5.

4. A spectral algorithm for seriation

Atkins et al. [1999] gave an algorithm, based on eigenvalues and eigenvectors, for finding a permutation matrix Σ such that ΣA is a P-matrix. It assumes that the original matrix A is a pre-P matrix, but it degrades gracefully in the absence of that condition (that is, its results are not greatly affected if the matrix is almost a pre-P matrix.)

We start with some definitions.

Definition. A matrix $S \in \mathbb{R}^{m\times m}$ is *reducible* if there exists a permutation matrix Σ such that

$$\Sigma S \Sigma^{-1} = \begin{pmatrix} B & C \\ 0 & D \end{pmatrix},$$

where $B \in \mathbb{R}^{r\times r}$, $D \in \mathbb{R}^{(m-r)\times(m-r)}$, and $0 < r < m$. If no such permutation exists, S is called *irreducible*.

Definition. Given an $m \times m$ symmetric matrix S and a diagonal matrix D such that $D(i,i) = \sum_{j=1}^{m} S(i,j)$ for $1 \le i \le m$, the *Laplacian* of S is $L = D - S$. It is easy to see that $e = (1,\dots,1)$ is an eigenvector of L, with eigenvalue 0. The minimum eigenvalue of L with an eigenvector orthogonal to e is called the *Fiedler value*, and a corresponding eigenvector is a *Fiedler vector*.

Definition. A square matrix S is called a *Robinson matrix* [1951], or *R-matrix*, if

$$S(i,j) \le S(i,k) \ \text{ for } \ j < k < i \quad \text{and} \quad S(i,j) \ge S(i,k) \ \text{ for } \ i < j < k.$$

If there is a permutation matrix Σ such that $\Sigma S \Sigma^{-1}$ is an R-matrix, S is called a *pre-R matrix*.

Theorem 3 [Atkins et al. 1999]. *Any R-matrix has a monotone Fiedler vector.*

Theorem 4 [Atkins et al. 1999]. *Let S be a pre-R matrix with a simple Fiedler value and a Fiedler vector with no repeated values. Let Σ_1 and Σ_2, respectively be the permutations induced by sorting the values in the Fiedler vector in increasing and decreasing order. Then $\Sigma_1 S \Sigma_1^{-1}$ and $\Sigma_2 S \Sigma_2^{-1}$ are R-matrices and no other permutations of S produce R-matrices.*

For the similarity matrix S of our running example (page 5), is irreducible with simple Fiedler value approximately equal to 3.517 and Fiedler vector

$$(-1.674, 0.722, -3.800, 1.238, 0.757, 0.757, 1, 1).$$

Since the Fiedler vector is not monotonic, by Theorem 3, S is not an R-matrix. Under the assumption that S is a pre-R matrix, one can find the permutation that puts the Fiedler vector in increasing order; however Theorem 4 cannot be applied due to the occurrence of repeated entries in the Fiedler vector. The following two theorems prove helpful if the similarity matrix is reducible or has a Fiedler vector with repeated values.

Lemma 5 [Atkins et al. 1999]. *Let S_k be the irreducible blocks of a pre-R matrix A and let Σ_k be permutations that make these blocks become R-matrices. Then any permutation obtained by concatenating the Σ_k will make A become an R-matrix.*

Theorem 6 [Atkins et al. 1999]. *Let S be a pre-R matrix with a simple Fiedler value and Fiedler vector x. Suppose that there is some repeated value β in x and define I, J and K to be the indices for which*

- *$x_i < \beta$ for all $i \in I$,*
- *$x_i = \beta$ for all $i \in J$,*
- *$x_i > \beta$ for all $i \in K$.*

Then ΣS is an R-matrix if and only if Σ or its reversal can be expressed as $(\Sigma_i, \Sigma_j, \Sigma_k)$, where Σ_j is an R-matrix ordering for the submatrix $S(J, J)$ of S induced by J and Σ_i and Σ_k are the restrictions of some R-matrix ordering for S to I and K respectively.

Applying Theorem 6 to our running example, the spectral algorithm provides the permutation ordering 48765213 for the tales, under the assumption that S is a pre-P matrix. This algorithm is much less time-consuming than the seriation techniques in Section 3; it took 0.062 seconds of CPU time in Mathematica 8, and properly grouped Plains and North Pacific tribes together.

5. Seriation results: the Woodlands region

In order to show the reader the strength of seriation while still staying within current computing capacity, we chose to limit *the Star Husband* tribes to the Woodlands

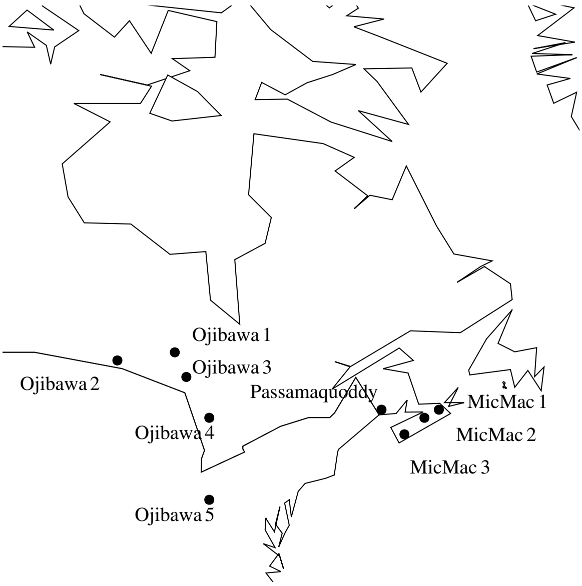


Figure 2. A map of the Woodland area tribe locations.

area, comprised of the Ojibwa, Micmac, and Passamaquoddy tribes, which contains 9 versions of the *Star Husband Tale* with 30 traits. A geographical map of the Woodlands area can be found in Figure 2.

In creating the incidence matrix for the folklore story, we used nine versions of the tale, numbered as follows:

1. Ojibwa 1
2. Ojibwa 2
3. Ojibwa 3
4. Ojibwa 4
5. Ojibwa 5
6. Micmac 1
7. Micmac 2
8. Micmac 3
9. Passamaquoddy

Note that if a trait within the nine tribes' tales was the same, the trait was eliminated from the incidence matrix all together. From the results, the minimum total dissimilarity L from all permutations is 192, while the maximum is 223. The minimum 1-content R is 86, while the maximum is 169. The range of total dissimilarities versus 1-contents can be visualized in Figure 3. Ideally, the best permutation would be the one with both the minimum L and minimum R. The permutations that satisfy this criterion are

687549312, 798651432, 798651423, 423561798, 312459687, 312459678.

In the majority of these permutations, the Ojibwa (tribes 1, 2, 3, and 4) and Mic Mac tribes (6, 7, and 8) are grouped together respectively based on their tales' characteristics. Also note that Ojibwa 4 and 5 as well as the Passamaquody tales

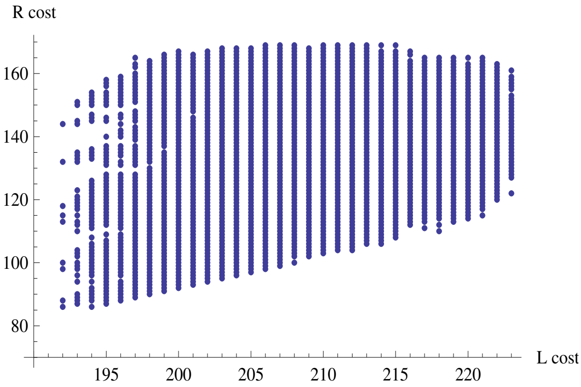


Figure 3. Cost functions length $L(\Sigma)$ versus 1-content $R(\Sigma)$ for all permutation matrices Σ.

are most often the transitional tales in the seriation. This is most likely due to the geographic central proximately of these tribes to the neighboring tribes. These are significant results as the evolution of the *Star Husband Tale* based strictly on the presence or absence of traits in the tales matches the geographic locality of these tribes as well.

As mentioned in the previous section, we introduce the spectral algorithm to find the evolution of the *Star Husband Tale* as an alternative to the seriation techniques that require the generation of all permutations of tribes. Note that both of the seriation techniques presented in Section 3 would need to make computations with 9! permutations as applied to the Woodlands tribe while the spectral algorithm looks strictly at the eigenvalues and eigenvectors of the similarity matrix.

Using the spectral algorithm technique to order the tales from the Woodlands area takes significantly less time then the traditional seriation techniques. This algorithm produces an ordering of

$$8, 7, 6, 9, 3, 5, 1, 4, 2,$$

grouping the Ojibwa tales together and Mic Mac with Passamaqudy tales which corresponds to the geographically locations of these tribes as well. In addition, the ordering puts tale 3, Ojibwa 3, closest to the Passamaqudy tale. Although geographically Ojibwa 4 is closer to the Passamaqudy tribe, Ojibwa 3 comes in a close second.

6. Seriation results: eighty-six tribes

This spectral algorithm also does a reasonable job in ordering the entire eighty-six versions of *Star Husband Tale* as well, something the other seriation techniques can

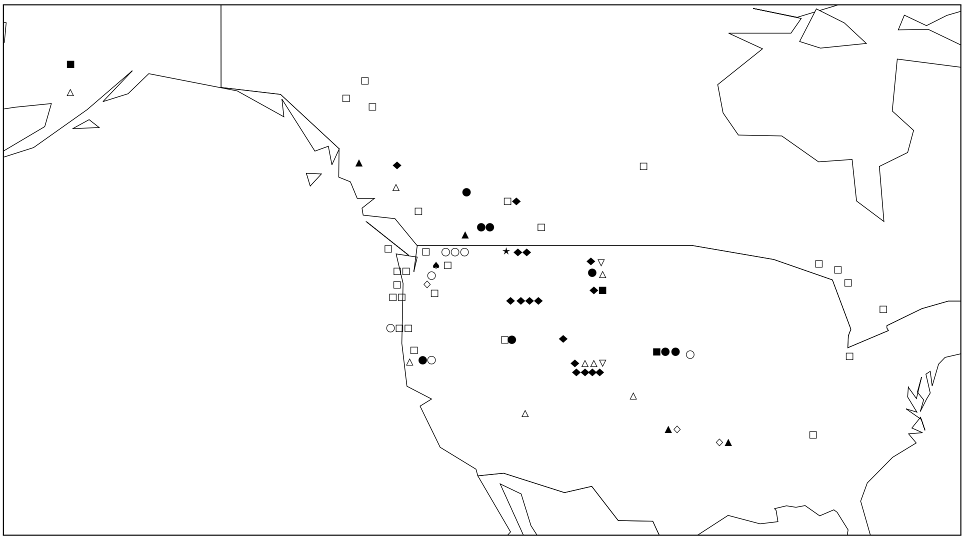


Figure 4. The geographic location and cluster based on spectral algorithm with imposed clustering. Order in which the tales represented by each symbol occurs in the seriation: □ = 1, ○ = 2, ♠ = 3, ◇ = 4, ★ = 5, ▲ = 6, ● = 7, ■ = 8, ◆ = 9, ∇ = 10, △ = 11.

not achieve computationally. The ordering produced with the eighty-six version data set reveals that *the Star Husband Tale* originated somewhere in the Plains region, possibly with the Cree tribe, and stayed in and around the Northwest border of the United States and Canada before spreading south and east to the California and Woodlands regions respectively. This ordering falls in line with Thompson's analysis [Dundes 1965], which claims that the tale did in fact originate in the Plains region.

Although the seriation techniques deal specifically with ordering and do not have a natural imposed clustering, *grouping*, if one did wish to cluster the seriated data note that tales within clusters should be similar. We impose a clustering by calculating the Hamming distance between adjacent tales in the spectral algorithm results. When the Hamming distance varies significantly a new cluster is created.

Figure 4 shows all eighty-six tale locations based on the spectral algorithm and imposed clustering. With both a seriation, ordering, and clustering, grouping, of the data one can analyze the progression of *the Star Husband Tale* between clusters. The significance of the results presented in Figure 4 is that the algorithm which produces a clustering based on characteristics of the tales also matches up geographically with the locations of the tribes.

The spectral algorithm with imposed clustering produces a first cluster of 31 tales. This cluster may be the most interesting to analyze. One can see two significant subclusters from this first cluster, one on the western coast and one along the northeastern border of the United States and Canada, and a large region in central Canada containing only one tribe, also contained in this cluster. It is highly possible that tribes from both subclusters shared common hunting grounds located in the plains region of Canada and thus producing similar versions of the tale.

We have presented here just a few algorithms for ordering *the Star Husband Tale*. With this particular data set the spectral algorithm was very successful; however all of the algorithms described assume that the data matrix is a pre-P matrix. Without this attribute, results could degrade quickly. One might consider applying clustering techniques to a similar data set. For this particular data set both agglomerative clustering and k-means clustering were explored but the results were much less attractive than those produced by the spectral algorithm with imposed clustering.

References

[Atkins et al. 1999] J. E. Atkins, E. G. Boman, and B. Hendrickson, "A spectral algorithm for seriation and the consecutive ones problem", *SIAM J. Comput.* **28**:1 (1999), 297–310. MR 99j:68049 Zbl 0930.05064

[Buneman 1971] P. Buneman, "The recovery of trees from measures of dissimilarity", pp. 387–395 in *Mathematics in the archaeological and historical sciences*, edited by F. R. Hodson et al., Edinburgh University Press, Edinburgh, 1971.

[Carroll 1979] M. P. Carroll, "A new look at Freud on myth: reanalyzing the star-husband tale", *Ethos* **7**:3 (1979), 189–205.

[Dundes 1965] A. Dundes (editor), *Introduction to Stith Thompson's "The star husband tale"*, pp. 414–415, Prentice-Hall, Englewood Cliffs, NJ, 1965.

[Huson and Bryant 2006] D. H. Huson and D. Bryant, "Application of phylogenetic networks in evolutionary studies", *Mol. Biol. Evol.* **23**:2 (2006), 254–267.

[Lyman et al. 1998] R. L. Lyman, S. Wolverton, and M. J. O'Brien, "Seriation, superposition, and interdigitation: a history of Americanist graphic depictions of culture change", *Amer. Antiquity* **63**:2 (1998), 239–261.

[O'Brien and Lyman 2002] M. J. O'Brien and R. L. Lyman, *Seriation, stratigraphy, and index fossils: the backbone of archaeological dating*, Kluwer Academic, New York, 2002.

[Petrie 1899] W. M. F. Petrie, "Sequences in prehistoric remains", *J. Anthropol. Inst.* **29**:3–4 (1899), 295–301.

[Rich 1971] G. W. Rich, "Rethinking the 'star-husbands"', *J. Amer. Folklore* **84**:334 (1971), 436–441.

[Robinson 1951] W. S. Robinson, "A method for chronologically ordering archaeological deposits", *Amer. Antiquity* **16**:4 (1951), 293–301.

[Shuchat 1984] A. Shuchat, "Matrix and network models in archaeology", *Math. Mag.* **57**:1 (1984), 3–14. MR 86a:00017 Zbl 0532.90097

[Young 1978] F. W. Young, "Folktales and social structure: a comparison of three analyses of the star-husband tale", *J. Amer. Folklore* **91**:360 (1978), 691–699.

Received: 2010-02-26 Revised: 2011-10-27 Accepted: 2011-12-03

ccoles@elon.edu
Department of Mathematics and Statistics, Elon University,
Elon, NC 27244, United States

tlee@elon.edu
Department of Mathematics and Statistics, Elon University,
Elon, NC 27244, United States

cborden2@elon.edu
Elon University, Elon, NC 27244, United States

mathematical sciences publishers

msp

INVOLVE 7:1 (2014)
dx.doi.org/10.2140/involve.2014.7.15

A simple agent-based model of malaria transmission investigating intervention methods and acquired immunity

Karen A. Yokley, J. Todd Lee, Amanda K. Brown,
Mary C. Minor and Gregory C. Mader

(Communicated by Suzanne Lenhart)

Malaria, an infectious disease prevalent in sub-Saharan Africa, is transmitted to humans through mosquito bites, and ordinary differential equation models have often been used to describe the spread of the disease. A basic agent-based model (ABM) of malaria transmission is established and compared to an ODE model of the disease in order to ascertain the similarity of the ABM to typical modeling approaches. Additionally, the ABM is described using protocol from current literature. In order to illustrate the flexibility of the ABM, the basic ABM is modified to incorporate the use of insecticide-treated bed nets (ITNs) and the effect of acquired immunity. The simulations incorporating acquired immunity and the use of ITNs show a decrease in the prevalence of the disease due to these factors. Additionally, the ABM can easily be modified to account for other complicated issues affecting malaria spread.

1. Introduction

Malaria, a blood-borne infectious disease widespread in sub-Saharan Africa, is characterized by cases of high fever, chills, nausea, sweating, and fatigue. According to the Center for Disease Control in Atlanta (CDC), each year there are 350 to 500 million reported cases of malaria, and around 1 million people die worldwide from this disease with 90% of these occurring in areas south of the Sahara [CDC 2012a]. Although malaria has not been eradicated, the spread of malaria can be controlled through both infection and disease prevention. Current and potential intervention methods for malaria that are used throughout the world include vaccination, insecticide-treated bed nets (ITNs) and insecticides such as indoor residual spray. Disease prevention efforts include antimalarial drugs that

MSC2010: 92-08, 92D25.
Keywords: malaria, agent-based modeling, population modeling.

are administered before infection and hinder the development of malaria parasites [CDC 2011].

Female *Anopheles* mosquitoes bite humans in order to obtain a blood meal for reproduction [Aron and May 1982]; the malaria parasite is transferred to humans through mosquito saliva and to mosquitoes via human blood taken in the blood meal. An infected mosquito will bite a susceptible (uninfected) human and transfer malaria sporozoites from the saliva into the human's blood, and the sporozoites then develop in a cycle in the human liver before causing symptoms. After a period of time (latency), the parasite is then prevalent in the human bloodstream and able to be passed to a mosquito drawing blood for reproduction. The parasite follows a cycle in the mosquito's gut and after some time is ready for transmission to a susceptible human [CDC 2012c].

Mathematical models of malaria have examined immunity [Aron 1988; 1983; De Zoysa et al. 1991; Gu et al. 2003; Gurarie and McKenzie 2007; Maire et al. 2006; Tumwiine et al. 2007], control methods [Chiyaka et al. 2008], climate [Dembele et al. 2009], drug resistance [Aneke 2002; Chiyaka et al. 2009; Koella and Antia 2003], vaccinations [Smith et al. 2006] and transmission parameters related to malaria spread [Chitnis et al. 2008]. Environmental, social and economic factors that contribute to the spread of malaria have also been modeled [Yang and Ferreira 2000], and several reviews and summaries have been written on existing mathematical models of malaria [Anderson and May 1991; Koella 1991; Nedelman 1985]. Ronald Ross generated one of the earliest mathematical models [Ross 1910; Spielman and D'Antonio 2001].

The basis for many of the deterministic models of malaria transmission is the Ross–MacDonald differential equation model [MacDonald 1957], but MacDonald himself also investigated more stochastic approaches [MacDonald et al. 1968] which modeled malaria through simulations based on four key epidemiologic parameters: the biting rate of the mosquito, the mosquito survival rate, the human recovery rate, and the reproduction number. These simulations incorporated seasonal changes but no incubation period of the infection. MacDonald et al. [1968] pursued computational approaches to malaria modeling "in order to adapt the model better to the detailed study of various preventive measures and to the process of eradication which cannot be handled by a deterministic model that deals only in numbers which never reach very low finite levels." However, the predictions from [MacDonald et al. 1968] were shown to have discrepancies with field research [Nájera 1974].

The current study is intended to investigate malaria transmission through an agent-based approach. Although many sophisticated malaria differential equation models have been developed, this investigation seeks to consider a simplistic agent-based approach and how that approach compares to ordinary differential equation modeling.

Agent-based approaches have been used previously for biological modeling of processes that are discrete [Castiglione et al. 2007; Eubank et al. 2004; Hinkelmann et al. 2011; Pogson et al. 2006; Wang et al. 2009]. This study uses an agent-based model (ABM) not only because actual infections in small populations may be more reasonably predicted by low finite numbers [MacDonald et al. 1968], but also because of the great potential for modeling complex aspects involved in malaria transmission.

The major objectives of the study include directing the focus of malaria modeling to agent-based approaches as in [MacDonald et al. 1968] in light of continuing computational advances and describing this approach in the language of current literature as in [Grimm et al. 2006]. The ABM presented in this study is very simplistic in order to establish a basic framework but still allow for easy incorporation of the many complex factors that affect the spread of malaria. A simple ODE system of malaria transmission based on the Ross–MacDonald model is used for comparison since ODE models are often used to model malaria. Because a simplistic ABM approach is considered, the ABM results are compared to output from the very basic and well established Ross–MacDonald ODE model. Two examples of how the ABM can be adapted to incorporate complexity are presented. These examples will investigate the effect of ITNs and acquired immunity on the spread of malaria.

2. Modeling malaria transmission

Most mathematical models representing the spread of malaria involve systems of differential equations [Aneke 2002; Chiyaka et al. 2009; 2007; 2008; Dembele et al. 2009; Koella 1991; Koella and Antia 2003; MacDonald 1957; Ngwa 2006; Tumwiine et al. 2007], and some have involved stochastic processes [Gu et al. 2003; Gurarie and McKenzie 2007; MacDonald et al. 1968; Maire et al. 2006; Smith et al. 2006]. The equations used in differential equation models describe the rates of change of the mosquito and human populations, most using standard SIR or related models. Because most malaria modeling involves deterministic differential equations, a basic ODE model is used for comparison purposes.

In addition to establishing a basic ODE model of malaria transmission, an ABM is created to describe the spread of the disease through simulated random interactions of population agents. The deterministic ODE and the ABM are compared in order to investigate how similar the modeling approaches are at a basic level. While models can be created involving both differential equations and stochastic processes, the current study investigates simplistic models that do not combine the two. Additionally, the ABM is modified for investigations related to preventative methods and immunity. All ABM simulations and ODE solutions in this study were computed using Mathematica (versions 6.0, 7.0, and 8.0).

2.1. ***Differential equation model.*** The ODE model used in this study is based on the deterministic Ross–MacDonald model [MacDonald 1957] and other quantitative models of malaria transmission [Daley and Gani 1999] with the addition of latency. This ODE model describes the population flow between three subgroups of humans and mosquitoes: those that are susceptible (without malaria), those that are latent (harboring the parasite but not yet able to transmit) and those that are infectious (infected and able to transmit).

In order to more accurately model the spread of malaria, a latency state for both humans and mosquitoes is added to the ODE system as in [Aneke 2002; Aron and May 1982]. Latency can be interpreted as the time between when a mosquito obtains a blood meal from a human and when the newly infected host can transmit the parasite. The incorporation of latency assists in accounting for hosts that carry the disease but cannot yet transmit the parasite. The time that it takes for a host to leave the latent state is referred to as the incubation time.

The population subgroups will be considered as percentages or proportions rather than in absolute numbers. The notation of lowercase h indicates a proportion of the human population with a subscript denoting which population (infected or latent) the proportion represents. The same notation with m in place of h will be used for the mosquito populations. Since the overall populations will be assumed to be constant, the susceptible population can be represented as the remainder of the total population ($h_s = 1 - h_l - h_i$, $m_s = 1 - m_l - m_i$). Hence, four differential equations are needed to describe the total population of humans and mosquitoes when latency is incorporated. The following equations (based on the system presented in [Daley and Gani 1999] with the addition of latency) represent the rates of change of the percentages of each population:

$$\frac{dh_l}{dt} = \gamma_{mh}\beta N m_i(1 - h_l - h_i) - \frac{1}{\lambda_h}h_l, \tag{1}$$

$$\frac{dh_i}{dt} = \frac{1}{\lambda_h}h_l - \frac{1}{\mu_h}h_i, \tag{2}$$

$$\frac{dm_l}{dt} = \gamma_{hm}\beta h_i(1 - m_l - m_i) - \frac{1}{\lambda_m}m_l - \frac{1}{\mu_m}m_l, \tag{3}$$

$$\frac{dm_i}{dt} = \frac{1}{\lambda_m}m_l - \frac{1}{\mu_m}m_i. \tag{4}$$

Each of the terms in the equations describes how individuals (proportionally) are entering into or exiting out of the particular population subgroup. The first term in (1) involves the interaction of infectious mosquitoes and susceptible humans. Susceptible humans become infected and enter the latent state based upon the ratio of mosquitoes to humans ($N = M_n/H_n$, where M_n is the total number of mosquitoes and H_n is the total number of humans), the rate mosquitoes bite humans β, and the transmission probability from mosquito to human γ_{mh}. These parameters are multiplied by the proportion of susceptible humans $1 - h_l - h_i$ and the proportion of infectious mosquitoes m_i. The second term in (1) represents the loss of latent humans to the infectious state based upon the human incubation time λ_h and also represents the same proportion as those moving into infectivity in (2). The human incubation time λ_h is the number of days in the latency period. The second term in (2) describes humans' recovering from malaria and returning to the susceptible population, and μ_h is the average number of days for human recovery. Equations (3) and (4) describe changes in the mosquito population using the same notation and structure as (1) and (2).

The following assumptions are used with both the ODE model and the ABM discussed in Section 2.2:

- Constant population sizes are assumed for both human and mosquito populations.
- Constant parameters are used and assumed to be sufficient for this modeling investigation.
- No individual experiences superinfection (the contraction of more than one strain of the parasite at a time).
- Climate and geography have no effect on the interactions of the populations.
- Only human and mosquito populations are considered, although mosquitoes do bite other mammals.

The parameters used in both the ODE model and ABM simulation (as described in Section 2.2) are presented in Table 1. The transmission rate from human to mosquito γ_{hm} is based on a probability-of-transfer parameter used in the original Ross–MacDonald model [MacDonald 1957]. The mosquito bite rate is based upon an assumption that mosquitoes breed on average once a week. The associated parameter β can be thought of as the overall bite rate times the proportion of human bites as in [Smith et al. 2007]. All simulations in this study used initial conditions reflecting the idea that 10% of infected humans encountered a currently uninfected mosquito population (initial proportion infected humans $h_0 = 0.1$; initial proportion infected mosquitoes $m_0 = 0$). The simulations produced for the ODE model are presented as a comparison for the ABM simulation output in Figures 1, 3, and 5.

A stability analysis was conducted on the model in (1)–(4) in order to further understand model behavior around equilibria. When evaluating the subpopulations using the parameters in Table 1, two equilibrium solutions are obtained. One of these equilibrium solutions is when all the populations are zero (the disease-free

Parameter	Definition	Value	Source
β	mosquito bite rate	$\frac{1}{7}$ bite/day	assumption
γ_{mh}	transmission probability $m \to h$	0.6	Spielman and D'Antonio 2001
γ_{hm}	transmission probability $h \to m$	1.0	MacDonald 1957
μ_m	mosquito life span	21 days	World Book 2008
μ_h	recovery time for humans	14 days	CDC 2012a
N	mosquito/human ratio	5	Shililu et al. 1998
λ_m	mosquito incubation time	7 days	CDC 2012b
λ_h	human incubation time	10 days	CDC 2012a

Table 1. Parameter values used in the ODE model and the ABM.

equilibrium, or DFE) and the other is $h_i = 0.478142$, $m_i = 0.441919$, $h_l = 0.34153$, $m_l = 0.147306$, $h_s = 0.180328$, and $m_s = 0.410774$ (the endemic equilibrium). The Jacobian matrix J for the ODE system is shown below:

$$J = \begin{pmatrix} \frac{\partial}{\partial h_l}\left[\frac{dh_l}{dt}\right] & \frac{\partial}{\partial h_i}\left[\frac{dh_l}{dt}\right] & \frac{\partial}{\partial m_l}\left[\frac{dh_l}{dt}\right] & \frac{\partial}{\partial m_i}\left[\frac{dh_l}{dt}\right] \\ \frac{\partial}{\partial h_l}\left[\frac{dh_i}{dt}\right] & \frac{\partial}{\partial h_i}\left[\frac{dh_i}{dt}\right] & \frac{\partial}{\partial m_l}\left[\frac{dh_i}{dt}\right] & \frac{\partial}{\partial m_i}\left[\frac{dh_i}{dt}\right] \\ \frac{\partial}{\partial h_l}\left[\frac{dm_l}{dt}\right] & \frac{\partial}{\partial h_i}\left[\frac{dm_l}{dt}\right] & \frac{\partial}{\partial m_l}\left[\frac{dm_l}{dt}\right] & \frac{\partial}{\partial m_i}\left[\frac{dm_l}{dt}\right] \\ \frac{\partial}{\partial h_l}\left[\frac{dm_1}{dt}\right] & \frac{\partial}{\partial h_i}\left[\frac{dm_i}{dt}\right] & \frac{\partial}{\partial m_l}\left[\frac{dm_i}{dt}\right] & \frac{\partial}{\partial m_i}\left[\frac{dm_i}{dt}\right] \end{pmatrix}$$

$$= \begin{pmatrix} -N\beta\gamma_{mh}m_i - \lambda_h & -N\beta\gamma_{mh}m_i & 0 & N\beta\gamma_{mh}(1-h_i-h_l) \\ \lambda_h & -\gamma_h & 0 & 0 \\ 0 & \beta\gamma_{hm}(1-m_i-m_l) & -\beta\gamma_{hm}h_i-\lambda_m-\gamma_m & -\beta\gamma_{hm}h_i \\ 0 & 0 & \lambda_m & -\gamma_m \end{pmatrix}.$$

When evaluating the matrix J at the DFE, the eigenvalues consist of two real roots of opposite sign and two complex roots with negative real parts; hence, the DFE is a saddle point (hyperbolic fixed point). At the endemic equilibrium, the eigenvalues consist of two negative real roots and two complex roots with negative real parts, indicating stability and attraction. If the dynamics of malaria can be controlled in such a way that the nonzero equilibrium point is closer to the origin, then the total number of overall cases of infection will most likely decrease.

2.2. *Agent-based model (ABM).*

Although an ODE model of malaria spread may reasonably model the spread of infection, the incorporation of some specific biological and environmental features of the disease (such as immunity) may result in very complex, nonlinear models. ABMs allow research to be performed by looking at the interaction of individuals in the simulated populations to model large-scale occurrences. The idea behind the ABM is that the simulation stores information about each individual mosquito and human and randomly simulates the interactions of these agents. Unless otherwise stated, the parameters used in the ABM will be denoted as they were in the ODE model discussed in Section 2.1 with values from Table 1. The ABM is conceptually similar to the work in [MacDonald et al. 1968], which uses a decision-based computer simulation to model the transmission of malaria. The study in [MacDonald et al. 1968] does use random numbers to simulate transmission of malaria in a finite population, but latency is not incorporated. The current ABM differs from [MacDonald et al. 1968] by incorporating latency, defining mosquitoes as individual agents, and not modeling seasonal effects.

An array of mosquitoes and an array of humans were created in the computer simulations and are referred to as *agents*. These agents have attributes regarding malarial infection and each are stored through advancements in the simulation representing the passage of time. Much of the ABM simulation involves random selection of two agents (one mosquito and one human) to interact and produce an infected mosquito, infected human, both, or neither (if neither is infected). Each agent has qualities stored in the code. The infective status and latent status of the agent is stored within the array (infectious, latent) with a 1 indicating yes and 0 indicating no. Hence, (1, 0) describes an agent that is infectious, (0, 1) describes an agent that is in the latent stage, and (0, 0) describes an agent with no infection.

The simulation begins by tracking the mosquitoes that take a blood meal. A human is then randomly selected as the target of the mosquito taking the blood meal. The simulation checks if either or both the selected agents are infectious and if transmission of the parasite occurs based on model probabilities. Whether an agent moves from the susceptible state to the latent stage through infection, moves to the infectious state from the latent state, or returns to the susceptible state from the infectious state is determined stochastically. The probabilities used in the ABM are based on the parameters from the ODE model, and how they are used is more specifically outlined in the Appendix.

Various methods have been used with ABMs and protocol to standardize descriptions has been suggested [Grimm et al. 2006]. A description following the overview, design concepts, and details (ODD) protocol [Grimm et al. 2006] of the basic ABM for malaria transmission is presented in the Appendix. The description in the Appendix includes flow charts and details for the basic simulation (Figures 6 and 7) and for the simulations involving preventative measures and immunity as presented in Section 3 (Figures 8 and 9). The primary method of investigation of the models in this study was through simulation, although alternate methods of analysis exist using other frameworks [Hinkelmann et al. 2011].

The ABM was simulated over a 6 month interval and the output proportions for the infectious mosquitoes are plotted versus the proportions of infectious humans in Figure 1. The corresponding output of the ODE model is also presented in Figure 1 for comparison. Noise and variation are apparent in the ABM simulation as is expected because of the stochastic nature of the model. The ODE solution

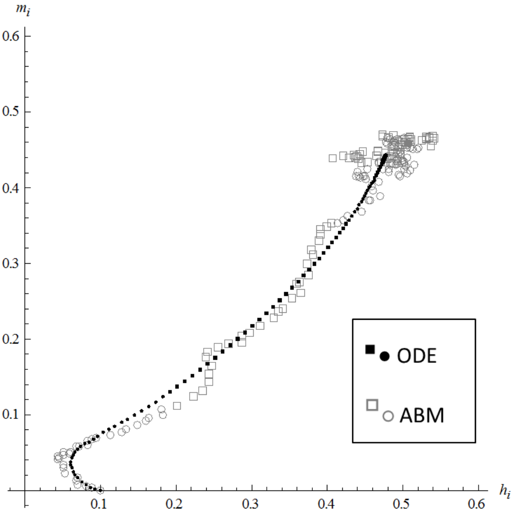


Figure 1. Plot of output from the ABM simulation, shown by the open gray shapes. The solid black shapes represent output from the ODE model with latency as described in Section 2.1. The simulations were generated over approximately six months (180 days). The horizontal axis represents proportions of infectious humans in the simulated population, and the vertical axis represents proportions of infectious mosquitoes in the simulated population. The symbols change from one shape to another at the end of each 30 day period.

moves toward an equilibrium point (which we know is stable based on analysis presented in Section 2.1). The ABM output also settles around a relatively similar point. Figure 1 illustrates that although the ABM and the ODE model do not have identical output, the two models make very similar predictions in value of output, shape of the output curve, and a nonzero settling point. Figure 1 presents the ABM output of one simulation in order to show the randomness of the approach. MacDonald et al. [1968] presented graphical representations of single simulations of their results but stated that several replicates were needed to describe the overall picture. Overall trends of the ABM using the parameters in Table 1 are presented in the black graphs in Figure 2.

3. ABM model investigations

3.1. ***ABM model sensitivity.*** The local sensitivity of the ABM was investigated visually by varying all the parameters in Table 1. Graphs were generated with ellipses surrounding each point representing potential deviation from equilibrium due to randomness of the simulation for three values of each parameter: 25% below the value in Table 1, the value in Table 1, and 25% above the value in Table 1. The two exceptions to this are for γ_{hm} and N. The transmission probability from humans to mosquitoes γ_{hm} is already at a maximum reasonable value of 1, and therefore the visual sensitivity analysis was performed with only the value from Table 1 and a value 25% below. Since the ABM uses virtual mosquitoes and humans, the value of N was kept as a whole number and sensitivity simulations were performed for two different sets of values of N (4,5,6 and 3,5,7). For each individual parameter investigation, all other parameter values were set to the values listed in Table 1.

Some parameters showed little change in the resulting ABM simulations. Neither set of simulations for varying N produced significant changes in model output. Only small changes in model output were seen when varying γ_{mh}. Slightly greater model output changes were seen with variation in γ_{hm}. The model output followed the same basic path when the values of β and λ_m were varied, but the settling points varied somewhat. The greatest sensitivity was observed when varying μ_m, μ_h, and λ_h, and graphs with these results are presented in Figure 2. Multiple simulations were run and averaged in order to obtain a more accurate representation of the trend of the simulations, and Figure 2 contains average model output over 500 simulations that were run for a six month time period. Each ellipse in Figure 2 is centered on the average value found for that point after 500 runs of the simulation plus one standard deviation in the direction of the major and minor axes.

3.2. ***Insecticide-treated bed nets.*** The ABM was modified to investigate the effects of the use of ITNs. In the ABM, the major assumptions made regarding the use of bed nets are as follows:

- A strict proportion of humans will receive bed nets in the initialization of the simulation and will continue use of bed nets until the model is run completely.
- Once a bed net is hung, it is assumed to stay intact and be used every night.
- A six month time frame (one season) is used unless otherwise indicated.
- ITNs are assumed to be 96% effective, meaning a mosquito has a 4% chance of continued interaction with the human when attempting to take a blood meal from an individual using an ITN. The level of effectiveness of ITNs is expected to be somewhere from 95% to 99%, taking into account efficacy as well as potential wear [Curtis et al. 1992; N'Guessan et al. 2001].

A more formal description of the ABM process incorporating the use of bed nets is presented in Appendix A.3.2.

Trends of infection were compared between populations with proportions of humans using bed nets ranging from 10% to 70%. The plots in Figure 3 show examples of simulations using a six month period where the given proportion of

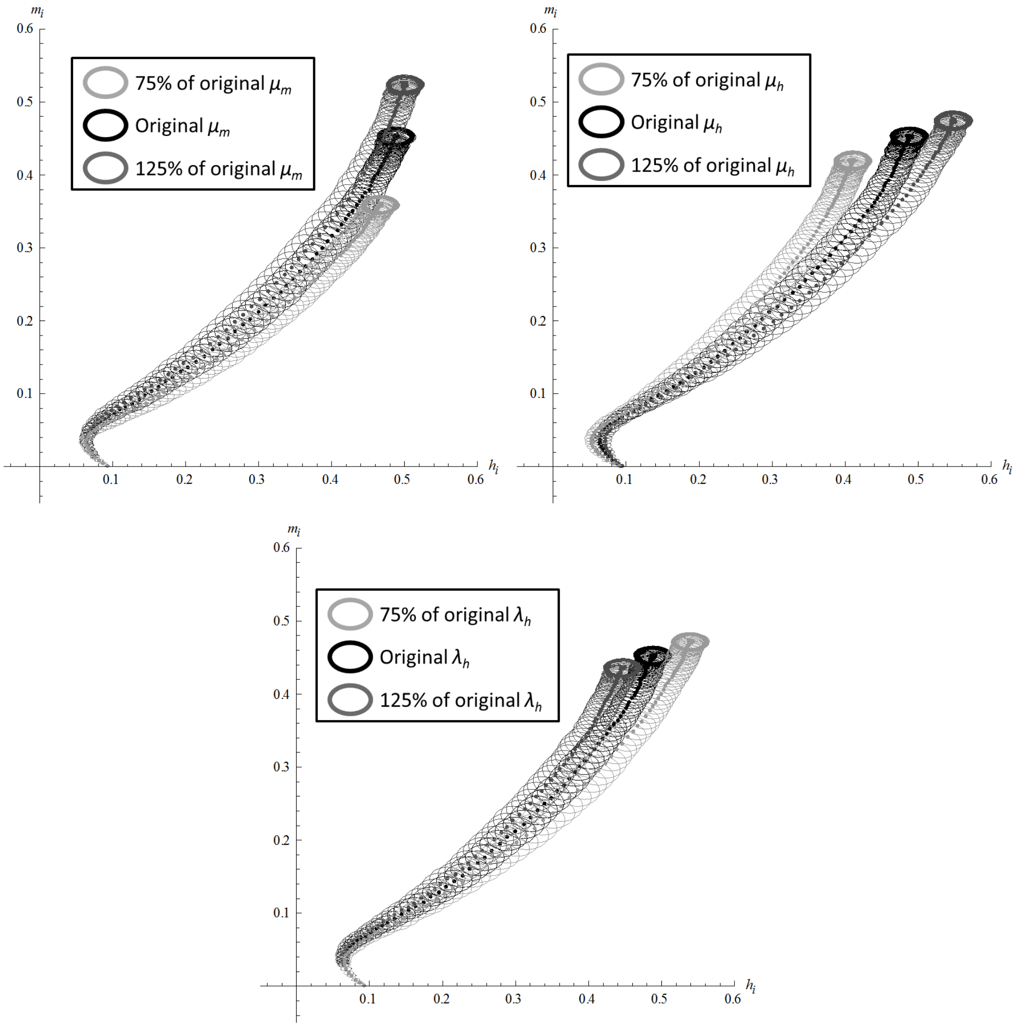


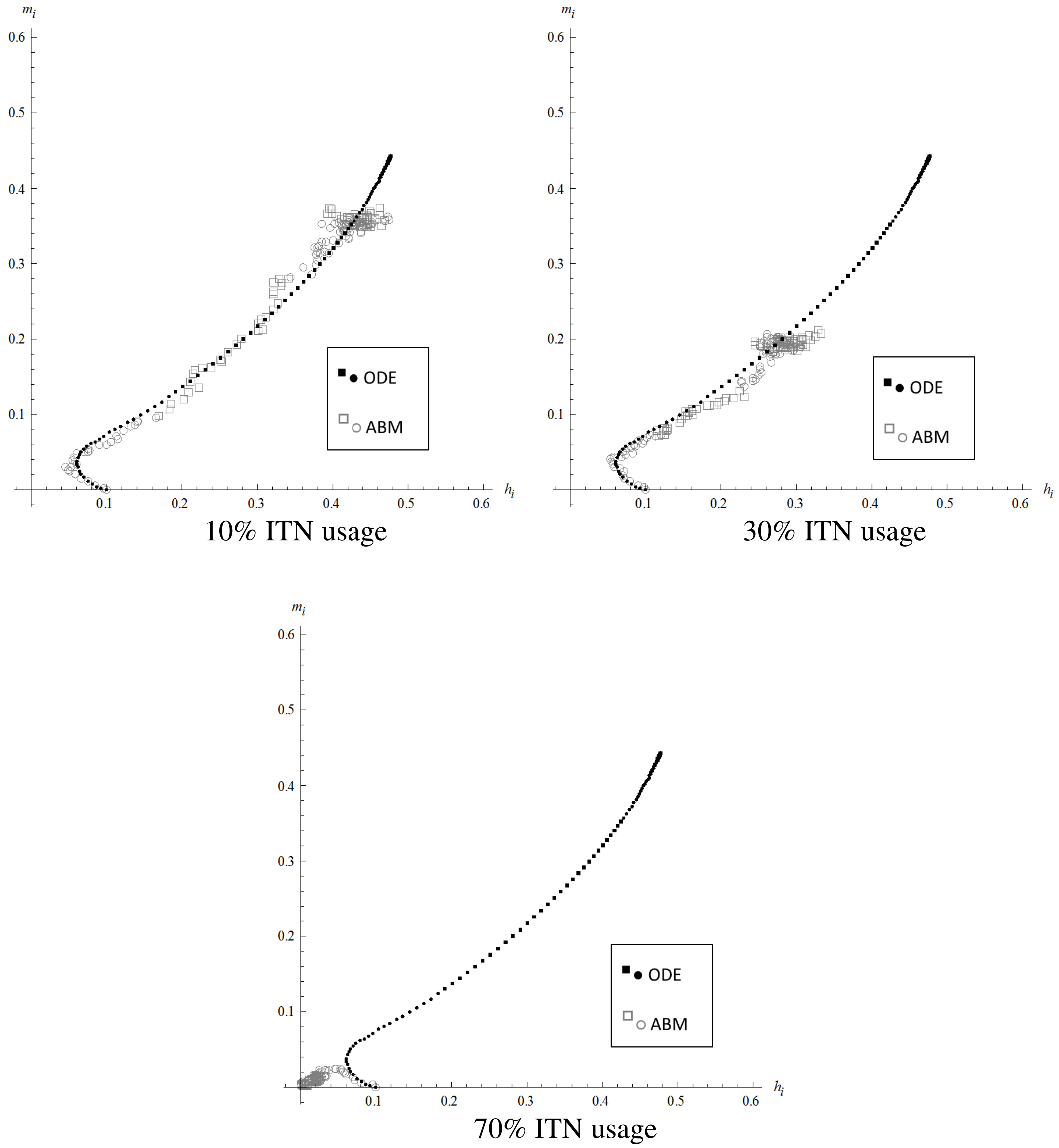
Figure 2. Plots of output from the ABM simulation for varying values of μ_m (top left), μ_h (top right), and λ_h (bottom). The simulations are based on a time span of six months and used averages from 500 runs of the simulation. The ellipses surrounding each point represent potential deviation from equilibrium due to randomness of the simulation. The black graph uses the value of the investigated parameter from Table 1, the light gray graph uses 75% of this value, and the dark gray graph uses 125% of this value. Parameters other than the investigated parameter were defined as in Table 1. The horizontal axis represents proportions of infectious humans in the simulated population, and the vertical axis represents proportions of infectious mosquitoes in the simulated population.

humans had bed nets. The corresponding output for the ODE model is also presented in the plots in Figure 3 to show the decreasing proportion of infected individuals with increasing ITN usage. The ABM predictions in Figure 3 are for individual runs of the simulation and have not been averaged since they are intended as examples of simulation output.

In order to demonstrate the overall trend of infection with increasing ITN usage in the current study, the average equilibrium points (h_i, m_i) for the ABM simulation were calculated for each proportion of bed net use by humans, from 1 to 100 percent. Again, several replicates should be simulated in order to describe the overall system [MacDonald et al. 1968]. The ABM was run 500 times to allow for variation with random numbers used in the simulation. A time period of two years was used in order to allow for more settling to the equilibrium points. The average equilibrium points were calculated using the ABM output for h_i and m_i in the last 500 days of each two-year simulation run, and the equilibrium points were averaged across various simulations. The simulation appears to settle after six months; hence, the output in the last 500 days of the two-year run represents output after the simulation has localized. Successive equilibrium points from 1 percent to 100 percent bed net usage are displayed in Figure 4 along with ellipses representing the error due to variability. The rightmost point on the plot represents the average equilibrium point with 1 percent of humans using bed nets, and this percentage increases moving right to left on the plot. The ellipses are centered around the mean equilibrium coordinates for humans and mosquitoes using major and minor axes with lengths of two standard deviations. The ellipses are presented to illustrate the variability in the individual simulation runs.

3.3. ***Acquired immunity.*** Acquired immunity is gained through repeated exposure to the malaria parasite, and the effects of acquired immunity have been previously modeled [Aron 1983; Chiyaka et al. 2007; Gu et al. 2003; Gurarie and McKenzie 2007; Milligan and Downham 1996; Tumwiine et al. 2007]. Previous models have considered acquired immunity as leading to milder forms of the disease [Tumwiine et al. 2007] and have defined acquired immunity in a host as protection against severe illness [Chiyaka et al. 2007]. Chiyaka et al. [2007] also asserted that while this immunity may be beneficial to the individual, these immune individuals disrupt the control strategies for the disease. When infection is mild the infected person may not seek medical attention which allows susceptible mosquitoes to become infected and spread the disease to other susceptible human hosts.

In modeling the spread of acquired immunity in the current ABM, human agents were assumed to gain immunity after a certain number of infections as in [Milligan and Downham 1996]. Once a host has reached a certain level of infections, the host was assumed to lose immunity at a particular rate (if not reinfected) [Aron



10% ITN usage

30% ITN usage

70% ITN usage

Figure 3. Plot of output from the ABM simulation incorporating ITN usage, shown by the open gray shapes. The solid black shapes represent output from the ODE model with latency as described in Section 2.1, shown for comparison. The simulations were generated over approximately 6 months (180 days). The horizontal axis represents proportions of infectious humans in the simulated population, and the vertical axis represents proportions of infectious mosquitoes in the simulated population. The symbols change from one shape to another at the end of each 30 day period.

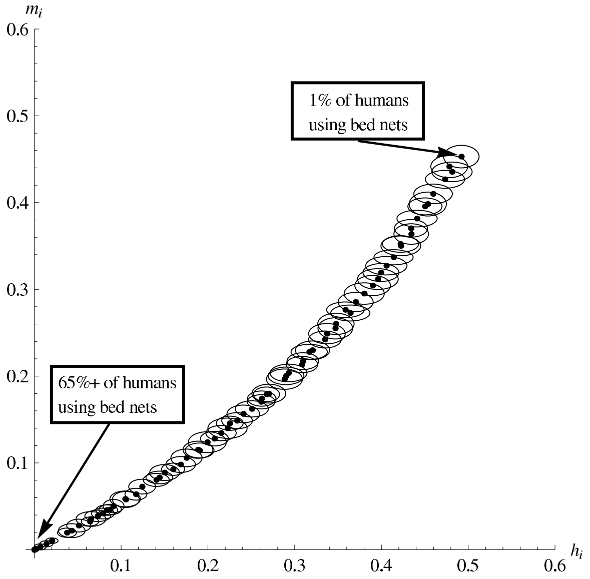


Figure 4. For each proportion of humans using ITNs, a settling point of the ABM was calculated based on an average of 500 simulation runs over a 2 year period. The equilibrium point for 1% ITN use is found on the top right with nearly 50% infectivity. As ITN usage increases, the equilibrium points trend toward (0, 0) or no infection. The ellipses surrounding each point represent potential deviation from equilibrium due to randomness of the simulation. The ellipses surrounding each point represent potential deviation from equilibrium due to randomness of the simulation.

1983; Milligan and Downham 1996]. Gu et al. [2003] investigated an ABM with acquired immunity, but their model did not incorporate latency and did allow for superinfection. Although acquired immunity could be described using only the number of infections a person has experienced as in [Gu et al. 2003], the ABM in the current study was constructed to model immunity as time-dependent.

A third characteristic is added to the array for humans in the ABM that represents a quantitative measure of acquired immunity. Each exposure to the disease is expected to add to this quantitative measure; once a certain level of exposure to malaria is reached, acquired immunity begins. In the ABM for acquired immunity a person is assumed to resist infection to malaria after roughly three infections. Each time the characteristic array of the human is changing from susceptible to latent, I_{ex} (which was set to 30) arbitrary units are added to the immunity characteristic. The quantity of this characteristic at which a person is expected to be immune I_c

is set to 70. Hence, the simulations allow for a person to be immune after three infections in a short time period.

Once a person has acquired immunity, the immunity is assumed to decrease with time. Once the human has returned to the susceptible state, one unit of immunity is lost as each iterate (or day) passes without another infection. Resistance to the disease is treated discretely; a person will not become latent if immunity is 70 or above and a person is equally susceptible to infection if immunity is anything between 0 and 69. Although acquired immunity may lead to milder forms of disease, the ABM was designed (in this initial investigation) to model immunity very simply. Individuals are also assumed to have a limit to the amount of immunity they can acquire, and the maximum quantity of immunity, I_{max} is set to 100. A more detailed description of the process of the ABM with acquired immunity is presented in Appendix A.3.2.

ABM simulation results incorporating acquired immunity are presented in Figure 5, and the numerical solution to the ODE model described in Section 2.1 is plotted for comparison. The ABM simulation does not appear to settle into an equilibrium point in the first six months, and ABM simulations were produced over two-year time periods to allow for more settling to potentially identify an equilibrium. As with the sensitivity analysis and the investigation into ITN usage, an overall trend was desired in line with suggestions from [MacDonald et al. 1968]. Figure 5 contains average model output over 500 simulations that were run for a two-year time period. Each ellipse in Figure 5 represents the average value found for that point after 500 runs of the simulation plus one standard deviation in the direction of the major and minor axes.

4. Discussion

The ABM of malaria transmission makes very similar predictions to the ODE model based on the work of Ross and MacDonald [1957] altered to account for latency. The ABM in the current study is similar in concept to the computational framework described in [MacDonald et al. 1968] with the addition of latency and without seasonal effects. The ABM is intended to be simple and straightforward and provides a convenient way to add complexity in modeling malaria transmission, such as incorporating the effects of bed nets and immunity. Since the output of the ODE model and the ABM are similar, the ABM simulations likely also have asymptotic behavior around the nonzero equilibrium point or endemic equilibrium. Strategies of reducing the spread of malaria could be determined by investigating changes that move the settling point of the ABM closer to the origin. The basic ODE model had an asymptotically stable endemic equilibrium, and all simulations of the ABM became localized after some length of time. By describing the populations in

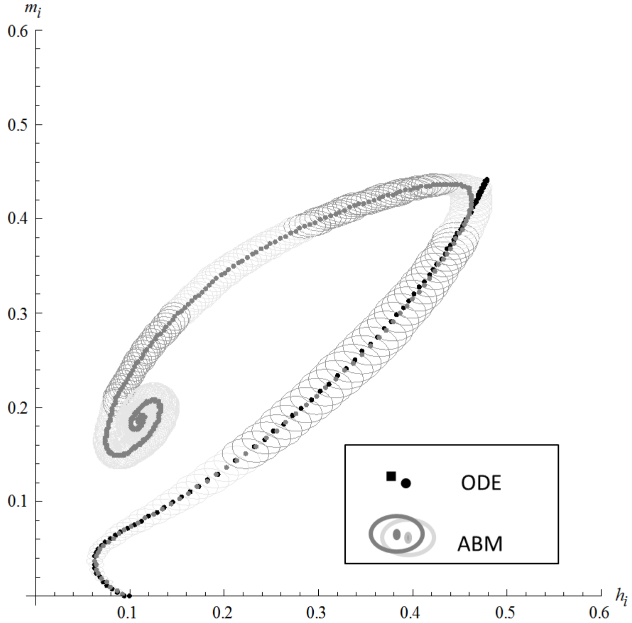


Figure 5. Plot of output from the ABM simulation incorporating acquired immunity based on a time span of 2 years and using averages from 500 runs of the simulation. The averages of model output of the two year period are shown by the gray dots and surrounding lighter gray ellipses. The ellipses change from one gray shade to another at the end of each 30 day period. The black dots represent output from the ODE model with latency as described in Section 2.1, shown for reference. The ellipses surrounding each point represent potential deviation from equilibrium due to randomness of the simulation. The horizontal axis represents proportions of infectious humans in the simulated population, and the vertical axis represents proportions of infectious mosquitoes in the simulated population.

terms of individual agents, the ABM may be better constructed to deal with disease eradication [MacDonald et al. 1968].

The addition of bed nets into the ABM demonstrated a clear decrease in infection compared with the base model. More specifically, increasing the amount of bed net usage resulted in a reduction in infection prevalence in both human and mosquito populations. Furthermore, infection is nearly eliminated from both human and mosquito populations if only 70% of the human population is protected with bed nets. This leads to the conclusion that bed nets, when used among the majority of a population, have the power of protecting more than just the individuals sleeping directly under them. Hence, the results suggest that protecting a given threshold of individuals within a population extends disease protection to everyone in that population. The ABM simulation did assume that once an ITN was put in place, the individual using it would continue to use the bed net, which may need to be more fully considered.

Additionally, the ABM is intended as a base framework for miscellaneous investigations; and through modification of the ABM, one could easily study the effects of bed nets combined with other considerations (seasonality, etc.). As described in Appendix A.3.2, the ABM simulation was modified by the addition of an element to the individual's characteristic array which indicates the presence of a bed net and a step to check if the ITN prevents a mosquito from biting a human. The use of ITNs can be incorporated into an ODE model, but the purpose of this investigation is not only to find specific results but also to illustrate an example of how the ABM can be easily adapted.

As also discussed in Appendix A.3.2, acquired immunity was modeled in the ABM through an element in the agent's characteristic array, and this element was affected only by infection history and time. Hence, the ABM did not require computational solving of a nonlinear system. A reduction in the proportion of infectious individuals was expected due to acquired immunity, and the ABM of malaria transmission was relatively easy to alter to incorporate this complicated issue of malaria transmission. As with the ABM investigation involving ITNs, the investigation incorporating acquired immunity is intended as an example of how the ABM can be easily modified for issues surrounding malaria or other diseases.

The ABM simulations incorporating acquired immunity show spiraling behavior as time increases as shown in Figure 5, and this model behavior may be worth investigating further. The description of acquired immunity may require additional sophistication in the ABM as the results presented in this study were based on the idea that an individual was either completely immune or completely susceptible, which is oversimplified and was warned against in [Gurarie and McKenzie 2007]. A probabilistic approach could be used to describe an agent as less likely (but not completely immune) to contract the disease after repeated exposures. The structure of the ABM is not currently constructed to investigate questions such as how mildly affected individuals' not seeking treatment would change transmission dynamics (as mentioned in [Chiyaka et al. 2007]), but the ABM could be modified to do so through larger arrays describing the agents. Additionally, combining the dynamics of ITN usage and acquired immunity in the ABM may provide even greater insight into the effect of preventative measures in the population. The results of a model combining these issues may suggest an even lower percentage of ITN usage is needed for a significant reduction in infection.

The ABM for malaria transmission does not currently incorporate spatial considerations. The interactions are all treated randomly, and the agents are assumed to be distributed in such a way that random interaction is reasonable. However, where humans and mosquitoes are located in an area likely does affect which humans are selected as blood-meal targets if the distributions of either group are not uniform. The ABM could be adapted to describe smaller populations within the larger community in order to account for this in some way. An additional characteristic could easily be added to represent a finite number of locations that have different parameters and interaction details associated with them.

Although the predictions from the ABM and the ODE model are similar in shape, they do differ somewhat with time. The predictions are not equal at each time point with the ABM simulations moving toward the settling point more quickly than the numerical solution of the ODE system. The reason for the time difference has not yet been identified and may be a focus for future work. This is an important issue since the time difference could have implications on what public health professionals should expect from real-world dynamics.

Many aspects of malaria transmission have not been incorporated into the ODE model or the ABM. Seasonality, climate, emigration, and other environmental factors are assumed to not have an impact on the model, but these factors do all affect the spread of the disease. Some of the simulations are computed over two years, which does involve seasonal changes; therefore, assuming climate has no affect is an oversimplification. However, the ABM uses a structure that is easily adaptable to different situations, and many of these aspects should be relatively straightforward to include in the ABM. In order to establish the ABM as a reasonable vehicle to describe malaria transmission, a basic ABM was created to compare to a fairly simple ODE model. The similarity of the two models has been established, and future work will incorporate greater complexity. Additional studies may also consider the comparison between more complicated models of both structures.

Appendix: ODD protocol

Various techniques and methods have been used in simulation models involving ABMs, and [Grimm et al. 2006] presents the overview, design concepts, and details (ODD) protocol for describing ABMs. The following is a description of the malaria transmission ABM used in this study based on the ODD protocol.

A.1. *ODD protocol: overview.*

A.1.1. *Purpose.* The purpose of the ABM malaria model in this study is to describe malaria infection on a population level in order to better understand transmission of the disease. Additionally, the framework of this ABM allows for fairly easy additions of very complicated factors of malaria spread.

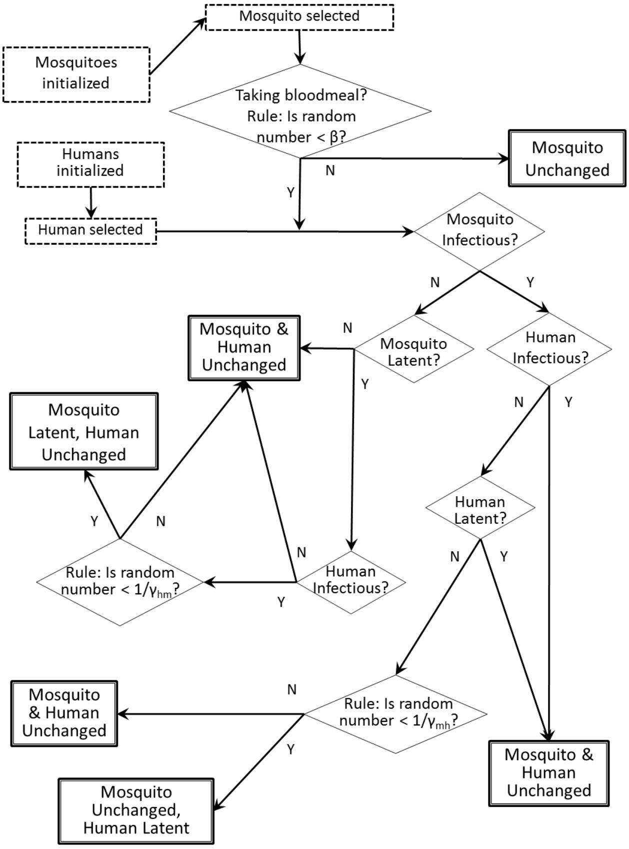

Mosquito selected
Mosquitoes initialized
Taking bloodmeal? Rule: Is random number < β?
Humans initialized
N
Mosquito Unchanged
Y
Human selected
Mosquito Infectious?
N
Y
Mosquito & Human Unchanged
N
Mosquito Latent?
Human Infectious?
Y
N
Y
Mosquito Latent, Human Unchanged
Human Latent?
Y
N
N
Y
Rule: Is random number < 1/γhm?
Human Infectious?
N
Y
Y
Mosquito & Human Unchanged
N
Rule: Is random number < 1/γmh?
Y
Mosquito & Human Unchanged
Mosquito Unchanged, Human Latent

Figure 6. Flow chart showing the ABM rules of malaria transmission during a blood meal. This chart describes how agents move from the susceptible stage to the latent stage, but does not account for the modeling of an agent's move from the latent stage to the infectious stage. The continued process of the ABM is shown in Figure 7. The notation used for parameters is the same as in Section 2.1 and the parameter values used are presented in Table 1.

A.1.2. *State variables.* The agents in the basic ABM are mosquitoes and humans in closed populations. Each agent has two characteristics, whether the individual is latent and whether the individual is infectious. The characteristics are indicated using 0 and 1; 0 indicates the individual does not have the particular characteristic (latent or infectious), and 1 indicates the individual does have the characteristic.

When modeling ITN usage, a third characteristic is added to the human individuals indicating whether or not the individual was given a bed net at the beginning of the simulation. The third number in the individual's characteristic array would

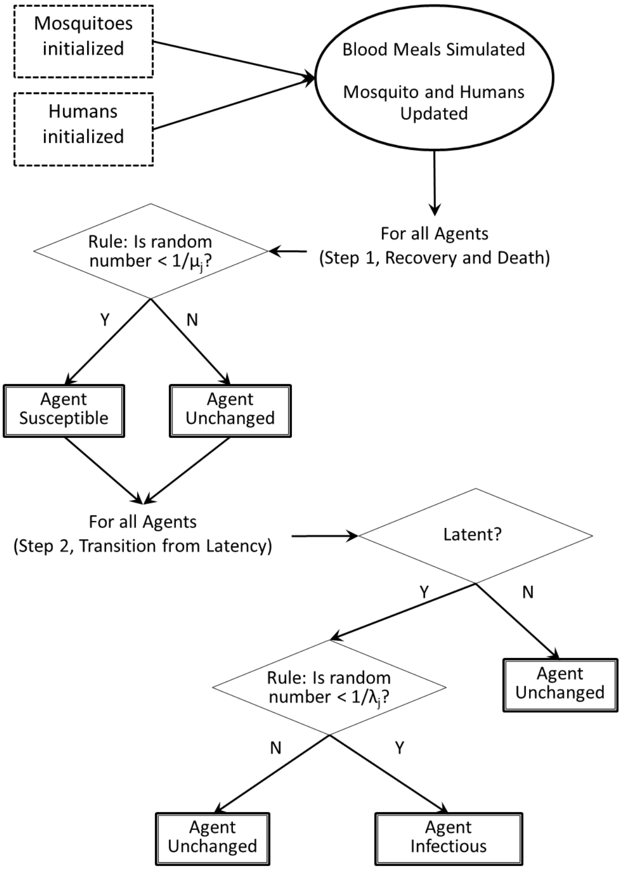


Figure 7. Flow chart showing the ABM rules defining transitions from latency to the infectious state and returning from the infectious state to the susceptible state. The details involved in the oval for blood meals is expanded in Figure 6. The subscript j is either m or h depending on whether the agent is a mosquito or a human. The notation used for parameters is the same as in Section 2.1 and the parameter values used are presented in Table 1.

then be 1 if the individual was using an ITN and 0 if not. In investigations with immunity, an third characteristic is also added to the human individuals indicating a level of acquired immunity which will be described more fully in Appendix A.3.2.

A.1.3. *Process overview.* The primary interaction modeled in the ABM for malaria transmission is the contact between a mosquito taking a blood meal and the individual human being bitten. Time is treated discretely, using steps of days. The beginning process of the ABM is outlined in Figure 6 showing how the individual biting events are modeled. The process of agents' moving to latency is presented in

Figure 7. The process in Figure 7 applies to both mosquito and human populations, with the rules based on appropriate parameters. The subscript j in the diagram indicates m or h depending on whether the agent is a mosquito or a human. Although the process is the same for both types of agents, the left side of the diagram (step 1) has different meaning depending on the type of agent. Since the population of mosquitoes is assumed to remain constant, if a mosquito is selected to "die" (the selected random number is less than $1/\mu_m$), then a new mosquito essentially takes its place by changing the characteristic array to describe a susceptible mosquito. The parameter μ_h indicates the recovery time of humans, so the change of the characteristic array to indicate susceptibility is assumed to (most likely) be describing the same human individual who has now recovered from the disease. The notation used for parameters is the same as in Section 2.1 and the parameter values used are presented in Table 1. All random numbers are selected from a uniform distribution between 0 and 1.

A.2. ***ODD protocol: design concepts.*** The interaction between mosquitoes and humans is modeled explicitly, and all rules of the ABM are based on probabilities. Mosquitoes and humans from the list or array of agents were chosen randomly and proximity or location was not incorporated.

A.3. *ODD protocol: details.*

A.3.1. *Initialization.* The parameters in Table 1 were initialized and kept fixed throughout the simulations of the ABM. The total number of humans simulated was 500 except in a few simulations investigating sensitivity as described in Section 3.1. (Note that since the ratio of mosquitoes to humans is a fixed parameter in the model, defining the number of humans also defines the total number of mosquitoes.) The initial proportion of infectious humans was set to be 10%, and the initial proportion of infectious mosquitoes was set to be 0%. No agents were initialized in the latent stage. All simulations used the same initial proportions.

The individual simulations involving ITNs and acquired immunity did involve more initialized values. The details of those simulations are presented in Appendix A.3.2.

A.3.2. *Submodels.*

Insecticide-treated bed nets. The ABM for malaria transmission was adapted to predict the spread of the disease when insecticide-treated bed nets are used. The process of simulated malaria transmission when bed nets are used is shown in Figure 8. In the initialization of the simulations with ITN usage, the percentage of how many humans would be using bed nets was set and fixed for the rest of that simulation. To ascertain the effect of ITNs on the spread of malaria, different simulations were varied using percentages of ITN usage. The parameter δ was not

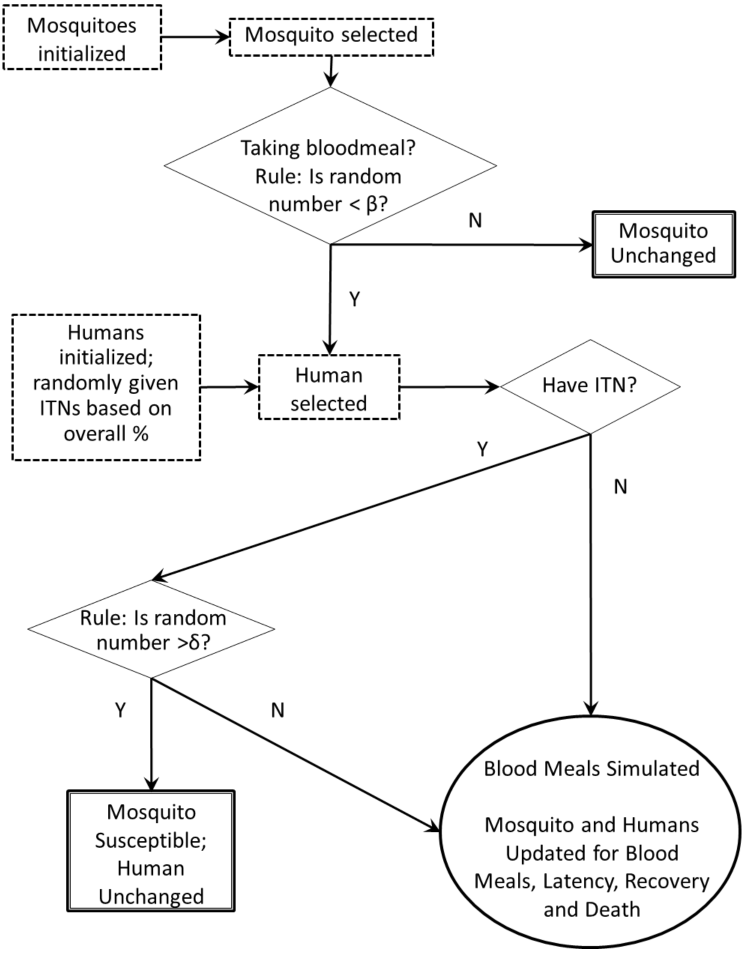

Mosquitoes initialized
Mosquito selected
Taking bloodmeal? Rule: Is random number < β?
N
Mosquito Unchanged
Y
Humans initialized; randomly given ITNs based on overall %
Human selected
Have ITN?
Y
N
Rule: Is random number >δ?
Y
N
Mosquito Susceptible; Human Unchanged
Blood Meals Simulated
Mosquito and Humans Updated for Blood Meals, Latency, Recovery and Death

Figure 8. Flow chart showing the ABM rules defining malaria transmission when bed nets are used by the human population. The notation used for parameters is the same as in Section 2.1 and the parameter values used are presented in Table 1 except for the parameter δ indicating the probability that a mosquito will survive once it tries to bite a human using a bed net.

in either original model and represents the probability that a mosquito will survive once it tries to take a blood meal from a human using a bed net. Since bed net usage was expected to be 96% effective (see Section 3.2), δ was set to be 0.04 for all simulations. After the simulation checks if the mosquito dies while trying to take a blood meal, the simulation proceeds as described in Section A.1.3. As in the basic simulation, if an agent "dies," its characteristic array is changed to represent

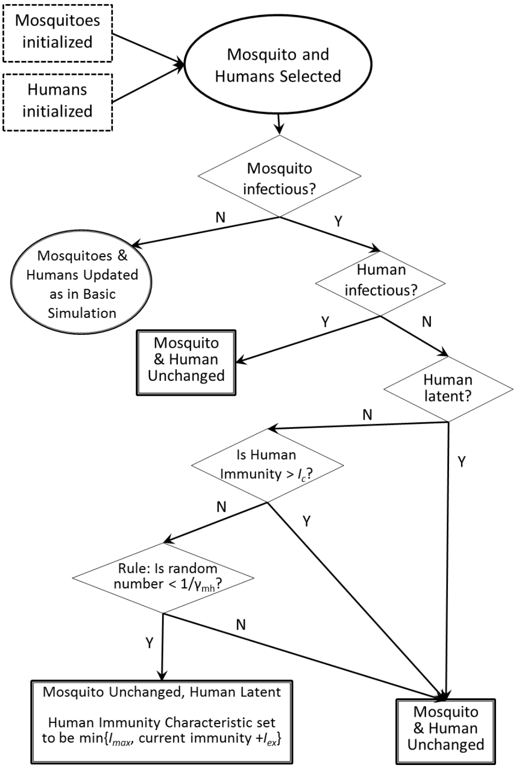

Mosquitoes initialized
Humans initialized
Mosquito and Humans Selected
Mosquito infectious?
N
Y
Mosquitoes & Humans Updated as in Basic Simulation
Human infectious?
Y
N
Mosquito & Human Unchanged
Human latent?
N
Y
Is Human Immunity > I_c?
N
Y
Rule: Is random number < $1/\gamma_{mh}$?
N
Y
Mosquito Unchanged, Human Latent
Human Immunity Characteristic set to be min{I_{max}, current immunity +I_{ex}}
Mosquito & Human Unchanged

Figure 9. Flow chart showing the ABM rules defining malaria transmission when human agents may be able to acquire immunity to the disease. The parameter I_c indicates the numerical immunity that is necessary to avoid infection, I_{ex} indicates the addition to the immunity characteristic when the human is successfully infected, and I_{max} is the largest allowable value for the immunity characteristic. Otherwise, the notation used for parameters is the same as in Section 2.1 and the parameter values used are presented in Table 1. This flow chart does not include the entire ABM simulation for malaria transmission with acquired immunity as it does not show steps modeling the move from latency to being infectious, modeling recovery or death, or modeling how individual immunity decreases.

a newly born, susceptible agent. The oval in Figure 8 contains all steps shown in Figure 6 and Figure 7.

Acquired immunity. In order to model malaria transmission, a third element was added to the agent characteristic array. This third characteristic quantifies the immunity of the individual human agent. A flow chart describing the ABM with malaria transmission when acquired immunity is incorporated is presented in Figure 9. Each time a human is successfully infected (which means that the human moves to the latent stage) an immunity exposure value I_{ex} is added to the immunity characteristic of that human. Human agents are expected to have a maximum to the immunity they can obtain; therefore, the human immunity characteristic is limited by a maximum value I_{max}. Once a human has immunity above a critical value I_c, that individual will be protected from transmission from an infectious mosquito. The flow chart in Figure 9 only shows the portion of the simulation through the blood-meal process, similar to the basic simulation flow shown in Figure 6. The simulations for the ABM with acquired immunity also proceed through the steps outlined in Figure 7. Additionally, each iteration of the simulation represents one day, and the immunity characteristic (if nonzero) of susceptible humans decreases by 1 with each iteration. In all simulations, we use $I_c = 70$, $I_{\text{ex}} = 30$, and $I_{\text{max}} = 90$ as is described more fully in Section 3.3.

Acknowledgements

The authors would like to thank Dr. Crista Arangala for helpful comments and discussions throughout the investigation.

References

[Anderson and May 1991] R. Anderson and R. May, *Infectious diseases of humans: dynamics and control*, Oxford University Press, 1991.

[Aneke 2002] S. J. Aneke, "Mathematical modelling of drug resistant malaria parasites and vector populations", *Math. Methods Appl. Sci.* **25**:4 (2002), 335–346. MR 2002m:92031 Zbl 0994.92025

[Aron 1983] J. L. Aron, "The dynamics of immunity boosted by exposure to infection", *Math. Biosci.* **64**:2 (1983), 249–259. Zbl 0515.92026

[Aron 1988] J. L. Aron, "Mathematical modelling of immunity to malaria", *Math. Biosci.* **90**:1-2 (1988), 385–396. MR 89f:92032 Zbl 0651.92018

[Aron and May 1982] J. L. Aron and R. May, "The population dynamics of malaria", pp. 139–179 in *The population dynamics of infectious disease: theory and applications*, edited by R. Anderson, Chapman and Hall, London, 1982.

[Castiglione et al. 2007] F. Castiglione, K. Duca, A. Jarrah, R. Laubenbacher, D. Hochberg, and D. Thorley-Lawson, "Simulating Epstein–Barr virus infection with C-ImmSim", *Bioinformatics* **23**:11 (2007), 1371–1377.

[CDC 2011] Centers for Disease Control and Prevention, "Drugs used in the prophylaxis of malaria", 2011. See http://wwwnc.cdc.gov/travel/yellowbook/2012/chapter-3-infectious-diseases-related-to-travel/malaria#1939.

[CDC 2012a] Centers for Disease Control and Prevention, "Malaria", 2012, http://www.cdc.gov/malaria.

[CDC 2012b] Centers for Disease Control and Prevention, "Malaria: biology", 2012, http://www.cdc.gov/malaria/about/biology.

[CDC 2012c] Centers for Disease Control and Prevention, "Malaria: frequently asked questions", 2012, http://www.cdc.gov/malaria/about/faqs.html.

[Chitnis et al. 2008] N. Chitnis, J. M. Hyman, and J. M. Cushing, "Determining important parameters in the spread of malaria through the sensitivity analysis of a mathematical model", *Bull. Math. Biol.* **70**:5 (2008), 1272–1296. MR 2009g:92111 Zbl 1142.92025

[Chiyaka et al. 2007] C. Chiyaka, W. Garira, and S. Dube, "Transmission model of endemic human malaria in a partially immune population", *Math. Comput. Model.* **46**:5-6 (2007), 806–822. MR 2008b:92131 Zbl 1126.92043

[Chiyaka et al. 2008] C. Chiyaka, J. M. Tchuenche, W. Garira, and S. Dube, "A mathematical analysis of the effects of control strategies on the transmission dynamics of malaria", *Appl. Math. Comput.* **195**:2 (2008), 641–662. MR 2008m:92052 Zbl 1128.92022

[Chiyaka et al. 2009] C. Chiyaka, W. Garira, and S. Dube, "Effects of treatment and drug resistance on the transmission dynamics of malaria in endemic areas", *Theor. Popul. Biol.* **75** (2009), 14–29. Zbl 1210.92005

[Curtis et al. 1992] C. F. Curtis, J. Myamba, and T. J. Wilkes, "Various pyrethroids on bednets and curtains", *Mem. Inst. Oswaldo Cruz* **87**:Supplement 3 (1992), 363–370.

[Daley and Gani 1999] D. J. Daley and J. Gani, *Epidemic modelling: an introduction*, Cambridge Studies in Mathematical Biology **15**, Cambridge University Press, 1999. MR 2000e:92042 Zbl 0922.92022

[De Zoysa et al. 1991] A. P. K. De Zoysa, C. Mendis, A. C. Gamage-Mendis, S. Weerasinghe, P. R. J. Herath, and K. N. Mendis, "A mathematical model for *Plasmodium vivax* malaria transmission: estimation of the impact of transmission-blocking immunity in an endemic area", *B. World Health Organ.* **69**:6 (1991), 725–734.

[Dembele et al. 2009] B. Dembele, A. Friedman, and A.-A. Yakubu, "Malaria model with periodic mosquito birth and death rates", *J. Biol. Dyn.* **3**:4 (2009), 430–445. MR 2011g:34094

[Eubank et al. 2004] S. Eubank, H. Guclu, V. S. Anil Kumar, M. V. Marathe, A. Srinivasan, Z. Toroczkai, and N. Wang, "Modelling disease outbreaks in realistic urban social networks", *Nature* **429**:6988 (2004), 180–184.

[Grimm et al. 2006] V. Grimm, U. Berger, F. Bastiansen, S. Eliassen, V. Ginot, J. Giske, J. Goss-Custard, T. Grand, S. K. Heinz, G. Huse, A. Huth, J. U. Jepsen, C. Jørgensen, W. M. Mooij, B. Müller, G. Pe'er, C. Piou, S. F. Railsback, A. M. Robbins, M. M. Robbins, E. Rossmanith, N. Rüger, E. Strand, S. Souissi, R. A. Stillman, R. Vabø, U. Visser, and D. L. DeAngelis, "A standard protocol for describing individual-based and agent-based models", *Ecol. Model.* **198**:1–2 (2006), 115–126.

[Gu et al. 2003] W. Gu, G. F. Killeen, C. M. Mbogo, J. L. Regens, J. I. Githure, and J. C. Beier, "An individual-based model of *Plasmodium falciparum* malaria transmission on the coast of Kenya", *Trans. R. Soc. Trop. Med. Hyg.* **97**:1 (2003), 43–50.

[Gurarie and McKenzie 2007] D. Gurarie and F. E. McKenzie, "A stochastic model of immune-modulated malaria infection and disease in children", *Math. Biosci.* **210**:2 (2007), 576–597. MR 2008k:92056 Zbl 1134.92024

[Hinkelmann et al. 2011] F. Hinkelmann, D. Murrugarra, A. S. Jarrah, and R. Laubenbacher, "A mathematical framework for agent based models of complex biological networks", *Bull. Math. Biol.* **73**:7 (2011), 1583–1602. MR 2012f:92003 Zbl 1225.92001

[Koella 1991] J. C. Koella, "On the use of mathematical models of malaria transmission", *Acta Tropica* **49**:1 (1991), 1–25.

[Koella and Antia 2003] J. C. Koella and R. Antia, "Epidemiological models for the spread of anti-malarial resistance", *Malaria J.* **2**:3 (2003).

[MacDonald 1957] G. MacDonald, *The epidemiology and control of malaria*, Oxford University Press, London, 1957.

[MacDonald et al. 1968] G. MacDonald, C. B. Cuellar, and C. V. Foll, “The dynamics of malaria”, *B. World Health Organ.* **38**:5 (1968), 743–755.

[Maire et al. 2006] N. Maire, T. Smith, A. Ross, S. Owusu-Agyei, K. Dietz, and L. Molineaux, “A model for natural immunity to asexual blood stages of *Plasmodium falciparum* malaria in endemic areas”, *Amer. J. Trop. Med. Hyg.* **75**:Supplement 2 (2006), 19–31.

[Milligan and Downham 1996] P. J. M. Milligan and D. Y. Downham, “Models of superinfection and acquired immunity to multiple parasite strains”, *J. Appl. Probab.* **33**:4 (1996), 915–932. MR 97g:92021 Zbl 0871.92025

[Nájera 1974] J. A. Nájera, “A critical review of the field application of a mathematical model of malaria eradication”, *B. World Health Organ.* **50**:5 (1974), 449–457.

[Nedelman 1985] J. Nedelman, “Introductory review: some new thoughts about some old malaria models”, *Math. Biosci.* **73**:2 (1985), 159–182. MR 86d:92022 Zbl 0567.92020

[N’Guessan et al. 2001] R. N’Guessan, F. Darriet, J. M. C. Doannio, F. Chandre, and P. Carnevale, “Olyset Net® efficacy against pyrethroid-resistant *Anopheles gambiae* and *Culex quinquefasciatus* after 3 years’ field use in Côte d’Ivoire”, *Med. Vet. Entomol.* **15**:1 (2001), 97–104.

[Ngwa 2006] G. A. Ngwa, “On the population dynamics of the malaria vector”, *Bull. Math. Biol.* **68**:8 (2006), 2161–2189. MR 2007k:92119

[Pogson et al. 2006] M. Pogson, R. Smallwood, E. Qwarnstrom, and M. Holcombe, “Formal agent-based modelling of intracellular chemical interactions”, *BioSystems* **85**:1 (2006), 37–45.

[Ross 1910] R. Ross, *The prevention of malaria*, Dutton, New York, 1910.

[Shililu et al. 1998] J. I. Shililu, W. A. Maier, H. M. Seitz, and A. S. Orago, “Seasonal density, sporozoite rates, and entomological inoculation rates of *Anopheles gambiae* and *Anopheles funestus* in a high-altitude sugarcane growing zone in Western Kenya”, *Trop. Med. Int. Health* **3**:9 (1998), 706–710.

[Smith et al. 2006] T. Smith, G. F. Killeen, N. Maire, A. Ross, L. Molineaux, F. Tediosi, G. Hutton, J. Utzinger, K. Dietz, and M. Tanner, “Mathematical modeling of the impact of malaria vaccines on the clinical epidemiology and natural history of *Plasmodium falciparum* malaria: overview”, *Amer. J. Trop. Med. Hyg.* **75**:Supplement 2 (2006), 1–10.

[Smith et al. 2007] D. L. Smith, F. E. McKenzie, R. W. Snow, and S. I. Hay, “Revisiting the basic reproductive number for malaria and its implications for malaria control”, *PLoS Biol.* **5**:3 (2007), 531–542.

[Spielman and D’Antonio 2001] A. Spielman and M. D’Antonio, *Mosquito: a natural history of our most persistent and deadly foe*, Hyperion, New York, 2001.

[Tumwiine et al. 2007] J. Tumwiine, J. Y. T. Mugisha, and L. S. Luboobi, “A mathematical model for the dynamics of malaria in a human host and mosquito vector with temporary immunity”, *Appl. Math. Comput.* **189**:2 (2007), 1953–1965. MR 2332148 Zbl 1117.92039

[Wang et al. 2009] Z. Wang, C. M. Birch, J. Sagotsky, and T. S. Deisboeck, “Cross-scale, cross-pathway evaluation using an agent-based non-small cell lung cancer model”, *Bioinformatics* **25**:18 (2009), 2389–2396.

[World Book 2008] World Book, “Mosquito”, in *The World Book Encyclopedia*, World Book, Chicago, 2008.

[Yang and Ferreira 2000] H. Yang and M. Ferreira, “Assessing the effects of global warming and local social and economic conditions on malaria transmission”, *Rev. Saúde Públ.* **34**:3 (2000), 214–222.

Received: 2012-05-04 Revised: 2013-02-08 Accepted: 2013-05-23

kyokley@elon.edu *Department of Mathematics and Statistics, Elon University, Elon, NC 27244, United States*

tlee@elon.edu *Department of Mathematics and Statistics, Elon University, CB 2320, Elon, NC 27244, United States*

akbrown19@gmail.com *University of Texas Health Science Center at Houston, Houston, TX 77030, United States*

mchristina.minor@gmail.com *Fitts Department of Industrial and Systems Engineering, North Carolina State University, Raleigh, NC 27695, United States*

gmader16@gmail.com *Department of Mathematics, North Carolina State University, Raleigh, NC 27695, United States*

INVOLVE 7:1 (2014)
dx.doi.org/10.2140/involve.2014.7.41

Slide-and-swap permutation groups

Onyebuchi Ekenta, Han Gil Jang and Jacob A. Siehler

(Communicated by Joseph A. Gallian)

We present a simple tile-sliding game that can be played on any 3-regular graph, generating a permutation group on the vertices. We classify the resulting permutation groups and obtain a novel presentation for the simple group of 168 elements.

From sliding tiles to simple groups

The sliding tiles of the notorious "fifteen puzzle" (arranged in a 4×4 array with one square missing) are an object lesson in parity: Which permutations of the numbered tiles can be achieved? Precisely the even permutations. Put another way, the moves of the fifteen puzzle generate the alternating group A_{15}. Aaron Archer [1999] gives us a tidy proof of this folkloric fact.

R. M. Wilson [1974] considers tile-sliding games on arbitrary graphs as a generalization of the fifteen puzzle, and classifies the permutation groups which can be generated by these games. Briefly, the permutation group for a tile-sliding game on a graph with k vertices is generally either the alternating group A_{k-1} if the graph is bipartite or the full symmetric group S_{k-1} if it is not. There is only one interesting exception, a 7-vertex graph which generates a group of just 120 permutations. Wilson presents this group as $\mathrm{PGL}(2, \mathbb{F}_5)$, the group of Möbius transformations over the field of five elements, but it is isomorphic to the symmetric group S_5. Fink and Guy [2009] give a thorough discussion of this interesting, exceptional case, which they refer to as the "tricky six" puzzle.

John Conway [1997; 2006] uses a tile-sliding game on a 13-point projective plane to generate the Mathieu group M_{12}. This game is dubbed M_{13}. In M_{13}, moving a tile to the open point also requires swapping two other tiles at the same time. However, the rules of the game are specific to the projective plane on which it is played, and it does not generalize in any obvious way to a family of games on larger projective planes or other combinatorial structures.

MSC2010: 20B15, 91A43.
Keywords: simple group, permutation group, primitive group, graph theory.

Here, we consider a permutation game which can be played on any 3-regular graph, using a "slide-and-swap" rule inspired by Conway's M_{13}. We classify the resulting permutation groups and obtain a general result similar to Wilson's theorem, with just one interesting exceptional case. The exception gives a novel and elementary presentation for the simple group of 168 elements.

Like Rubik's cube and the other permutation games we have mentioned, ours can be treated purely as a puzzle, where you scramble pieces by moving them about and then try to return them to their initial configuration — or, perhaps, try to achieve some other "goal" configuration that has been posed as a problem. We have made a playable version of the game [Siehler 2011] as an aid to understanding the rules. In this article, we do not give any algorithms for unscrambling the pieces on a given graph, so solving the puzzle in that sense should remain an enjoyable challenge if you are so inclined.

Rules of the slide-and-swap game

Let Γ be a 3-regular graph (which we may as well assume to be connected) with labeled vertices. The game on Γ begins with one vertex uncovered and each of the remaining vertices covered by a tile with the same label. At any stage in the game, you make a move as follows: Choose a vertex v adjacent to the current uncovered vertex and slide the tile on v into the uncovered position (uncovering v in the process); and, at the same time, swap the tiles on the other two neighbors of v. See the first two diagrams in Figure 1, representing a single move where a tile slides from vertex b into the uncovered vertex and the tiles on the other two neighbors of vertex b are swapped.

We denote the move of sliding a tile from vertex b to vertex a by $[a, b]$. If this seems counterintuitive, think of the "hole" itself as a special blank tile which moves along the vertices in the order that they are listed when the move is played. Longer move sequences are expressed similarly: $[a, b, c, \ldots]$ is the move sequence which starts with the hole on vertex a, moves it to vertex b, from there to vertex c, and so on. Figure 1 shows a sequence of four moves played on the 8-vertex cube graph. The resulting permutation of tiles can be expressed in cycle form as $(b\ g\ c\ h\ d\ f\ e)$ — when this sequence is played, the tile which begins on vertex b moves to vertex g; the tile on vertex g moves to vertex c; and so on.

If $P = a, b, c, \ldots$ is a path in Γ, then "playing P" means playing the move sequence $[a, b, c, \ldots]$. In terms of tile movements, that means first moving the tile on vertex b to the vacant vertex a, then moving the tile on vertex c to the newly vacated vertex b, and so on, with the accompanying swaps at each step. The hole itself proceeds from a to b to c and so on, in the order they are listed.

Unlike the fifteen puzzle or M_{13}, the scrambling that happens as a result of

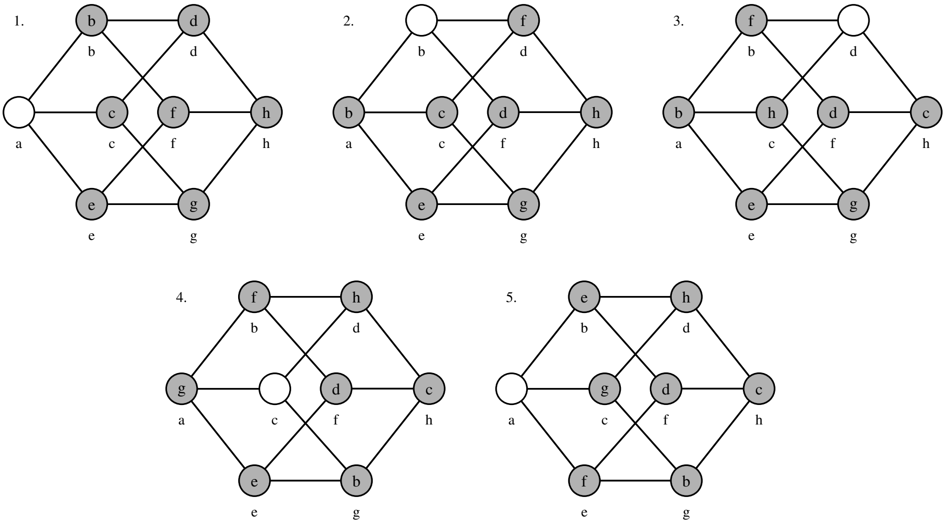


Figure 1. $[a, b, d, c, a] = (b\ g\ c\ h\ d\ f\ e)$ on the 8-vertex cube graph.

playing a path P in this game cannot be undone simply by playing its reverse (meaning the same path as P, just traversed in the opposite direction). However, we do have a basic result about invertibility:

Proposition 1. *Any legal move sequence in the slide-and-swap game can be undone* (*returning the tiles to their position before the sequence was played*) *by another legal move sequence.*

Proof. Suppose vertex x_0 is initially empty and we permute the tiles by playing $P = [x_0, x_1, \ldots, x_n]$. Let $R = [x_n, x_{n-1}, \ldots, x_0]$. The result of P followed by R is a permutation (of finite order, since there are only finitely many tiles!) which returns the hole to x_0. From that point we can play "P followed by R" repeatedly until all the tiles (and the hole) have returned to their initial position.

You may also note that a single slide $[x_0, x_1]$ can be undone by $[x_1, x_0, x_1, x_0]$. Consequently, a longer move sequence $[x_0, x_1, \ldots, x_n]$ can be undone one step at a time, as follows: First undo the final slide by playing $[x_n, x_{n-1}, x_n, x_{n-1}]$, then undo the one before that by $[x_{n-1}, x_{n-2}, x_{n-1}, x_{n-2}]$, and so on. □

Slide-and-swap loop groups

Now, suppose we begin a game on a connected, 3-regular graph Γ with k vertices, vertex a initially uncovered. The permutations which return the hole to its initial vertex (as in Figure 1) form a group under composition. We call this the *loop group* for Γ based at a and denote it $\mathcal{G}_a$. Since the basic moves in the game are all double transpositions, $\mathcal{G}_a$ is always a subgroup of the alternating group A_{k-1}. The notation suggests that the loop group depends on the choice of the initial uncovered vertex, but up to isomorphism the choice does not matter.

Proposition 2. *For any vertices a and b in Γ, the groups $\mathcal{G}_a$ and $\mathcal{G}_b$ are isomorphic.*

Proof. Since Γ is connected, choose a path P from a to b and let π be the permutation induced by playing P. The mapping $\alpha \mapsto \pi\alpha\pi^{-1}$ defines a homomorphism from $\mathcal{G}_a$ to $\mathcal{G}_b$. This homomorphism has $\beta \mapsto \pi^{-1}\beta\pi$ as its inverse, so the two groups are isomorphic (indeed, they are conjugate inside the symmetric group S_k). □

For this reason, we will henceforth omit any reference to the uncovered vertex and refer to *the* loop group associated to a graph.

Proposition 3. *The loop group of the tetrahedron (that is, the complete graph on 4 vertices) is trivial.*

The proof of this is left as an exercise. Larger graphs generate nontrivial groups, however, and a natural algebraic problem is to determine, up to isomorphism, which permutation groups can be realized as slide-and-swap loop groups. That problem is entirely resolved by the following theorems, which we will prove in the subsequent sections.

Notation. From now on, Γ will always denote a connected, 3-regular graph on k vertices, and $\mathcal{G}$ will denote its loop group (with the understanding that the choice of empty vertex doesn't matter).

Theorem 1. *If Γ is not the cube or tetrahedron, then $\mathcal{G}$ is isomorphic to the alternating group A_{k-1}.*

Theorem 2. *The loop group of the cube is isomorphic to $\mathrm{GL}(3, \mathbb{F}_2)$, the simple group of 168 elements.*

The resemblance to Wilson's results for "ordinary" tile-sliding games on graphs seems uncanny.

Fundamental terms and propositions

Dixon's problem book [1973] is a handy reference for the elementary theory of permutation groups, and the material is developed in depth in [Dixon and Mortimer 1996]. Here, we need only a few basic definitions and properties.

Let G be a group of permutations on a set X. The *orbit* of an element $x \in X$ is $\{\sigma(x) \mid \sigma \in G\}$. These orbits form a partition of X. If there is only one orbit (which contains all the elements of X), then G is said to be *transitive*.

A nonempty set $B \subseteq X$ is called a *block* for G if for every $\sigma \in G$, either $\sigma(B) = B$ or $\sigma(B) \cap B = \varnothing$. The set X itself is evidently a block, as are all singleton subsets of X; these are called *trivial* blocks. G is said to be *primitive* if all blocks for G are trivial; otherwise, if nontrivial blocks exist, G is *imprimitive*.

If G is transitive and B is any block, then the sets $\sigma(B)$, where $\sigma \in G$, partition X into disjoint, nonempty sets, each of which is a block. Such a partition of X is called a *system of imprimitivity* for G.

We consider $\mathcal{G}$ to be a group of permutations on the nonempty vertices of Γ. The following proposition is less obvious for slide-and-swap games than it is for ordinary tile-sliding games. Some time spent with a playable version of the game [Siehler 2011], trying to move a given tile to a given vertex, may be helpful in understanding the difficulties.

Proposition 4. *If* Γ *is not the tetrahedron*, *then* $\mathcal{G}$ *is transitive.*

Like Proposition 2, this is a useful fact to realize from the outset. It follows from the proof of Proposition 9, however, so we do not include a separate proof at this point. Our proof of Theorem 1 depends on the following general result:

Proposition 5 [Wilson 1974]. *Let G be a transitive permutation group on a set X and suppose that G contains a* 3*-cycle. If G is primitive*, *then G contains the alternating group on X.*

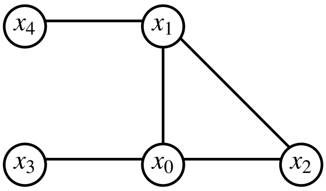
Generating the alternating group

The next few results show that, in general, our loop groups satisfy the hypotheses of Proposition 5. First, we establish the presence of 3-cycles.

Proposition 6. *If* Γ *is not the tetrahedron or cube*, *then* $\mathcal{G}$ *contains* 3*-cycles.*

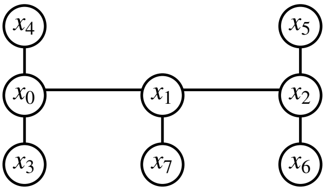
Proof. First, note that the isomorphism in Proposition 2 is realized by conjugation within the symmetric group, which preserves cycle types. For this reason, if the group based at any vertex contains a 3-cycle, then this is true at every vertex, and so we can choose the empty vertex at our convenience. In the following, the labels (x_1, x_2, and so on) name the vertices of the graph. It doesn't matter which tiles are on which vertices, except that the vertex labeled x_0 is the initially open vertex. We consider three cases.

Case 1. If Γ contains a triangle $\{x_0, x_1, x_2\}$, then (since Γ is not a tetrahedron) two vertices in the triangle must have distinct neighbors. Suppose x_0 and x_1 have neighbors x_3 and x_4, like this:



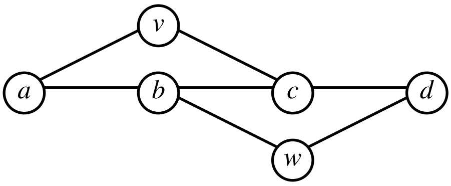
In this case, $[x_0, x_1, x_0] = (x_2\ x_4\ x_3)$.

Case 2. Suppose Γ does not contain a triangle, but contains a path x_0, x_1, x_2 where x_0 and x_2 have no neighbors in common except x_1:

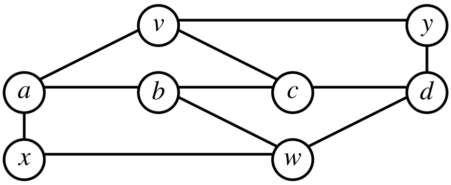


Then $[x_0, x_1, x_2, x_1, x_0]^2 = (x_1\ x_2\ x_7)$. Similarly, if Γ has no triangle but has a path x_0, x_1, x_2 where x_0 and x_2 have all three neighbors in common (say, $x_4 = x_5$ and x_3=x_6 in the figure), then $[x_0, x_1, x_2, x_1, x_0] = (x_1\ x_7\ x_2)$.

Case 3. The only remaining case to consider is a graph Γ with no triangles, in which the endpoints of every path of length two have exactly two neighbors in common. In *An atlas of graphs*, Read and Wilson [1998] show that there are only six 3-regular graphs of diameter less than three, including the tetrahedron, and none of them satisfy these hypotheses. These are all small graphs and the claim is easy to verify by inspection. Therefore we may assume that the diameter of Γ is at least 3. We claim that in this case Γ can only be the cube. Begin with a path a, b, c, d, where the distance from a to d is 3 (so there is no shorter path from a to d). Let v be the other common neighbor of a and c, and w the other common neighbor of b and d:



Now the path a, b, w implies that a and w have another common neighbor x. Similarly the path d, c, v implies that d and v must have another common neighbor y. This brings us to the following situation:



Since the graph is 3-regular, x needs another edge. If x were connected to some other vertex z not already shown, then the path z, x, a would imply that there is another vertex adjacent to both a and z. This would imply that either a, or one of the vertices adjacent to a in the preceding figure, has an additional edge, which is impossible; each of those vertices already has three edges. It follows that x and y are adjacent. At this point all the vertices have all three edges accounted for, and the resulting graph is the cube, as claimed. □

Proposition 7. *Suppose that for every pair of adjacent vertices x and y in Γ there is a p-cycle $\sigma \in \mathscr{G}$ where p is prime and $\sigma(x) = y$. Then $\mathscr{G}$ is primitive.*

Proof. Suppose the nonempty vertices are partitioned into blocks $B_1, B_2, \ldots, B_n$ forming a system of imprimitivity for $\mathscr{G}$. Let B_i and B_j be two blocks which contain adjacent vertices w_i and w_j, respectively. By hypothesis, we may choose a p-cycle $\sigma \in \mathscr{G}$ with $\sigma(w_i) = w_j$. Considered as a permutation of blocks, σ acts nontrivially (sending B_i to B_j), and since p is prime, this implies that σ acts with order p. However, σ only moves p vertices, so there can only be p blocks involved and only one vertex in each of those blocks, including B_i.

This argument can applied with any block playing the role of B_i and any adjacent block playing the role of B_j. Thus, since Γ is connected, either all blocks are singletons or there is only a single block. Since no nontrivial blocks are possible, $\mathscr{G}$ is primitive. □

Proposition 8. *Given any two adjacent, nonempty vertices x and y in Γ, there is a move sequence which leaves the tiles on x and y fixed while positioning the hole adjacent to one of those two vertices.*

Proof. If the empty vertex is already adjacent to x or y, no moves are needed. Otherwise, let Q be the shortest possible path beginning at the empty vertex, with the property that the ending vertex of Q has a distance of two from x or a distance of two from y. No point on Q can be adjacent to either x or y, since the previous point would have a distance of two, and a shorter path could be constructed. It follows that the permutation induced by playing Q leaves the tiles on x and y fixed. So, assume that we have played Q, ending with the hole on a vertex x_0 which is distance two from one of the given vertices — without loss of generality, say it's distance two from y. The goal now is to move the hole onto a vertex x_1 which is adjacent to y, still without disturbing the tiles on x and y. There are three local configurations to consider, and the proper moves to proceed in each case are shown in Figures 2–4.

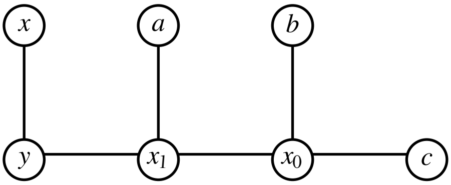


Figure 2. $[x_0, x_1, x_0, x_1] = (x_0\ x_1)(b\ c)$.

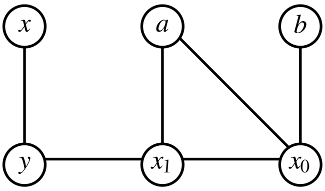


Figure 3. $[x_0, x_1, x_0, x_1, x_0, x_1] = (x_0\ x_1)(a\ b)$.

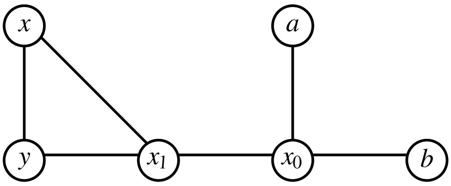
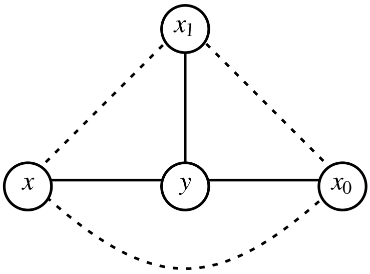


Figure 4. $[x_0, x_1, x_0, x_1] = (x_0\ x_1)(a\ b)$.

In each case, we get the desired result. Vertex x_1 (adjacent to y) is vacated, while the tiles on x and y remain fixed. We justify the claim that these three are the only cases as follows: Since x_0 is not adjacent to either x or y it must have two neighbors other than x, y or x_1. And x_1 must have one additional neighbor other than y or x_0. This additional neighbor may be x, or one of the neighbors of x_0, or a point distinct from both x and the neighbors of x_0, precisely the three cases we have considered. □

Proposition 9. *If* Γ *is not the tetrahedron, then the hypotheses of Proposition 7 are satisfied, and* $\mathscr{G}$ *is primitive.*

Proof. Once again, let x and y be any two adjacent, nonempty vertices in Γ. Assume the empty vertex is adjacent to y. Now, the possible configurations of the graph near x and y are summarized in the following figure, where x_0 is the empty vertex and x_1 is a vertex adjacent to y other than x and x_0.



Each dotted edge may or may not exist. If all edges exist simultaneously then Γ is the tetrahedron. That leaves seven nontrivial cases to consider.

There are six cases in which at least one dotted edge exists; they are shown in

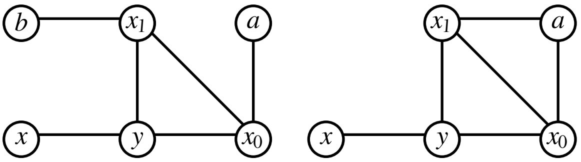


Figure 5. An x_0-x_1 edge. $[x_0, x_1, y, x_0]$ induces either $(x\ y\ b\ a\ x_1)$ or $(x\ y\ x_1)$.

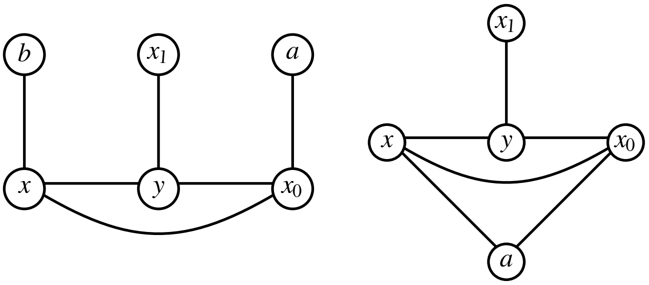


Figure 6. An x-x_0 edge. $[x_0, x, y, x_0]$ induces either $(x\ x_1\ y\ b\ a)$ or $(x\ x_1\ y)$.

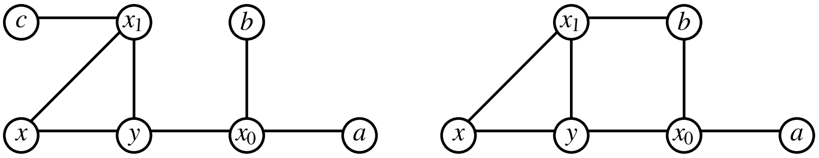


Figure 7. An x-x_1 edge. $[x_0, y, x_0, y, x_1, y, x_0]$ induces either $(x\ c\ y)$ or $(x\ a\ y)$.

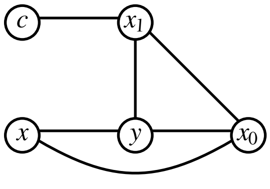


Figure 8. x-x_1 and x_0-x_1 edges. $[x_0, x_1, x_0] = (x\ y\ c)$.

Figures 5–10. In each case we exhibit a path which can be played to generate a cycle of prime length (either a 3- or a 5-cycle) sending x to y. Note that if a dotted edge is omitted, the vertices that edge connects must each have an edge to some other vertex not appearing in the figure at the bottom of previous page. These "extra" vertices are not necessarily distinct, so there are some subcases to be considered. In our figures, vertices not on the path or adjacent to a point on the path are not

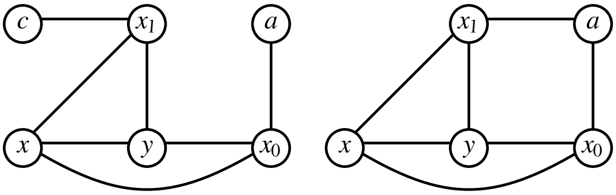


Figure 9. x-x_1 and x-x_0 edges. In the first case, $[x_0, y, x_1, y, x_0] = (x\ x_1\ c\ y\ a)$. In the second case, $[x_0, y, x_0, y, x_0, y, x_1, y, x_0] = (x\ y\ a)$.

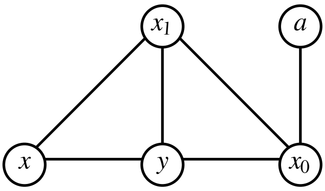


Figure 10. x-x_1 and x_0-x_1 edges. $[x_0, x_1, x_0] = (x\ a\ y)$.

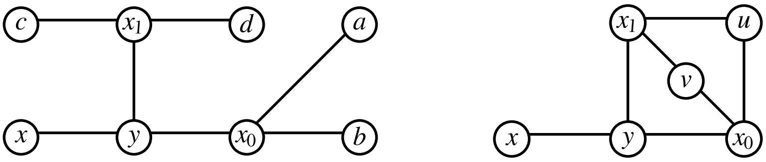


Figure 11. No edges among $\{x, x_0, x_1\}$. $[x_0, y, x_1, y, x_0]$ induces either $(a\ b)(c\ d)(x\ x_1\ y)$ (which can be squared to get the desired 3-cycle) or $(x\ x_1\ y)$.

drawn because they have no effect on the resulting permutation.

Figure 11 shows the case where none of the dotted edges are present, and x_1 and x_0 share 0 or 2 neighbors other than y.

That leaves the case where none of the dotted edges are present, and x_1 and x_0 have exactly one common neighbor u. The third neighbor of u is either one of the points on the graph other than y, or a point off the graph. So there are a total of four subcases to deal with in this case and they are shown in Figures 12–15.

In every case, we produce a cycle α of prime length which sends x to y. But we began with the provisional assumption that the initially empty vertex is adjacent to y. In general, however, Proposition 8 can be applied to produce a permutation σ which moves the empty vertex adjacent to y while leaving x and y fixed. Conjugating α by σ gives the desired cycle in $\mathscr{G}$. □

Remark. Since any vertex may be moved to any adjacent vertex by means of these cycles, $\mathscr{G}$ is transitive, as we asserted in Proposition 4.

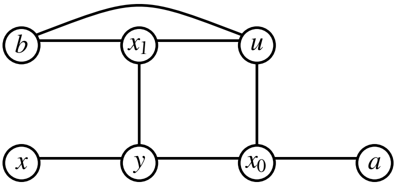


Figure 12. $[x_0, u, x_1, y, x_0] = (a\ u\ x\ y\ b)$.

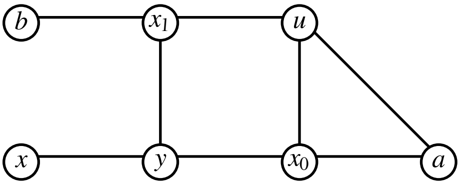


Figure 13. $[x_0, u, x_1, y, x_0] = (x\ y\ b\ x_1\ u)$.

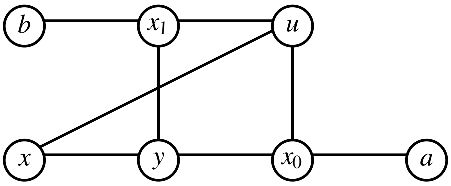


Figure 14. $[x_0, y, x_0, u, x_1, u, x_0] = (x\ u\ y\ b\ a)$.

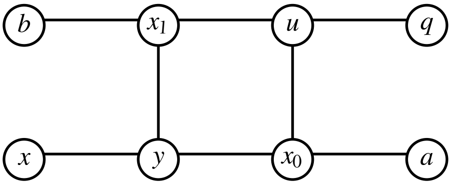


Figure 15. $[x_0, u, x_1, y, x_0] = (a\ u\ x\ y\ b\ x_1\ q)$.

Proof of Theorem 1. Our main result now follows quickly. If Γ is not the cube or tetrahedron, then Proposition 6 shows that $\mathscr{G}$ contains 3-cycles. Proposition 9 shows that $\mathscr{G}$ is transitive and primitive. Thus Proposition 5 applies to $\mathscr{G}$ and we conclude that $\mathscr{G}$ contains all even permutations of the nonempty vertices. □

The exceptional cube

Now, we analyze the loop group of the cube. Initially, a computer calculation revealed that this group has only 168 elements (instead of the expected $7!/2 = 2520$ in A_7). The number 168 is familiar to algebraists as the order of $\mathrm{GL}(3, \mathbb{F}_2)$, the group

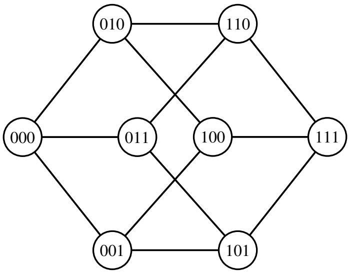


Figure 16. Vertices of the cube labeled with vectors of $\mathbb{F}_2^3$.

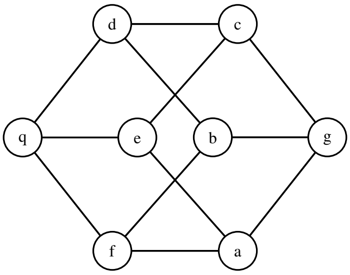


Figure 17. Vertices of the cube labeled with unknown vectors.

of invertible 3 × 3 matrices over the field of two elements, and the second-smallest nonabelian simple group.

To establish a connection between this group and the cube, we label the vertices with 3-dimensional vectors over $\mathbb{F}_2$, as in Figure 16. For brevity, we write a vector $\langle b_1, b_2, b_3 \rangle$ as a 3-bit binary string $b_1b_2b_3$. With such a labeling, moves in the game can be interpreted as permutations of $\mathbb{F}_2^3$. The particular labeling in Figure 16 has the property that the sum (mod 2, of course) of any tile together with its three adjacent tiles is zero, and we will call any arrangement of vector tiles with this property a *locally zero* arrangement.

Proposition 10. *The permutation of $\mathbb{F}_2^3$ induced by a single move on a cube with a locally zero arrangement of vectors is an affine transformation.*

Proof. Let $a, \ldots, g$ and q be the eight vectors of $\mathbb{F}_2^3$, labeling the vertices of the cube in a locally zero arrangement as in Figure 17. If we define

$$\alpha = a + q, \quad \beta = b + q, \quad \gamma = c + q,$$

the following additional relations follow quickly from the locally zero condition:

$$\beta + \gamma = d + q, \quad \alpha + \gamma = e + q, \quad \alpha + \beta = f + q, \quad \alpha + \beta + \gamma = g + q.$$

Note that all the linear combinations of α, β, and γ are distinct, so $\{\alpha, \beta, \gamma\}$ is a linearly independent set (and, in fact, a basis for $\mathbb{F}_2^3$).

Now, consider a slide-and-swap move. By symmetry we can assume that q is empty and we slide a tile into the hole from d. This induces a permutation φ with $\varphi(q)=d,\ \varphi(d)=q,\ \varphi(b)=c,\ \varphi(c)=b,$ and the other points remaining fixed. Define $\hat{\varphi}$ by $\hat{\varphi}(x)=\varphi(x+q)+d$. Applying $\hat{\varphi}$ to our basis elements gives

$$\hat{\varphi}(\alpha)=\alpha+\beta+\gamma, \quad \hat{\varphi}(\beta)=\beta, \quad \hat{\varphi}(\gamma)=\gamma,$$

from which we can verify the hard way (that is, by checking every linear combination of basis elements) that $\hat{\varphi}$ is linear:

$$\begin{aligned}
\hat{\varphi}(\alpha+\beta) &= \hat{\varphi}(f+q) = f+d = \alpha+\gamma = \hat{\varphi}(\alpha)+\hat{\varphi}(\beta),\\
\hat{\varphi}(\alpha+\gamma) &= \hat{\varphi}(e+q) = e+d = \alpha+\beta = \hat{\varphi}(\alpha)+\hat{\varphi}(\gamma),\\
\hat{\varphi}(\beta+\gamma) &= \hat{\varphi}(d+q) = q+d = \beta+\gamma = \hat{\varphi}(\beta)+\hat{\varphi}(\gamma),\\
\hat{\varphi}(\alpha+\beta+\gamma) &= \hat{\varphi}(g+q) = g+d = \qquad \alpha = \hat{\varphi}(\alpha)+\hat{\varphi}(\beta)+\hat{\varphi}(\gamma),
\end{aligned}$$

and of course $\hat{\varphi}(0)=\varphi(q)+d=d+d=0$. Thus, with respect to the basis $\{\alpha,\ \beta,\ \gamma\}$, $\hat{\varphi}$ is represented by the matrix

$$M=\begin{pmatrix}1&0&0\\1&1&0\\1&0&1\end{pmatrix},$$

and φ is described by the formula $\varphi(x)=M(x+q)+d$, or $\varphi(x)=Mx+(Mq+d)$, an affine transformation as claimed. □

Proof of Theorem 2. Begin the game with vertices labeled by vectors in a locally zero arrangement and the 000 vertex open. By the preceding proposition, any sequence of slides is a composition of affine transformations (which is again an affine transformation). The location of the hole always reveals the translation part of the transformation, so the loop group (corresponding to slides where the hole returns to 000) consists of linear transformations and is contained in $\mathrm{GL}(3, \mathbb{F}_2)$.

To complete the proof, we simply exhibit a few elements of the group. Returning to the cube in Figure 17 and supposing q is initially open, consider the elements

$$\begin{aligned}
[q,e,q,d,b,f,b,f,q] &= (d\ e)(a\ g\ b\ c),\\
[q,f,q] &= (d\ e)(a\ b),
\end{aligned}$$

which generate a dihedral group of eight elements. Also,

$$\begin{aligned}
[q,e,c,e,q] &= (f\ g\ e)(d\ c\ b),\\
[q,d,b,f,q] &= (f\ g\ b\ c\ e\ d\ a).
\end{aligned}$$

Subgroups of order 8, 3, and 7 imply a group of order at least 168, and so the loop group of the cube is not just contained in, but equal to $\mathrm{GL}(3, \mathbb{F}_2)$. □

Efficient solutions and other questions

In principle, the method of proof that we used to classify the loop groups could likely be turned into an algorithm for solving a scrambled puzzle on any given graph, since we show how to produce small, localized cycles at any point on the graph, which could be used to migrate pieces to their appropriate locations. In practice, this sort of solution takes many more moves than the optimal solution. In ordinary tile-sliding games, the problem of finding an optimal solution for a scrambled state is NP-complete [Goldreich 1984; Ratner and Warmuth 1990]. We do not know if the same is true for slide-and-swap games; this is a question for a future paper. The slide-and-swap variant is also unusual in that the number of moves required to achieve a position from start may be different from the number of moves required to return it to start. This aspect of the puzzle has no counterpart in other sliding or twisting permutation puzzles that we are familiar with, and the relationship between distance to and distance from start is worthy of some further analysis.

Conway, Elkies and Martin [Conway et al. 2006] have shown how to use duality of the projective plane to produce an outer automorphism of the Mathieu group M_{12}. It would be interesting if the symmetries of the cube and the slide-and-swap game rules allowed a similar construction of outer automorphisms for $\mathrm{GL}(3, \mathbb{F}_2)$, but we have not yet discovered how to do this, and it remains a subject for further investigation.

References

[Archer 1999] A. F. Archer, “A modern treatment of the 15 puzzle”, *Amer. Math. Monthly* **106**:9 (1999), 793–799. MR 2001a:05004 Zbl 1007.00006

[Conway 1997] J. H. Conway, “M_{13}”, pp. 1–11 in *Surveys in combinatorics, 1997*, edited by R. A. Bailey, London Math. Soc. Lecture Note Ser. **241**, Cambridge Univ. Press, 1997. MR 98i:20003 Zbl 0887.05002

[Conway et al. 2006] J. H. Conway, N. D. Elkies, and J. L. Martin, “The Mathieu group M_{12} and its pseudogroup extension M_{13}”, *Experiment. Math.* **15**:2 (2006), 223–236. MR 2007k:20005 Zbl 1112.20003

[Dixon 1973] J. D. Dixon, *Problems in group theory*, Dover, 1973.

[Dixon and Mortimer 1996] J. D. Dixon and B. Mortimer, *Permutation groups*, Graduate Texts in Mathematics **163**, Springer, New York, 1996. MR 98m:20003 Zbl 0951.20001

[Fink and Guy 2009] A. Fink and R. Guy, “Rick’s tricky six puzzle: S_5 sits specially in S_6”, *Math. Mag.* **82**:2 (2009), 83–102. MR 2512593 Zbl 1223.05113

[Goldreich 1984] O. Goldreich, “Finding the shortest move-sequence in the graph-generalized 15-puzzle is NP-hard”, 1984, Available at http://tinyurl/15puzz-pdf. preprint.

[Ratner and Warmuth 1990] D. Ratner and M. Warmuth, “The $(n^2 - 1)$-puzzle and related relocation problems”, *J. Symbolic Comput.* **10**:2 (1990), 111–137. MR 91i:68138 Zbl 0704.68057

[Read and Wilson 1998] R. C. Read and R. J. Wilson, *An atlas of graphs*, The Clarendon Press Oxford University Press, New York, 1998. MR 2000a:05001 Zbl 0908.05001

[Siehler 2011] J. A. Siehler, "Slide and swap on cubic graphs", website, 2011, Available at http://tinyurl.com/sandscubic.

[Wilson 1974] R. M. Wilson, "Graph puzzles, homotopy, and the alternating group", *J. Combinatorial Theory Ser. B* **16** (1974), 86–96. MR 48 #10882 Zbl 0285.05110

Received: 2012-07-16 Revised: 2013-05-10 Accepted: 2013-05-25

ekentao15@mail.wlu.edu *Department of Mathematics, Washington and Lee University, Lexington, VA 24450, United States*

jangha15@mail.wlu.edu *Department of Mathematics, Washington and Lee University, Lexington, VA 24450, United States*

jsiehler@gmail.com

INVOLVE 7:1 (2014)
dx.doi.org/10.2140/involve.2014.7.57

Comparing a series to an integral

Leon Siegel

(Communicated by Andrew Granville)

We consider the difference between the definite integral $\int_0^\infty u^x e^{-u}\,\mathrm{d}u$, where x is a real parameter, and the approximating sum $\sum_{k=1}^\infty k^x e^{-k}$. We use properties of Bernoulli numbers to show that this difference is unbounded and has infinitely many zeros. We also conjecture that the sign of the difference at any positive integer n is determined by the sign of $\cos\bigl((n+1)\arctan(2\pi)\bigr)$.

1. Introduction

There are a variety of situations where it is necessary to examine differences of sums and integrals. The Euler–Maclaurin summation formula is the usual tool for estimating $\int_{u\le Y} g(u)\,\mathrm{d}u - \sum_{n\le Y} g(n)$ [Abramowitz and Stegun 1964, p. 806], but it can also be interesting to develop exact formulas for particular choices of $g(u)$. For instance, the Euler–Mascheroni constant arises if we set $g(u) = 1/u$ and consider the limit as $Y \to \infty$ [Wells 1986, p. 12]. The purpose of this paper is to examine the function

$$f(x) := \sum_{k=1}^{\infty} k^x e^{-k} - \int_0^\infty t^x e^{-t}\,\mathrm{d}t.$$

The integral on the right equals $\Gamma(x+1)$, where $\Gamma(x)$ is the gamma function, and the infinite series converges absolutely for all values of x. We can obtain an exact expression for $f(n)$ when $n \ge 1$ by using classical formulas for polylogarithms of negative order [Weisstein 2013]:

$$f(n) = -n! + \sum_{k=0}^{n} \frac{1}{(e-1)^{k+1}} \sum_{j=0}^{k} (-1)^j \binom{k}{j} (k+1-j)^n. \tag{1}$$

The main goal of this paper is to prove that $f(x)$ has infinitely many positive real zeros, and that the function becomes unbounded as $x \to \infty$. Further, in Conjecture 1 we hypothesize that $f(n)$ has the same sign as $\cos\bigl((n+1)\arctan(2\pi)\bigr)$ whenever

MSC2010: 33B15.
Keywords: polylogarithms, gamma function, Bernoulli numbers.

n is a positive integer. We prove that the conjecture is true with finitely many exceptions, provided that $\arctan(2\pi)/\pi$ has finite irrationality measure. If we expand $\cos\bigl((n+1)\arctan(2\pi)\bigr)$ using trigonometric identities, then we obtain the equivalent conjecture that the following identity holds for all positive integers n:

$$\begin{aligned}\operatorname{sign}&\Biggl[\sum_{j=0}^{n+1}(-1)^j\binom{n+1}{2j}(2\pi)^{2j}\Biggr]\\ &=\operatorname{sign}\Biggl[-(e-1)^{n+1}n!+\sum_{k=0}^{n}(e-1)^{n-k}\sum_{j=0}^{k}(-1)^j\binom{k}{j}(k+1-j)^n\Biggr].\end{aligned}\tag{2}$$

The left-hand side of (2) is a polynomial in π, while the right-hand side is a polynomial in e. Based on numerical experiments, we conjecture that (π, e) is the unique, nontrivial (i.e., $\neq (0, 1)$) tuple of real numbers which makes (2) valid for all positive integers n. When we choose values close to π and e respectively, we notice that (2) is false for some n in all considered cases. Surprisingly, (2) is valid for $n \leq 128$ if you insert $(\pi + 0.015, e)$, but only for $n \leq 2$ in the case of $(\pi, e + 0.015)$. So the equation seems to be a lot more sensitive to small modifications in the argument on the right-hand side. Also, choosing various random tuples (x, y) further away from (π, e), we always found an n such that (2) was wrong.

2. Elementary properties of $f(x)$

In this section we prove that $f(x)$ is an unbounded function by showing that the sequence $\{f(n)\}_{n=1}^{\infty}$ is unbounded as $n \to \infty$. Our proof uses properties of Bernoulli numbers. The n-th Bernoulli number is defined by

$$\frac{x}{e^x-1}=\sum_{n=0}^{\infty}B_n\frac{x^n}{n!},\tag{3}$$

and the generating series converges for $|x| < 2\pi$. It is known that Bernoulli numbers are always rational, and that $B_n = 0$ if $n > 1$ is odd. Bernoulli numbers have many interesting combinatorial properties [Abramowitz and Stegun 1964], and the following asymptotic holds for large values of n:

$$|B_{2n}| \sim \frac{n^{2n}}{(\pi e)^{2n}}.\tag{4}$$

This property will be used later. We begin by deriving a new formula for B_n. Then in Theorem 1, we use our formula to prove that $f(x)$ is unbounded.

Lemma 1. $$B_n=\sum_{k=n}^{\infty}\frac{f(k)-kf(k-1)}{(k-n)!}\quad \textit{for } n \geq 2.$$

Proof. Consider the generating function of the Bernoulli numbers,

$$g(x) := \frac{x}{e^x - 1},$$

whose Taylor series at $x = -1$ is

$$g(x) = \frac{e}{e-1} - \frac{e(e-2)}{(e-1)^2}(x+1) + \sum_{n=2}^{\infty} \frac{f(n) - nf(n-1)}{n!}(x+1)^n. \tag{5}$$

The Taylor coefficients at $n = 0$ and $n = 1$ are calculated directly. To obtain the coefficients when $n \geq 2$, we use

$$\begin{aligned} g^{(n)}(-1) =& \frac{\mathrm{d}^n}{\mathrm{d}x^n}\left[\frac{-x}{1-e^x}\right]_{x=-1} = \frac{\mathrm{d}^n}{\mathrm{d}x^n}\left[-x\sum_{m=0}^{\infty} e^{mx}\right]_{x=-1} \\ =& \sum_{m=1}^{\infty} m^n e^{-m} - n\sum_{m=1}^{\infty} m^{n-1}e^{-m} \\ =& \left(\sum_{m=1}^{\infty} m^n e^{-m} - n!\right) - n\left(\sum_{m=1}^{\infty} m^{n-1}e^{-m} - (n-1)!\right) \\ =& f(n) - nf(n-1). \end{aligned} \tag{6}$$

Since formula (3) is also valid when x lies in a neighborhood of -1, we can equate the two results:

$$\begin{aligned} g(x) = \sum_{n=0}^{\infty} \frac{B_n}{n!}x^n &= \sum_{n=0}^{\infty} \frac{g^{(n)}(-1)}{n!}(x+1)^n \\ &= \sum_{n=0}^{\infty} \frac{g^{(n)}(-1)}{n!}\sum_{k=0}^{n}\binom{n}{k}x^k = \sum_{n=0}^{\infty}\left[\sum_{k=n}^{\infty}\frac{g^{(k)}(-1)}{(k-n)!}\right]\frac{x^n}{n!}. \end{aligned}$$

Comparing coefficients and then applying (6), we find that for $n \geq 2$,

$$B_n = \sum_{k=n}^{\infty} \frac{g^{(k)}(-1)}{(k-n)!} = \sum_{k=n}^{\infty} \frac{f(k) - kf(k-1)}{(k-n)!}. \qquad \square$$

Theorem 1. *The sequence* $\{f(n)\}_{n=1}^{\infty}$ *is unbounded.*

Proof. We construct a proof by contradiction. Assume that $|f(n)| < C$ for some $C > 0$ and every $n \in \mathbb{N}$. By Lemma 1 and the triangle inequality, we have

$$|B_n| \leq \sum_{k=n}^{\infty} \frac{|f(k) - kf(k-1)|}{(k-n)!} \leq \sum_{k=n}^{\infty} \frac{C(1+k)}{(k-n)!} \leq Ce(n+2).$$

This contradicts the asymptotic $|B_{2n}| \sim n^{2n}/(\pi e)^{2n}$, which holds for n sufficiently large. □

Remark. Despite the fact that $f(n)$ is unbounded as $n \to \infty$, the ratio $f(n)/n!$ converges to zero. To prove this, we can use residue calculus to show that

$$\frac{f(n)}{n!} = \frac{1}{2\pi i} \oint_{\gamma} \frac{z^{-n-1}}{1-e^{z-1}} \, \mathrm{d}z,$$

where $\gamma = \{z \in \mathbb{C} \colon |z| = 2\}$. We then employ the triangle inequality and numerical integration to obtain the crude upper bound $|f(n)|/n! \le 0.82 \times 2^{-n}$. In fact, it is possible to develop a much sharper upper bound using formula (12) below.

Theorem 2. *The function $f(x)$ has infinitely many zeros.*

Proof. First notice that $f(2) \approx -0.0077$ and $f(3) \approx 0.0065$, so by continuity $f(x)$ has at least one zero in the interval $(2, 3)$. To prove that the function has infinitely many zeros, we proceed by contradiction.

Assume that f has only finitely many zeros. Then for any sufficiently large integer m, the elements of the set $\{f(m), f(m+1), f(m+2), \dots\}$ all have the same sign. Now consider the function

$$h(x) := \frac{1}{x} - \frac{1}{e^x - 1},$$

which has the Taylor series

$$h(x) = \frac{1}{e-1} + \sum_{k=1}^{\infty} \frac{f(k)}{k!}(x+1)^k. \tag{7}$$

Differentiating m times gives

$$h^{(m)}(x) = \sum_{k=m}^{\infty} \frac{f(k)}{(k-m)!}(x+1)^{k-m}. \tag{8}$$

If the elements of the set $\{f(m), f(m+1), \dots\}$ are strictly positive, then (8) becomes a sum over positive numbers whenever $x \in (-1, 0)$, and it follows that $h^{(m)}(x)$ is strictly positive. If we notice that

$$h(x) = \frac{1}{x} - \frac{1}{x}\frac{x}{e^x - 1} = -\sum_{n=1}^{\infty} \frac{B_n}{n!} x^{n-1},$$

then we also have

$$h^{(m)}(x) = -\sum_{n=m+1}^{\infty} \frac{B_n}{n!} \frac{(n-1)!}{(n-m-1)!} x^{n-m-1}. \tag{9}$$

The key observation is that formulas (8) and (9) have overlapping domains of convergence on the negative real axis near the origin. If x is a sufficiently small negative real number, then (9) implies

$$h^{(m)}(x) \approx -\frac{B_{m+1}}{m+1},$$

but (8) guarantees

$$h^{(m)}(x) > 0.$$

This is a contradiction, because Bernoulli numbers assume both positive and negative values as m increases. We can deal with the case where $\{f(m), f(m+1), \dots\}$ are strictly negative in a similar manner. □

In Theorem 2 we proved that $f(x)$ has infinitely many real zeros. In fact, we can be much more precise about the locations of the zeros. If x_j denotes the j-th positive real zero of $f(x)$ such that $f(x_j) = 0$, then we expect that

$$x_j \approx -1 + \frac{\pi(2j+1)}{2\arctan(2\pi)}. \tag{10}$$

The first approximation gives $x_1 \approx 2.335\dots$, and this is reasonably close to the true value $x_1 = 2.306\dots$. We have observed numerically that the approximations become more accurate for large values of j. To derive (10), consider an identity which is valid for $\mathrm{Re}(x) > 0$ and $\mathrm{Re}(\mu) > 0$:

$$\frac{1}{\Gamma(x+1)} \sum_{k=1}^{\infty} k^x e^{-\mu k} = \sum_{k=-\infty}^{\infty} \frac{1}{(\mu + 2\pi i k)^{x+1}}. \tag{11}$$

Formula (11) is a special case of an identity due to Lipschitz [Rademacher 1973, p. 77], and follows from the Poisson summation formula. Set $\mu = 1$ and take the real part of both sides to obtain

$$\frac{f(x)}{\Gamma(x+1)} = 2\sum_{k=1}^{\infty} \frac{\cos\big((x+1)\arctan(2\pi k)\big)}{(1+4\pi^2k^2)^{(x+1)/2}}. \tag{12}$$

Equation (12) converges rapidly, and we can approximate $f(x)$ by truncating the series. The first term gives

$$\frac{f(x)}{\Gamma(x+1)} \approx 2\frac{\cos\big((x+1)\arctan(2\pi)\big)}{(1+4\pi^2)^{(x+1)/2}}, \tag{13}$$

and we immediately recover (10). It is somewhat subtle to determine how often (13) actually provides a good approximation of $f(x)$, and we touch on this point in the next section.

3. A conjecture on the sign of $f(n)$

A second observation from (13) is that the sign of $f(n)$ should always equal the sign of $\cos\bigl((n+1)\arctan(2\pi)\bigr)$. We have verified this numerically for $n \le 5000$ in Maple, and as a result we have the following conjecture:

Conjecture 1. *For all positive integers n,*

$$\operatorname{sign} f(n) = \operatorname{sign}\cos\bigl((n+1)\arctan(2\pi)\bigr). \tag{14}$$

Equivalently, for every positive integer n,

$$\begin{aligned}\operatorname{sign}&\left[\sum_{j=0}^{n+1}(-1)^j\binom{n+1}{2j}(2\pi)^{2j}\right]\\ &= \operatorname{sign}\left[-(e-1)^{n+1}n! + \sum_{k=0}^{n}(e-1)^{n-k}\sum_{j=0}^{k}(-1)^j\binom{k}{j}(k+1-j)^n\right].\end{aligned} \tag{15}$$

Conjecture 1 is easy to check numerically. The main difficulty in actually proving the conjecture is to determine how often (13) leads to a good approximation of $f(n)$. The reason that (14) might fail is because $(n+1)\arctan(2\pi)$ is unreasonably close to a half-integer multiple of π. This would cause the first term of the infinite series in (12) to nearly vanish, in which case higher-order terms would dominate and the estimate in (13) would fail. Thus we need to rule out the possibility that $(n+1)\arctan(2\pi)$ is unreasonably close to a half-integer multiple of π. This is equivalent to ruling out the possibility that $\arctan(2\pi)/\pi$ is unreasonably well approximated by rational numbers. Before proceeding, we note that $\arctan(2\pi)/\pi$ is trivially irrational, because otherwise we would have an identity of the form $2\pi = \tan(p\pi/q)$ for some $(p, q) \in \mathbb{Z}^2$, contradicting the transcendence of π.

Lemma 2. *Equation* (14) *is true for any positive integer n which satisfies*

$$\left|\cos\bigl((n+1)\arctan(2\pi)\bigr)\right| > \frac{2.6}{1.98^{n+1}}. \tag{16}$$

Proof. First, rewrite (12) as

$$\frac{f(n)}{n!} = 2\frac{\cos\bigl((n+1)\arctan(2\pi)\bigr)}{(1+4\pi^2)^{(n+1)/2}} + 2\sum_{k=2}^{\infty}\frac{\cos\bigl((n+1)\arctan(2\pi k)\bigr)}{(1+4\pi^2k^2)^{(n+1)/2}}.$$

If the first term on the right dominates, then it follows easily that

$$\operatorname{sign}\frac{f(n)}{n!} = \operatorname{sign}\frac{2\cos\bigl((n+1)\arctan(2\pi)\bigr)}{(1+4\pi^2)^{(n+1)/2}},$$

and this is equivalent to Conjecture 1. Thus we need to prove

$$\left|2\frac{\cos\bigl((n+1)\arctan(2\pi)\bigr)}{(1+4\pi^2)^{(n+1)/2}}\right| > \left|2\sum_{k=2}^{\infty}\frac{\cos\bigl((n+1)\arctan(2\pi k)\bigr)}{(1+4\pi^2k^2)^{(n+1)/2}}\right|. \tag{17}$$

Equation (16) easily implies that

$$\left|2\frac{\cos\bigl((n+1)\arctan(2\pi)\bigr)}{(1+4\pi^2)^{(n+1)/2}}\right| > \frac{5.2}{1.98^{n+1}(1+4\pi^2)^{(n+1)/2}} > \frac{5.2}{12.59^{n+1}}. \tag{18}$$

On the other hand, by the triangle inequality

$$\begin{aligned}\left|2\sum_{k=2}^{\infty}\frac{\cos\bigl((n+1)\arctan(2\pi k)\bigr)}{(1+4\pi^2k^2)^{(n+1)/2}}\right| &\le 2\sum_{k=2}^{\infty}\frac{1}{(1+4\pi^2k^2)^{(n+1)/2}}\\ &< \frac{2}{(1+16\pi^2)^{(n-1)/2}}\sum_{k=2}^{\infty}\frac{1}{1+4\pi^2k^2}\\ &< \frac{5.2}{(1+16\pi^2)^{(n+1)/2}} < \frac{5.2}{12.6^{n+1}}.\end{aligned} \tag{19}$$

Thus combining (19) and (18) shows that

$$\begin{aligned}\left|\frac{2\cos\bigl((n+1)\arctan(2\pi)\bigr)}{(1+4\pi^2)^{(n+1)/2}}\right| - \left|2\sum_{k=2}^{\infty}\frac{\cos\bigl((n+1)\arctan(2\pi k)\bigr)}{(1+4\pi^2k^2)^{(n+1)/2}}\right|&\\ > \frac{5.2}{12.59^{n+1}} - \frac{5.2}{12.6^{n+1}} > 0,&\end{aligned}$$

and (17) follows immediately. Therefore Conjecture 1 is true whenever n is a positive integer for which (16) holds. □

It is typically very tricky to determine how well a particular number θ can be approximated by rational numbers. We say that θ has irrationality measure μ if μ is the smallest real number such that

$$\left|\theta - \frac{p}{q}\right| > \frac{1}{q^{\mu}}$$

for all but finitely many pairs $(p, q) \in \mathbb{Z}^2$ with $q > 0$. The Thue–Roth–Siegel theorem guarantees that $\mu = 2$ whenever θ is algebraic and irrational [Roth 1955]. An easy consequence of this theorem is that θ can never be algebraic and have irrationality measure greater than 2. The typical method for proving that particular numbers are *transcendental* is to construct infinite sequences of rational numbers which approximate them too well. Liouville gave the first examples of transcendental numbers in 1851 [Niven 1956, p. 93]. He proved that numbers like

$$\theta_0 = \sum_{n=1}^{\infty} \frac{1}{10^{n!}}$$

are always transcendental. Notice that if we set $p_N = \sum_{n=1}^{N} 10^{N!-n!}$ and $q_N = 10^{N!}$, then it is easy to show that

$$\left|\theta_0 - \frac{p_N}{q_N}\right| \le \frac{2}{q_N^{N+1}}.$$

Given any $k > 0$, this allows us to construct infinite sequences of rational numbers so that $|\theta_0 - p/q| < 1/q^k$. Numbers with this property are called *Liouville numbers* and are said to have infinite irrationality measure. While a simple counting argument shows that almost all numbers are irrational, the set of Liouville numbers has measure zero inside the irrational numbers. Irrational numbers typically have finite irrationality measures; it is known that π has irrationality measure at most 7.6063 [Salikhov 2008], and $\log 2$ has irrationality measure at most 3.57455391 [Marcovecchio 2009].

Theorem 3. *Assume that* $\arctan(2\pi)/\pi$ *has finite irrationality measure. Then Conjecture 1 is true for n sufficiently large.*

Proof. Assume that (16) fails for some integer n. Then we have

$$\frac{2.6}{1.98^{n+1}} \ge \bigl|\cos\bigl((n+1)\arctan(2\pi)\bigr)\bigr| = \bigl|\sin\bigl((n+1)\arctan(2\pi) - \tfrac{\pi}{2} - \pi j\bigr)\bigr|$$

for any integer j. Select j so that $z \in [-\pi/2, \pi/2]$, where z is the argument of the sine function. Elementary estimates show that $|\sin z| \ge 2|z|/\pi$. Thus

$$\frac{2.6}{1.98^{n+1}} \ge \frac{2}{\pi}\left|(n+1)\arctan(2\pi) - \frac{\pi}{2} - \pi j\right|,$$

and rearranging gives

$$\frac{1.3}{(n+1)1.98^{n+1}} \ge \left|\frac{\arctan(2\pi)}{\pi} - \frac{2j+1}{2(n+1)}\right|. \tag{20}$$

If $\arctan(2\pi)/\pi$ has finite irrationality measure, then (20) can only hold for finitely many values of n. We conclude that (16) holds for n sufficiently large, which implies that Conjecture 1 is also true for n sufficiently large. □

Acknowledgements

The author thanks Andrew Granville and the referee for their help.

References

[Abramowitz and Stegun 1964] M. Abramowitz and I. A. Stegun (editors), *Handbook of mathematical functions with formulas, graphs, and mathematical tables*, National Bureau of Standards Applied Mathematics Series **55**, U.S. Government Printing Office, Washington, DC, 1964. Reprinted by Dover, New York, 1974. MR 29 #4914 Zbl 0171.38503

[Marcovecchio 2009] R. Marcovecchio, "The Rhin–Viola method for log 2", *Acta Arith.* **139**:2 (2009), 147–184. MR 2010j:11114 Zbl 1197.11083

[Niven 1956] I. Niven, *Irrational numbers*, The Carus Mathematical Monographs **11**, MAA/Wiley, New York, 1956. MR 18,195c Zbl 0070.27101

[Rademacher 1973] H. Rademacher, *Topics in analytic number theory*, edited by E. Grosswald et al., Grundlehren der mathematischen Wissenschaften **169**, Springer, New York, 1973. MR 51 #358 Zbl 0253.10002

[Roth 1955] K. F. Roth, "Rational approximations to algebraic numbers", *Mathematika* **2** (1955), 1–20; corrigendum, 168. MR 17,242d Zbl 0064.28501

[Salikhov 2008] V. K. Salikhov, "О мере иррациональности числа π", *Uspekhi Mat. Nauk* **63**:3 (2008), 163–164. Translated as "On the irrationality measure of π" in *Russ. Math. Surv.* **63**:3 (2008), 570–572. MR 2010b:11082 Zbl 1208.11086

[Weisstein 2013] E. W. Weisstein, "Polylogarithm", MathWorld: a Wolfram web resource, 2013, Available at http://mathworld.wolfram.com/Polylogarithm.html.

[Wells 1986] D. Wells, *The Penguin dictionary of curious and interesting numbers*, Penguin, New York, 1986.

Received: 2012-07-17 Revised: 2013-05-25 Accepted: 2013-05-25

alkadash_leon@web.de *Christian-Albrechts-Universität zu Kiel,*
Christian-Albrechts-Platz 4, 24118 Kiel, Germany

INVOLVE 7:1 (2014)
dx.doi.org/10.2140/involve.2014.7.67

Some investigations on a class of nonlinear integrodifferential equations on the half-line

Mariateresa Basile, Woula Themistoclakis and Antonia Vecchio

(Communicated by Kenneth S. Berenhaut)

We consider the second-order integrodifferential boundary value problem

$$\begin{cases} \nu(y)g(y) - \int_0^{+\infty} k(x)g(x)\,dx\,[D(y)g'(y)]' = p(y) & \text{for } y \geq 0, \\ g'(0)=0, \quad g(+\infty)=0, \end{cases}$$

arising from the kinetic theory of dusty plasmas, and we provide information on the existence and other qualitative properties of the solution that have been essential in the numerical investigation.

1. Introduction

In this paper we present an analytical study of a particular class of nonlinear integrodifferential equations given by

$$\begin{cases} \nu(y)g(y) - \int_0^{+\infty} k(x)g(x)\,dx\,[D(y)g'(y)]' = p(y) & \text{for } y \geq 0, \\ g'(0)=0, \quad g(+\infty)=0, \end{cases} \tag{1.1}$$

that is, a second-order boundary value problem on the half-line where the coefficients of the derivatives of the unknown function depend on the function itself by means of an integral over the semiaxis. This kind of problem arises from important applications such as kinetics of plasma, population dynamics and thermodynamical equilibrium [Laitinen and Tiihonen 1998; Ratynskaia et al. 2007; Ricci et al. 2001; de Angelis et al. 2006; Takeuchi et al. 2007; Cannon and Galiffa 2008; 2011; Junghanns et al. 2014]. However, the above mentioned dependence makes it difficult to approach both from an analytical and a numerical point of view.

The aim of the present investigation is a theoretical analysis of problem (1.1), which provides useful information about the solution itself and represents an essential preparation for the numerical approach to the problem [Basile et al. 2012].

MSC2010: 34B40, 34K05, 45J05.

Keywords: boundary value problems, nonlinear integrodifferential equations, half-line.

A complete analysis of some related problems has been performed in [Cannon and Galiffa 2008; 2011]. The problems taken into account in such papers are of the type

$$\alpha\left(\int_0^1 g(x)\,dx\right)g''(y) = p(y), \tag{1.2}$$

$$\alpha\left(\int_0^1 g(x)\,dx\right)g''(y) + (g(y))^{2\eta+1} = 0, \tag{1.3}$$

where $0 < y < 1$, $g(0)$ and $g(1)$ are given, η is a nonnegative integer, and α is a given function. Although these problems contain the same peculiarity as problem (1.1), there are some different characteristics (for example, the function α in (1.2) and (1.3), the nonlinearity $g^{2\eta+1}$ in (1.3), and the presence of g' and the infinite domain of integration in (1.1)). Because of these differences, the existence and uniqueness results obtained in [Cannon and Galiffa 2008; 2011] cannot be directly applied to our case, and a specific analysis is carried out in the following sections. First we set

$$q = \int_0^{+\infty} k(x)g(x)\,dx$$

and rewrite (1.1) as

$$\begin{cases} v(y)g(y) - q[D(y)g'(y)]' = p(y), & y \geq 0, \\ g'(0) = 0, \quad g(+\infty) = 0. \end{cases} \tag{1.4}$$

Of course the solution of (1.4) depends on q, and when we want to underscore this dependence we will use the notation $g(y, q)$ instead of $g(y)$. In Section 2 we will examine problem (1.4), taking $q > 0$ as a fixed parameter, and report results about existence, uniqueness, positiveness, regularity and boundedness of the solution. This section contains an elaboration of known results (see, for example, [Granas et al. 1978; 1986]) and will serve the investigations on the complete problem (1.1) carried out in Section 3, where we prove the existence of a nonnegative solution g that is uniformly bounded together with its derivatives. The properties of the solutions of (1.1) reported in that section are helpful in the comprehension of the problem itself and its numerical analysis.

2. Analysis of the solution of problem (1.4)

In this section we consider q fixed and positive so that our equation reduces to a classical Sturm–Liouville boundary value problem. We report some results on the existence and the uniqueness of the solution $g(y)$ of problem (1.4) along with the analysis of other useful properties such as the sign of g and the boundedness of $g(y)$, $g'(y)$ and $g''(y)$.

In order to make the aim of this paper clear, we want to specify that most of the results reported here are already known in the literature (and this will be evident throughout). However an effort has been made to fit them into the form of problem (1.4) in order to prepare the basis for the analysis of problem (1.1), which will be performed in the following section.

From now on we will consider boundary value problems of the kind (1.4) with q positive and fixed, and we make the following assumptions on the involved functions:

(1) $D \in C^1([0, +\infty))$, ν, $p \in C([0, +\infty))$,

(2) $0 < D_{\inf} \le D(y) \le D_{\sup}$, $y \ge 0$,

(3) $0 < \sup_{y \ge 0} |D'(y)/D(y)| < +\infty$,

(4) $0 < \nu_{\inf} \le \nu(y) \le \nu_{\sup}$, $y \ge 0$,

(5) $0 \le p(y) \le P$, $y \ge 0$,

(6) $\int_0^{+\infty} p(y)\, dy < +\infty$,

(7) $\lim_{y \to +\infty} p(y) = 0$.

Theorem 2.1. *Assume* (1)–(7) *are satisfied. Then for any* $q > 0$, *the boundary value problem* (1.4) *has a unique nonnegative solution* $g \in C^2([0, +\infty))$.

Proof. Following a standard procedure (see, for example, the proof of Theorem 2.2 in [Granas et al. 1986]), starting from the solutions g_n of

$$\begin{cases} \nu(y) g_n(y) - q[D(y) g_n'(y)]' = p(y) & \text{for } 0 \le y \le n, n \in N, \\ g_n'(0) = 0, \quad g_n(n) = 0 \end{cases}$$

and applying inductively the Ascoli–Arzelà theorem we get the existence of a solution $g \in C^2([0, +\infty))$ of (1.4) satisfying only the first boundary condition.

Then using hypotheses (2) and (4), we set

$$m := (q D_{\inf} \nu_{\inf})^{1/2},$$

and consider the function [Agarwal and O'Regan 2001]

$$\omega(y) := \frac{1}{2m} e^{-\frac{m}{q} \int_0^y \frac{ds}{D(s)}} \left(\int_0^{+\infty} p(\tau) e^{-\frac{m}{q} \int_0^\tau \frac{ds}{D(s)}} \, d\tau + \int_0^y p(\tau) e^{\frac{m}{q} \int_0^\tau \frac{ds}{D(s)}} \, d\tau \right) + \frac{1}{2m} e^{\frac{m}{q} \int_0^y \frac{ds}{D(s)}} \int_y^{+\infty} p(\tau) e^{-\frac{m}{q} \int_0^\tau \frac{ds}{D(s)}} \, d\tau.$$

Hence, proceeding as in the proof of Theorem 1.8.3 in [Agarwal and O'Regan 2001], it can be proved that

$$0 \le g(y) \le \omega(y) \quad \text{and} \quad \lim_{y \to +\infty} \omega(y) = 0,$$

which ensures that the second boundary condition in (1.4) is satisfied too. Finally, the uniqueness of the solution of problem (1.4) can be proved by standard arguments showing that the homogeneous problem has only the trivial solution. □

In order to show the boundedness of $g(y)$ and its first and second derivatives, we note that hypotheses (2)–(5) allow us to define the constants

$$A := \left\| \frac{D'}{D} \right\|_\infty, \quad B := \left\| \frac{D'}{D} \right\|_\infty + \left\| \frac{\nu}{qD} \right\|_\infty \left\| \frac{p}{\nu} \right\|_\infty + \left\| \frac{p}{qD} \right\|_\infty, \tag{2.5}$$

where, as usual, $\|f\|_\infty = \sup_{y\geq 0} |f(y)|$. Moreover, we set

$$r_0 := \left\| \frac{p}{\nu} \right\|_\infty, \tag{2.6}$$

$$r_1 := \left[\frac{B}{A}(e^{2Ar_0} - 1) \right]^{1/2}, \tag{2.7}$$

$$r_2 := \left\| \frac{D'}{D} \right\|_\infty r_1 + \left\| \frac{\nu}{qD} \right\|_\infty r_0 + \left\| \frac{p}{qD} \right\|_\infty. \tag{2.8}$$

Theorem 2.2. *Under the assumptions* (1)–(7), *for any fixed* $q > 0$, *the unique solution* $g(y)$ *of* (1.4) *satisfies the bounds*

$$g(y) \leq r_0 \quad \textit{for } y \geq 0, \tag{2.9}$$

$$|g'(y)| \leq r_1 \quad \textit{for } y \geq 0, \tag{2.10}$$

$$|g''(y)| \leq r_2 \quad \textit{for } y \geq 0. \tag{2.11}$$

Proof. The boundedness of g follows by observing that $g(+\infty) = 0$ and that g cannot have a local maximum point $\bar{y} \geq 0$ such that $g(\bar{y}) > r_0$ (see, for example, [Agarwal and O'Regan 2001, page 18]). The remaining part of the proof is based on an idea developed in [Granas et al. 1978, page 71]. Therefore, we give here only a brief sketch of it. From (1.4) and hypothesis (2), we have

$$g''(y) = -\frac{D'(y)}{D(y)} g'(y) + \frac{\nu(y)}{qD(y)} g(y) - \frac{p(y)}{qD(y)} \quad \text{for } y \geq 0. \tag{2.12}$$

Then we deduce the *Bernstein growth condition*

$$|g''(y)| \leq Ag'^2(y) + B \quad \text{for } y \geq 0,$$

which, multiplying both sides by $2A|g'(y)|$, gives

$$\frac{2Ag'(y)g''(y)}{Ag'^2(y) + B} \leq 2A|g'(y)| \quad \text{for } y \geq 0, \tag{2.13}$$

with A and B defined in (2.5).

Observe that any $y > 0$ such that $g'(y) \neq 0$ belongs to an interval $[a, b]$ where g' does not change sign and $g'(a) = 0$. Hence, (2.10) follows by integrating (2.13) from a to y and by recalling that $0 \leq g(y) \leq r_0$ for all $y \geq 0$.

Finally, from (2.9), (2.10), (2.12) and the hypotheses (2)–(5), we easily obtain (2.11). □

Remark. The positiveness of the solution g arises from the positiveness of the right side p in (1.4). However, if no information on the sign of p is given, we can still say that a unique solution $g(y)$ of the problem (1.4) exists, and (2.9) becomes

$$|g(y)| < r_0 \quad \text{for } y \geq 0. \tag{2.14}$$

Corollary 2.3. *Assume* (1)–(7) *hold. Then* $\lim_{y\to+\infty} g'(y) = 0$.

Proof. From Theorem 2.2 we know that g'' is bounded for any fixed value of the parameter q; hence, g' is uniformly continuous for all $y \geq 0$. From here and Barbalat's lemma [Sun 2009], we have $\lim_{y\to+\infty} g'(y) = 0$. □

We now prove other useful properties of g. Denote by $\mathrm{BC}^{(r)}[0, +\infty)$ the space of functions $f(x)$ with $f^{(j)}(x)$, $j = 0, 1, \ldots, r$, bounded and continuous on $[0, +\infty)$. Then observe that if $g(y)$ is a solution of (1.4) it satisfies (2.12); hence, the proof of the following theorem is straightforward.

Theorem 2.4. *Let $r \in N$. In addition to* (1)–(7), *assume $p, v \in \mathrm{BC}^r[0, +\infty)$ and $D \in \mathrm{BC}^{r+1}[0, +\infty)$. Then, for any fixed $q > 0$, the solution g of* (1.4) *is in $\mathrm{BC}^{r+2}[0, +\infty)$. Moreover, for any $\bar{q} > 0$, the derivatives $g^{(j)}(y)$, $j = 0, \ldots, r$, are uniformly bounded with respect to $q \in [\bar{q}, +\infty)$.*

All these properties together with the uniform continuity of g as function of q, that we are going to prove in the following section represent the basic material to deal with the difficult task of proving the existence of the solution of the original problem (1.1).

3. Existence of the solution of problem (1.1)

In this section, the focus of our attention will be the model (1.1), whose analysis requires all the results already described for (1.4). Whereas the results in Section 2 for problem (1.4) with fixed $q > 0$ are mainly obtained by elaborations of existing studies, the investigations we start in this section represent new contributions.

In Section 2, we showed that for any q fixed and positive there exists a unique solution $g(y, q)$ of (1.4). Thus, the function

$$F(q) := q - \int_0^{+\infty} k(x) g(x, q)\, dx \quad \text{for } q > 0 \tag{3.15}$$

is well defined, where the kernel k is assumed to satisfy

(8) $k \in C^2([0,+\infty))$,

(9) $\int_0^{+\infty} k(x)\,dx < +\infty$,

(10) $k(x) \geq 0$ for $x \in [0,+\infty)$.

To prove the existence of a solution of (1.1), we show that there exists a solution of the equation $F(q) = 0$. Specifically, if F is continuous and there exist two positive values a and b such that $F(a)F(b) < 0$, then by the intermediate value theorem, equation $F(q) = 0$ has at least one solution q^* and the corresponding function $g(y,q^*)$ is the solution of (1.1).

Theorem 3.1. *Assume that hypotheses* (1)–(10) *hold. Then* $F(q)$ *is uniformly continuous on* $[\bar{q},+\infty)$, *for all* $\bar{q} > 0$.

Proof. Let us prove that $g(y,q)$ is uniformly continuous with respect to $q \geq \bar{q}$ and $y \geq 0$; that is, for all $\epsilon > 0$ there exists $\delta_\epsilon > 0$ such that

$$|g(y,q_1) - g(y,q_2)| < \epsilon \quad \forall q_1, q_2 \text{ such that } |q_1 - q_2| < \delta_\epsilon \text{ and } \forall y \geq 0. \tag{3.16}$$

Let q_1, $q_2 \geq \bar{q}$ be arbitrarily fixed. Then the functions $g(y,q_1)$ and $g(y,q_2)$ satisfy, respectively,

$$\begin{cases} \nu(y)g(y,q_1) = q_1[D(y)g'(y,q_1)]' + p(y) & \text{for } y \geq 0, \\ g'(0,q_1) = 0, \qquad g(+\infty,q_1) = 0, \end{cases} \tag{3.17}$$

$$\begin{cases} \nu(y)g(y,q_2) = q_2[D(y)g'(y,q_2)]' + p(y) & \text{for } y \geq 0, \\ g'(0,q_2) = 0, \qquad g(+\infty,q_2) = 0, \end{cases} \tag{3.18}$$

Subtracting both sides of (3.17) and (3.18), we see that

$$e(y) = g(y,q_1) - g(y,q_2)$$

is a solution of

$$\begin{cases} \nu(y)e(y) = q_1[D(y)e'(y)]' + (q_1 - q_2)[D(y)g'(y,q_2)]' & \text{for } y \geq 0, \\ e'(0) = 0, \qquad e(+\infty) = 0, \end{cases}$$

Hence, thanks to (2.14) the inequality

$$|e(y)| \leq |q_1 - q_2| \sup_{y \geq 0} \frac{\big|[D(y)g'(y,q_2)]'\big|}{\nu(y)} \tag{3.19}$$

holds, where, by using hypotheses (2) and (4) and Theorem 2.4, it comes out that

$$\sup_{y \geq 0} \frac{\big|[D(y)g'(y,q_2)]'\big|}{\nu(y)} \leq M_0, \tag{3.20}$$

with M_0 independent of the parameters q_1, $q_2 \geq \bar{q}$. Thus, we conclude that for all $\epsilon > 0$, there exists $\delta_\epsilon = \epsilon/M_0 > 0$ such that (3.16) holds. The desired result on F is achieved by noting that from (2.9) and hypotheses (9) and (10), the improper integral $\int_0^{+\infty} k(x)g(x,q)\,dx$ is uniformly convergent with respect to q and we are allowed to take the limit under the integral. □

In the following theorem, we find an interval $[a,b]$ where the function F changes its sign. In order to provide the explicit values for a and b, the additional hypotheses (11)–(14) are required.

Theorem 3.2. *Let $F(q)$ be the function defined in* (3.15), *assume that hypotheses* (1)–(10) *hold*, *and that*

(11) $\nu \in C^2([0,+\infty))$,

(12) $|\nu'(y)| < c \quad$ *for all* $y \geq 0$,

(13) $|k'(y)| < k_1 \quad$ *for all* $y \geq 0$,

(14) $\int_0^{+\infty} \left| \left[\left(\frac{k(y)}{\nu(y)} \right)' D(y) \right]' \right| dy = C_1 < \infty, \quad \int_0^{+\infty} \frac{k(y)}{\nu(y)} p(y)\,dy = C_2 < \infty.$

Then there exist $a, b \in (0,+\infty)$ *such that* $F(a)F(b) \leq 0$.

Proof. By (1.4) we have

$$\begin{aligned} F(q) &= q - \int_0^{+\infty} k(x)g(x,q)\,dx \\ &= q\left(1 - \int_0^{+\infty} \frac{k(x)}{\nu(x)} \big(D(x)g'(x,q)\big)'\,dx\right) - \int_0^{+\infty} \frac{k(x)}{\nu(x)} p(x)\,dx. \end{aligned}$$

Integrating twice by parts, by (11), (13) and Corollary 2.3 we get

$$\begin{aligned} &\int_0^{+\infty} \frac{k(x)}{\nu(x)} \big(D(x)g'(x,q)\big)'\,dx \\ &\qquad = \left[\frac{k}{\nu}\right]'(0)g(0)D(0) + \int_0^{+\infty} \left[\left(\frac{k(x)}{\nu(x)} \right)' D(x) \right]' g(x,q)\,dx. \end{aligned}$$

Thus

$$F(q) \leq q\big[1 + r_0(C_0 + C_1)\big] - C_2, \tag{3.21}$$

where r_0 and C_1, C_2 are defined, respectively, in (2.6) and (14), and

$$C_0 = \left| \left[\frac{k}{\nu}\right]'(0) \right| D(0).$$

By (3.21), $F(q) \leq 0$ for any $q \leq a$ with

$$a := \frac{C_2}{1 + r_0(C_0 + C_1)}. \tag{3.22}$$

Finally observe that from (2.9) and hypothesis (10) we have

$$F(q) = q - \int_0^{+\infty} k(x) g(x,q)\,dx \geq q - r_0 \int_0^{+\infty} k(x)\,dx, \tag{3.23}$$

which gives $F(q) \geq 0$ for any $q \geq b$ with

$$b := r_0 \int_0^{+\infty} k(x)\,dx, \tag{3.24}$$

completing the proof. □

From Theorem 3.2, by using the intermediate value theorem, we get our main result.

Theorem 3.3. *Assume that* (1)–(14) *hold. Then there exists at least one solution g of problem* (1.1) *such that*

$$a \leq \int_0^{+\infty} k(x) g(x)\,dx \leq b,$$

where a and b are defined in (3.22) *and* (3.24).

Observe that this theorem requires only the continuity of F, and it gives the existence but does not assure the uniqueness of the solution of (1.1). By exploiting the uniform continuity of F the following uniqueness result can be proved.

Theorem 3.4. *Assume that* (1)–(10) *and*

(15) $M_0 \|k\|_1 \leq 1$

hold, where M_0 is given in (3.20). *Then problem* (1.1) *has a unique solution.*

Proof. The statement follows easily from (3.19), (15) and the Banach fixed point theorem. □

Since a solution of (1.1) is a solution of (1.4), it satisfies all the properties reported in Section 2. In particular, under the hypotheses of Theorem 3.2 and from (2.6), we define

$$r1_a := \left\{ \left[1 + \left(\left\| \frac{\nu}{aD} \right\|_\infty r_0 + \left\| \frac{p}{aD} \right\|_\infty \right) \left\| \frac{D'}{D} \right\|_\infty^{-1} \right] \left(e^{2\|D'/D\|_\infty r_0} - 1 \right) \right\}^{1/2},$$

$$r2_a := \left\| \frac{D'}{D} \right\|_\infty r1_a + \left\| \frac{\nu}{aD} \right\|_\infty r_0 + \left\| \frac{p}{aD} \right\|_\infty.$$

Thus we have for $y \geq 0$

$$0 \leq g(y) \leq r_0, \quad |g'(y)| \leq r1_a, \quad |g''(y)| \leq r2_a.$$

Compared to (2.6)–(2.8) these bounds are independent of q and they turn to be useful in the numerical analysis of the problem that we carried out in [Basile et al. 2012].

References

[Agarwal and O'Regan 2001] R. P. Agarwal and D. O'Regan, *Infinite interval problems for differential, difference and integral equations*, Kluwer, Dordrecht, 2001. MR 2002g:34058 Zbl 0988.34002

[de Angelis et al. 2006] U. de Angelis, G. Capobianco, C. Marmolino, and C. Castaldo, "Fluctuations in dusty plasmas", *Plasma Phys. Control. Fusion* **48**:12B (2006), 91–98.

[Basile et al. 2012] M. Basile, E. Messina, W. Themistoclakis, and A. Vecchio, "A numerical method for a class of non-linear integro-differential equations on the half line", *Comput. Math. Appl.* **64**:7 (2012), 2354–2363. MR 2966871

[Cannon and Galiffa 2008] J. R. Cannon and D. J. Galiffa, "A numerical method for a nonlocal elliptic boundary value problem", *J. Integral Equations Appl.* **20**:2 (2008), 243–261. MR 2009m:35121 Zbl 1149.65099

[Cannon and Galiffa 2011] J. R. Cannon and D. J. Galiffa, "On a numerical method for a homogeneous, nonlinear, nonlocal, elliptic boundary value problem", *Nonlinear Anal.* **74**:5 (2011), 1702–1713. MR 2012b:65099 Zbl 1236.34028

[Granas et al. 1978] A. Granas, R. B. Guenther, and J. W. Lee, "On a theorem of S. Bernstein", *Pacific J. Math.* **74**:1 (1978), 67–82. MR 57 #10068 Zbl 0377.34003

[Granas et al. 1986] A. Granas, R. B. Guenther, J. W. Lee, and D. O'Regan, "Boundary value problems on infinite intervals and semiconductor devices", *J. Math. Anal. Appl.* **116**:2 (1986), 335–348. MR 87m:34013 Zbl 0594.34019

[Junghanns et al. 2014] P. Junghanns, W. Themistoclakis, and A. Vecchio, "Fixed point iterations for a class of nonstandard Sturm–Liouville boundary value problems", *Nonlinear Analysis* (2014). To appear.

[Laitinen and Tiihonen 1998] M. T. Laitinen and T. Tiihonen, "Integro-differential equation modelling heat transfer in conducting, radiating and semitransparent materials", *Math. Methods Appl. Sci.* **21**:5 (1998), 375–392. MR 98m:45016 Zbl 0958.80003

[Ratynskaia et al. 2007] S. Ratynskaia, M. De Angeli, U. de Angelis, C. Marmolino, G. Capobianco, M. Lontano, E. Lazzaro, G. E. Morfill, and G. Gervasini, "Observation of the effects of dust particles on plasma fluctuation spectra", *Phys. Rev. Lett.* **99**:7 (2007). Article ID # 075002.

[Ricci et al. 2001] P. Ricci, G. Lapenta, U. de Angelis, and V. N. Tsytovich, "Plasma kinetics in dusty plasmas", *Phys. Plasmas* **8**:3 (2001), 769–776.

[Sun 2009] M. Sun, "A Barbalat-like lemma with its application to learning control", *IEEE Trans. Automat. Control* **54**:9 (2009), 2222–2225. MR 2010i:93091

[Takeuchi et al. 2007] Y. Takeuchi, Y. Iwasa, and K. Sato (editors), *Mathematics for life science and medicine*, Springer, Berlin, 2007. MR 2008a:92006 Zbl 1103.92301

Received: 2012-12-03 Revised: 2013-02-27 Accepted: 2013-04-12

mariateresa.basile@unina.it
Dipartimento di Matematica e Applicazioni,
Università degli Studi di Napoli "Federico II",
Via Cintia, Monte San Angelo, I-80126 Napoli, Italy

woula.themistoclakis@cnr.it
Istituto per le Applicazioni del Calcolo "Mauro Picone",
CNR National Research Council of Italy,
Via P. Castellino, 111, I-80131 Napoli, Italy

antonia.vecchio@cnr.it
Istituto per le Applicazioni del Calcolo "Mauro Picone",
CNR National Research Council of Italy,
Via P. Castellino, 111, I-80131 Napoli, Italy

INVOLVE 7:1 (2014)
dx.doi.org/10.2140/involve.2014.7.77

Homogenization of a nonsymmetric embedding-dimension-three numerical semigroup

Seham Abdelnaby Taha and Pedro A. García-Sánchez

(Communicated by Scott T. Chapman)

Let n_1, n_2, n_3 be positive integers with $\gcd(n_1, n_2, n_3) = 1$. For $S = \langle n_1, n_2, n_3\rangle$ nonsymmetric, we give an alternative description, using elementary techniques, of a minimal presentation of its homogenization $\bar{S} = \langle(1, 0), (1, n_1), (1, n_2), (1, n_3)\rangle$. As a consequence, we show that this minimal presentation is unique. We recover Bresinsky's characterization of the Cohen–Macaulay property of $\bar{S}$ and present a procedure to compute all possible catenary degrees of the elements of $\bar{S}$.

Introduction

An *affine semigroup* is a finitely generated submonoid of $\mathbb{N}^k$ for some positive integer k, where $\mathbb{N}$ stands for the set of nonnegative integers. Every affine semigroup admits a unique minimal generating system (see Exercise 6 in [Rosales and García-Sánchez 1999, Chapter 3]). Let S be an affine semigroup and let $A = \{n_1, \dots, n_e\}$ be its unique minimal generating system. Then the monoid morphism $\varphi\colon \mathbb{N}^e \to S$ induced by $e_i \mapsto n_i$ (e_i stands for the i-th row of the $e \times e$ identity matrix) is an epimorphism. Therefore S is isomorphic as a monoid to $\mathbb{N}^e/\ker\varphi$, where $\ker\varphi = \{(a, b) \in \mathbb{N}^e \times \mathbb{N}^e \mid \varphi(a) = \varphi(b)\}$ is the kernel congruence of S. A generating set for $\ker\varphi$ is known as a presentation for S, and it is a *minimal presentation* if it is minimal with respect to set inclusion (or equivalently, if it is minimal with respect to cardinality in view of [Rosales and García-Sánchez 1999, Corollary 9.5], which is finite). The monoid S is said to be uniquely presented if it has a unique minimal presentation (see [García-Sánchez and Ojeda 2010]).

The monoid morphism φ is sometimes called the factorization morphism associated to S. This is because for $s \in S$, the set $\mathsf{Z}(s) = \varphi^{-1}(s)$ corresponds with

MSC2010: 20M14, 20M25.

Keywords: numerical semigroup, catenary degree, projective monomial curve, homogeneous catenary degree.

Taha is supported by the Spanish AECID program. García-Sánchez is supported by the projects MTM2010-15595, FQM-343 and FQM-5849, and FEDER funds. The authors would like to thank Ignacio Ojeda, Aureliano M. Robles-Pérez and the referee for their comments and suggestions. This manuscript will be part of Taha's master's thesis.

the *set of factorizations* of s if we identify the free monoid on A with $\mathbb{N}^e$ (the elements in A are sometimes called the atoms or irreducible elements of S). The set of factorizations of s has finitely many elements (see, for instance, [Rosales and García-Sánchez 1999, Lemma 9.1]), and corresponds to the set of nonnegative integer solutions of a system of linear Diophantine equations $xB = s$ (where B denotes the matrix whose rows are $n_1, \dots, n_e$). An element $s \in S$ is said to have *unique expression* if the cardinality of $\mathsf{Z}(s)$ is one. If every element has unique expression, the monoid is *factorial*; in this case, $\ker\varphi$ is trivial and S is isomorphic to $\mathbb{N}^e$.

For a factorization $x = (x_1, \dots, x_e) \in \mathsf{Z}(s)$, its *support* is the set

$$\operatorname{supp}(x) = \{n_i \mid x_i \neq 0\},$$

that is, it is the set of atoms involved in the factorization x. For a given factorization $x = (x_1, \dots, x_e) \in \mathsf{Z}(s)$, its *length* is $|x| = x_1 + \cdots + x_e$. The *set of lengths* of s is $\mathsf{L}(s) = \{|x| \mid x \in \mathsf{Z}(s)\}$. When the set of lengths of all the elements have cardinality one, then the monoid is said to be *half-factorial*.

A minimal presentation of S can be computed as described in [Rosales and García-Sánchez 1999, Chapter 9]. We briefly explain this procedure. For $s \in S$, define the graph G_s whose vertices are

$$\mathrm{V}(\mathrm{G}_s) = \{a \in A \mid s - a \in S\}$$

(the atoms "dividing" s), and edges

$$\mathrm{E}(\mathrm{G}_s) = \{ab \mid a, b \in A \text{ and } s - (a + b) \in S\}.$$

On $\mathsf{Z}(s)$ define the relation $\mathcal{R}$ as follows: $x \mathcal{R} y$ if there exists $x_1, \dots, x_k \in \mathsf{Z}(s)$ such that

- $x_1 = x$, $x_k = y$, and
- for every $i \in \{1, \dots, k-1\}$, $x_i \cdot x_{i+1} \neq 0$ (or equivalently, $\operatorname{supp}(x_i) \cap \operatorname{supp}(x_{i+1})$ is not empty).

Proposition 9.7 in [Rosales and García-Sánchez 1999] states that there is a bijective map between the set of $\mathcal{R}$-classes of $\mathsf{Z}(s)$ and the set of nonconnected components of G_s: for every connected component C of G_s, there exists $x \in \mathsf{Z}(s)$ whose support is contained in the vertices of C; the map sends C to the $\mathcal{R}$-class containing x. Let $R_1, \dots, R_t$ be the different $\mathcal{R}$-classes of $\mathsf{Z}(s)$, and take $x_i \in R_i$ for every i. Define $\rho_s = \{(x_1, x_2), \dots, (x_{t-1}, x_t)\}$ (actually, one can choose any set of pairs corresponding to the edges of a spanning tree of the complete graph with vertices $\{x_1, \dots, x_t\}$; if $t = 1$, then $\rho_i = \varnothing$). Then

$$\rho = \bigcup_{s \in S} \rho_s$$

is a minimal presentation of S. This union in fact ranges only over the elements $s \in S$ such that G_s is not connected. These elements are called *Betti elements* of S, and the set of Betti elements of S will be denoted by $\mathrm{Betti}(S)$.

Let k be a field. The semigroup ring associated to S is $k[S] = \bigoplus_{s \in S} kt^s$, where t is an indeterminate. Addition is performed componentwise, while the product is defined by distributivity and the rule $t^s t^{s'} = t^{s+s'}$. The monoid morphism φ has a ring analog $\bar{\varphi}\colon k[x_1, \dots, x_e] \to k[S]$, which is the morphism induced by $x_i \mapsto t^{n_i}$, $i \in \{1, \dots, e\}$, where $x_1, \dots, x_e$ are unknowns. Its kernel I_S is generated by

$$\bigl\{x_1^{a_1} \cdots x_e^{a_e} - x_1^{b_1} \cdots x_e^{b_e} \bigm| \bigl((a_1, \dots, a_e), (b_1, \dots, b_e)\bigr) \in \ker \varphi\bigr\}.$$

Indeed, σ is a minimal presentation if and only if

$$\bigl\{x_1^{a_1} \cdots x_e^{a_e} - x_1^{b_1} \cdots x_e^{b_e} \bigm| \bigl((a_1, \dots, a_e), (b_1, \dots, b_e)\bigr) \in \sigma\bigr\}$$

is a minimal generating system of I_S (see [Herzog 1970]).

Let S be a *numerical semigroup*, that is, a submonoid of $\mathbb{N}$ with finite complement in $\mathbb{N}$ (or equivalently, $\gcd(S) = 1$). It is easy to show that S admits a unique *minimal generating set* with finitely many elements, and thus every numerical semigroup is an affine semigroup. The cardinality of the minimal generating set of S is known as the *embedding dimension* of S. The largest integer not belonging to S is the *Frobenius number* of S, denoted $\mathrm{F}(S)$. The numerical semigroup S is *symmetric* if for every integer z not in S, $\mathrm{F}(S) - z \in S$.

Let S be a numerical semigroup minimally generated by $\{n_1, n_2, n_3\}$, where $n_1 < n_2 < n_3$. Define

$$c_i = \min\bigl\{k \in \mathbb{N} \setminus \{0\} \mid kn_i \in \langle n_j, n_k\rangle\bigr\},$$

where $\{i, j, k\} = \{1, 2, 3\}$. Thus there exists $r_{ij} \in \mathbb{N}$ such that

$$c_i n_i = r_{ij} n_j + r_{ik} n_k.$$

Also, we have $\mathrm{Betti}(S) = \{c_1 n_1, c_2 n_2, c_3 n_3\}$ [Rosales and García-Sánchez 2009, Example 8.23]. If S is not symmetric, then these r_{ij} are unique (see [Herzog 1970]) and

$$\sigma = \bigl\{\bigl((c_1, 0, 0), (0, r_{12}, r_{13})\bigr), \bigl((0, c_2, 0), (r_{21}, 0, r_{23})\bigr), \bigl((0, 0, c_3), (r_{31}, r_{32}, 0)\bigr)\bigr\}$$

is essentially the unique minimal presentation of S (that is, if τ is any other minimal presentation and $(a, b) \in \tau$, then either $(a, b) \in \sigma$ or $(b, a) \in \sigma$). Moreover, we have

$$\begin{aligned}
\mathsf{Z}(c_1 n_1) &= \{(c_1, 0, 0), (0, r_{12}, r_{13})\},\\
\mathsf{Z}(c_2 n_2) &= \{(0, c_2, 0), (r_{21}, 0, r_{23})\},\\
\mathsf{Z}(c_3 n_3) &= \{(0, 0, c_3), (r_{31}, r_{32}, 0)\}.
\end{aligned}$$

We also have the following relations.

- Since $c_1n_1 = r_{12}n_2 + r_{13}n_3$, we have $c_1n_1 > r_{12}n_1 + r_{13}n_1$. Hence
$$c_1 > r_{12} + r_{13},$$
and we set $\lambda = c_1 - r_{12} - r_{13}$.
- Since $c_3n_3 = r_{31}n_1 + r_{32}n_2$, we have $c_3n_3 < r_{31}n_3 + r_{32}n_3$. Hence
$$c_3 < r_{31} + r_{32},$$
and we set $\nu = r_{31} + r_{32} - c_3$.
- $c_i = r_{ji} + r_{ki}$ for every $\{i, j, k\} = \{1, 2, 3\}$ [Rosales and García-Sánchez 2009, Lemma 10.19].

Define $\bar{n}_i = (1, n_i)$, $i \in \{1, 2, 3\}$ and $\bar{n}_0 = (1, 0)$. Set $\bar{S} = \langle \bar{n}_0, \bar{n}_1, \bar{n}_2, \bar{n}_3\rangle$, which we call the homogenization of S since $I_{\bar{S}}$ corresponds with the homogenization of I_S (see [Cox et al. 2007, Chapter 8]; with the notation introduced there, $I_{\bar{S}} = I_S^h$). The ring $k[\bar{S}]$ is the coordinate ring of a monomial curve on $\mathbb{P}^3$.

We start with an example that illustrates Bresinsky's algorithm [1984] for computing a minimal presentation (and thus the Betti elements) of $\bar{S}$. We are going to make use of the Apéry set associated to an element in S. Let $m \in S \setminus \{0\}$. The *Apéry set* of m in S is defined as

$$\mathrm{Ap}(S, m) = \{s \in S \mid s - m \notin S\},$$

and has exactly m elements, one for each congruent class modulo m. (See [Rosales and García-Sánchez 2009, Chapter 1]; clearly, this definition applies to any monoid. We will use it later for $\bar{S}$, though in the general case this set might have infinitely many elements.)

Example 1. Let S_k be the numerical semigroup minimally generated by

$$\langle 10, 17 + 10k, 19 + 10k\rangle, \quad k \in \mathbb{N}.$$

In this setting, $n_1 = 10$, $n_2 = 17 + 10k$, and $n_3 = 19 + 10k$. This semigroup is not symmetric since its minimal generators are pairwise coprime (see [Rosales and García-Sánchez 2009, Chapter 9]).

First, we compute the values of $c_1, c_2, c_3, \lambda, \delta, \nu$ and r_{ij} for all k. Let us denote them with the superindex k. A minimal presentation for $S = S_0$ is

$$\big\{\big((4, 1, 0), (0, 0, 3)\big), \big((3, 0, 2), (0, 4, 0)\big), \big((7, 0, 0), (0, 3, 1)\big)\big\},$$

and thus we know these values for $k = 0$. Also it is easy to check that

$$\mathrm{Ap}(S, 10) = \{0, n_2, 2n_2, 3n_2, n_3, 2n_3, n_2 + n_3, 2n_2 + n_3, n_2 + 2n_3, 2n_2 + 2n_4\}$$

(one can use the package `numericalsgps` [Delgado et al. 2013] to do these computations).

Now let $k \geq 1$.

- $c_1^k = 7 + k4$. Observe that $(7+4k)10 = 3(17+10k) + (19+10k)$, which gives us $c_1^k \leq 7+4k$. If $x10 = a(17+10k) + b(19+10k)$, with $0 \neq x, a, b \in \mathbb{N}$, then we have $x10 = a17 + b19 + (a+b)k10$. We can deduce that if $x \leq (a+b)k$, then $a17+b19+(ak+bk-x)10 = 0$, and this implies that $a = 0$, $b = 0$ and $x = 0$, and this is impossible. If $x > (a+b)k$, then $(x-(a+b)k)10 = a17+b19$. This shows that $x-(a+b)k \geq c_1^0 = 7$. Hence $x \geq 7+(a+b)k$, so it remains to show that $a+b \geq 4$. So assume to the contrary that $a+b \leq 3$. Clearly $a17+b19 = (x-(a+b)k)10$ and $x-(a+b)k \geq 0$ imply that $a17+b19 \notin \operatorname{Ap}(S, 10)$. According to the shape of $\operatorname{Ap}(S, 10)$, this forces $a = 0$ and $b = 3$. However $3 \times 19 \neq (x-3k)10$ for any k. This proves that $x \geq 7+4k$, and consequently $c_1^k = 7+k4$. Since S^k is uniquely presented, we also have $r_{12}^k = 3$ and $r_{13}^k = 1$, whence $\lambda = 3+4k$.

- $c_2^k = 4$. Note that $4(17+10k) = (3+2k)10 + 2(19+10k)$. Assume that $y(17+10k) = a10 + b(19+10k)$ for some $0 \neq y, a, b \in \mathbb{N}$. Then $y17 = (a+bk-yk)10 + b19$. If $a+bk-yk \geq 0$, this implies that $y \geq c_2^0 = 4$. For $a+bk-yk < 0$, we get $b19 = y17 + (yk-a-bk)10$. Thus $b \geq c_3^0 = 3$. It follows that $y > a/k + b > b \geq 3$, and thus $y \geq 4$. Hence $c_2^k = 4$. Also we obtain that $r_{21}^k = 3+2k$, $r_{23}^k = 2$ and $\delta = 1+2k$.

- $c_3^k = 3$. We already know that $c_3^k = r_{13}^k + r_{23}^k = 1+2 = 3$.

Hence, we have

$$(7+4k)n_1 = 3n_2 + n_3, \quad 4n_2 = (3+2k)n_1 + 2n_3, \quad 3n_3 = (4+2k)n_1 + n_2,$$

and a minimal presentation for S^k is

$$\{((7+4k, 0, 0), (0, 3, 1)), ((0, 4, 0), (3+2k, 0, 2)), ((0, 0, 3), (4+2k, 1, 0))\}.$$

If we apply Bresinsky's algorithm to these equalities, from $3n_3 = (4+2k)n_1 + n_2$ and $4n_2 = (3+2k)n_1 + 2n_3$ $(4+2k \geq 3+3k)$ we obtain $5n_3 = n_1 + 5n_2$. We now proceed with $4n_2 = (3+2k)n_1 + 2n_3$ and $5n_3 = n_1 + 5n_2$, getting

$$(5+4)n_2 = (3+2k-1)n_1 + (5+2)n_3.$$

Then we continue with $(5+4)n_2 = (3+2k-1)n_1 + (5+2)n_3$ and $5n_3 = n_1 + 5n_2$, obtaining $(2 \times 5+4)n_2 = (3+2k-2)n_1 + (2 \times 5+2)n_3$. By repeating these steps we obtain the general term $(5i+4)n_2 = (3+2k-i)n_1 + (5i+2)n_3$, and we must stop whenever $5i+4 \geq 3+2k-i+5i+2$, or equivalently $i \geq 2k+1$. Hence we need $2k+1$ steps to end after the initial step $5n_3 = n_1 + 5n_2$, which together with the three initial relations yield $2k+5$ relators in a minimal presentation of $\bar{S}_k$.

Observe that each of these relations come from a different element in $\bar{S}_k$, and thus we also deduce that $\#\operatorname{Betti}(\bar{S}_k) = 2k+5$ for all $k \in \mathbb{N}$.

In particular this also shows that even if the cardinality of a minimal presentation of a nonsymmetric embedding-dimension-three numerical semigroup S is always three, the cardinality of a minimal presentation of $\bar{S}$ can be arbitrarily large.

Alternatively, we can use Theorem 4 in [Cox et al. 2007, Chapter 8] to compute a presentation of $\bar{S}$ from a minimal presentation of S.

Example 2. Let $S = \langle 10, 17, 19\rangle$. A minimal presentation for S is

$$\big\{\big((4,1,0),(0,0,3)\big),\big((3,0,2),(0,4,0)\big),\big((7,0,0),(0,3,1)\big)\big\}.$$

Hence, a minimal generating system of I_S is

$$\{x_1^4x_2 - x_3^3,\, x_1^3x_3^2 - x_2^4,\, x_1^7 - x_2^3x_3\}.$$

We compute a Gröbner basis of I_S with respect to the graded lexicographic ordering and obtain

$$\{x_1^4x_2 - x_3^3,\, x_1^3x_3^2 - x_2^4,\, x_1^7 - x_2^3x_3,\, x_1x_2^5 - x_3^5,\, x_1^2x_3^7 - x_2^9,\, x_2^{14} - x_1x_3^{12}\}.$$

Hence

$$\{x_1^4x_2 - x_0^2x_3^3,\, x_1^3x_3^2 - x_0x_2^4,\, x_1^7 - x_0^3x_2^3x_3,\, x_1x_2^5 - x_0x_3^5,\, x_1^2x_3^7 - x_2^9,\, x_2^{14} - x_0x_1x_3^{12}\}$$

is a generating system for $I_{\bar{S}}$. By Herzog's correspondence,

$$\big\{\big((0,4,1,0),(2,0,0,3)\big),\big((0,3,0,2),(1,0,4,0)\big),\big((0,7,0,0),(3,0,3,1)\big),\\ \big((0,1,5,0),(1,0,0,5)\big),\big((0,2,0,7),(0,0,9,0)\big),\big((0,0,14,0),(1,1,0,12)\big)\big\}$$

is a presentation of $\bar{S}$, though not a minimal presentation, since we saw in Example 1 that the cardinality of a minimal presentation is 5.

If we use the graded inverse lexicographic ordering instead, we obtain

$$\{x_1^4x_2 - x_3^3,\, x_1^3x_3^2 - x_2^4,\, x_1^7 - x_2^3x_3,\, x_1x_2^5 - x_3^5,\, x_1^2x_3^7 - x_2^9\},$$

which yields a minimal presentation for $\bar{S}$:

$$\big\{\big((0,4,1,0),(2,0,0,3)\big),\big((0,3,0,2),(1,0,4,0)\big),\big((0,7,0,0),(3,0,3,1)\big),\\ \big((0,1,5,0),(1,0,0,5)\big),\big((0,2,0,7),(0,0,9,0)\big)\big\}.$$

The Gröbner basis computations in this example have been performed with Maxima (http://maxima.sourceforge.net).

In the first section we describe the Betti elements of $\bar{S}$ and its unique minimal presentation. The second section recovers a test due to Bresinsky for the Cohen–Macaulay property of $\bar{S}$. Section 3 shows how the catenary degree of $\bar{S}$ (and thus the homogeneous catenary degree of S) can be computed.

1. Determining the set of Betti elements

In this section we depict $\mathrm{Betti}(\bar{S})$, the set of elements $\bar{n} \in \bar{S}$ such that $\mathrm{G}_{\bar{n}}$ is not connected, or equivalently, $\mathsf{Z}(\bar{n})$ has more than one $\mathcal{R}$-class. Theorems 2.7 and 2.9 in [Li et al. 2012] determine $\mathrm{Betti}(\bar{S})$ just by imposing that $\gcd\{n_1, n_2, n_3\} = 1$ (notice that $\bar{S}$ is isomorphic to $\langle (n_3, 0), (n_3 - n_1, n_1), (n_3 - n - 2, n_2), (0, n_3) \rangle$ [Rosales et al. 1998, Example 1.4]). Here we present an alternative description for the case $S = \langle n_1, n_2, n_3 \rangle$ is a nonsymmetric embedding-three numerical semigroup, and we obtain that in this setting $\bar{S}$ is uniquely presented.

Lemma 3. $\mathsf{Z}(c_1\bar{n}_1) = \{(0, c_1, 0, 0), (\lambda, 0, r_{12}, r_{13})\}$. *In particular, the graph* $\mathrm{G}_{c_1\bar{n}_1}$ *is not connected.*

Proof. We already know that $\{(0, c_1, 0, 0), (\lambda, 0, r_{12}, r_{13})\} \subseteq \mathsf{Z}(c_1\bar{n}_1)$. So assume that $(a_0, a_1, a_2, a_3) \in \mathsf{Z}(c_1\bar{n}_1)$. Then

$$a_0\bar{n}_0 + a_1\bar{n}_1 + a_2\bar{n}_2 + a_3\bar{n}_3 = c_1\bar{n}_1 = \lambda\bar{n}_0 + r_{12}\bar{n}_2 + r_{13}\bar{n}_3,$$

and in particular $c_1 n_1 = a_1 n_1 + a_2 n_2 + a_3 n_3$, which means that

$$(a_1, a_2, a_3) \in \mathsf{Z}(c_1 n_1) = \{(c_1, 0, 0), (0, r_{12}, r_{13})\}.$$

It follows that if $(a_1, a_2, a_3) = (c_1, 0, 0)$, then $(a_0, a_1, a_2, a_3) = (0, c_1, 0, 0)$, and if $(a_1, a_2, a_3) = (0, r_{12}, r_{13})$, we get $(a_0, a_1, a_2, a_3) = (\lambda, 0, r_{12}, r_{13})$. □

Lemma 4. *Let* $\bar{n} = a_0\bar{n}_0 + a_1\bar{n}_1 \neq c_1\bar{n}_1$, $a_0, a_1 \in \mathbb{N}$. *Then the graph* $\mathrm{G}_{\bar{n}}$ *is connected.*

Proof. Notice that if $a_1 = c_1$, then

$$a_0\bar{n}_0 + a_1\bar{n}_1 = a_0\bar{n}_0 + c_1\bar{n}_1 = (\lambda + a_0)\bar{n}_0 + r_{21}\bar{n}_2 + r_{13}\bar{n}_3.$$

As $\bar{n} \neq c_1\bar{n}_1$, $a_0 > 0$, and we get that $\mathrm{V}(\mathrm{G}_{\bar{n}}) = \{\bar{n}_0, \bar{n}_1, \bar{n}_2, \bar{n}_3\}$, and $\bar{n}_0\bar{n}_2$, $\bar{n}_0\bar{n}_3$, $\bar{n}_0\bar{n}_1 \in \mathrm{E}(\mathrm{G}_{\bar{n}})$, and thus $\mathrm{G}_{\bar{n}}$ is connected.

If $a_1 < c_1$, then $\bar{n}$ has unique expression, since if

$$a_0\bar{n}_0 + a_1\bar{n}_1 = b_0\bar{n}_0 + b_1\bar{n}_1 + b_2\bar{n}_2 + b_3\bar{n}_3$$

for some $b_0, b_1, b_2, b_3 \in \mathbb{N}$, then $a_1 n_1 = b_1 n_1 + b_2 n_2 + b_3 n_3$. By the minimality of c_1, we deduce that $b_1 \geq a_1$. But then $0 = (b_1 - a_1)n_1 + b_2 n_2 + b_3 n_3$, which leads to $a_1 = b_1$, $b_2 = b_3 = 0$. Since $\bar{n}$ has unique expression, the graph $\mathrm{G}_{\bar{n}}$ is connected.

Finally, if $a_1 > c_1$, then $a_0\bar{n}_0 + a_1\bar{n}_1 = (a_0+\lambda)\bar{n}_0 + (a_1 - c_1)\bar{n}_1 + r_{21}\bar{n}_2 + r_{13}\bar{n}_3$. In this setting, the graph $\mathrm{G}_{\bar{n}}$ is K_4, the complete graph on four vertices, whence connected. □

Lemma 5. $\mathsf{Z}(\nu\bar{n}_0 + c_3\bar{n}_3) = \{(r_{31}, r_{32}, 0, 0), (\nu, 0, 0, c_3)\}$. *In particular, the graph* $\mathrm{G}_{\nu\bar{n}_0+c_3\bar{n}_3}$ *is not connected.*

Proof. The proof goes as in Lemma 3. □

Lemma 6. *For every positive integer k, we have $k\bar{n}_3 \notin \langle \bar{n}_0, \bar{n}_1, \bar{n}_2\rangle$.*

Proof. This is because $\bar{n}_3$ is not in the cone spanned by $\{\bar{n}_0, \bar{n}_1, \bar{n}_2\}$ (which is the cone spanned by $\{\bar{n}_0, \bar{n}_2\}$). □

Let

$$c_2' = \min\{k \in \mathbb{N} \setminus \{0\} \mid k\bar{n}_2 \in \langle \bar{n}_0, \bar{n}_1, \bar{n}_3\rangle\}.$$

Assume that

$$c_2'\bar{n}_2 = \gamma\bar{n}_0 + r_{21}'\bar{n}_1 + r_{23}'\bar{n}_3,$$

with $\gamma, r_{21}', r_{23}' \in \mathbb{N}$.

Lemma 7. $\mathsf{Z}(c_2'\bar{n}_2) = \{(0, 0, c_2', 0), (\gamma, r_{21}', 0, r_{23}')\}$. *In particular,* $\mathrm{G}_{c_2'\bar{n}_2}$ *is not connected. Moreover,*

(1) $r_{23}' \neq 0$,

(2) *if* $r_{21}' = 0$, *then*

$$c_2' = \frac{n_3}{\gcd\{n_2, n_3\}} \quad \text{and} \quad r_{23}' = \frac{n_2}{\gcd\{n_2, n_3\}}.$$

Proof. Assume that $c_2'\bar{n}_2 = a_0\bar{n}_0 + a_1\bar{n}_1 + a_2\bar{n}_2 + a_3\bar{n}_3$ for some $a_0, a_1, a_2, a_3 \in \mathbb{N}$. The minimality of c_2' forces $a_2 = 0$. If $(a_0, a_1, a_3) \neq (\gamma, r_{21}', r_{23}')$, then assume without loss of generality that $a_0 \leq \gamma$. Then $(\gamma - a_0)\bar{n}_0 + r_{21}'\bar{n}_1 + r_{23}'\bar{n}_3 = a_1\bar{n}_1 + a_3\bar{n}_3$. Notice that $(a_1, a_3) \not\leq (r_{21}', r_{23}')$, since otherwise we would obtain

$$(\gamma - a_0)\bar{n}_0 + (r_{21}' - a_1)\bar{n}_1 + (r_{23}' - a_3)\bar{n}_3 = 0,$$

and consequently $(a_0, a_1, a_3) = (\gamma, r_{21}', r_{23}')$, a contradiction. Hence either $a_1 \geq r_{21}'$ and $a_3 < r_{23}'$, or $a_1 < r_{21}'$ and $a_3 \geq r_{23}'$. By Lemma 6, we have $a_1 \not\leq r_{21}'$. This leads to $a_3 \leq r_{23}'$ and $(a_1 - r_{21}')\bar{n}_1 = (\gamma - a_0)\bar{n}_0 + (r_{23}' - a_3)\bar{n}_3$. Hence $a_1 \geq c_1$, and consequently $c_2'\bar{n}_2 = (a_0+\lambda)\bar{n}_0 + (a_1 - c_1)\bar{n}_1 + r_{12}\bar{n}_2 + (a_3 + r_{13})\bar{n}_3$. But $r_{13} \neq 0$, and we have that $r_{12} \neq 0$, and this forces $c_2' > r_{12}$. Hence

$$(c_2' - r_{12})\bar{n}_2 = (a_0+\lambda)\bar{n}_0 + (a_1 - c_1)\bar{n}_1 + r_{12}\bar{n}_2 + r_{13}\bar{n}_3,$$

contradicting once more the minimality of c_2'. This shows that

$$\mathsf{Z}(c_2'\bar{n}_2) = \{(0, 0, c_2', 0), (\gamma, r_{21}', 0, r_{23}')\}.$$

Observe that $r'_{23} \neq 0$, since otherwise on the one hand $c'_2 = \gamma + r'_{21} \geq r'_{21}$, while on the other $c'_2 n_2 = r'_{21} n_1 < r'_{21} n_2$, which leads to $c'_2 < r'_{21}$, a contradiction.

If $r'_{21} = 0$, then $c'_2 n_2 = r'_{23} n_3$. Whenever $a_2 n_2 = a_3 n_3$ for some $a_2, a_3 \in \mathbb{N}$, we get $a_2 n_2 = a_3 n_3 > a_3 n_2$, whence $a_2 > a_3$. So $c'_2 n_2$ is the least multiple of n_2 that is a multiple of n_3, and we obtain $c'_2 = n_3 / \gcd\{n_2, n_3\}$. □

Lemma 8. *Let $a_0, a_2 \in \mathbb{N}$, with $a_2 > c'_2$. Then $\mathrm{G}_{a_0\bar{n}_0 + a_2\bar{n}_2}$ is connected.*

Proof. Set $\bar{n} = a_0 \bar{n}_0 + a_2 \bar{n}_2$.

Observe that $a_0\bar{n}_0 + a_2\bar{n}_2 = (a_0 + \gamma)\bar{n}_0 + r'_{21}\bar{n}_1 + (a_2 - c'_2)\bar{n}_2 + r'_{23}\bar{n}_3$, and thus $\bar{n}_0$, $\bar{n}_2$ and $\bar{n}_3$ are in the same connected component (and so is $\bar{n}_1$ if $r'_{21} \neq 0$).

We distinguish two cases.

- If $\bar{n}_1 \notin \mathrm{V}(\mathrm{G}_{\bar{n}})$, then r'_{21} must be zero and $\mathrm{G}_{\bar{n}}$ is connected with set of vertices $\{\bar{n}_0, \bar{n}_2, \bar{n}_3\}$.
- If $\bar{n}_1 \in \mathrm{V}(\mathrm{G}_{\bar{n}})$, then there must exist $b_0, b_1, b_2, b_3 \in \mathbb{N}$, $b_1 \neq 0$, such that $\bar{n} = b_0\bar{n}_0 + b_1\bar{n}_1 + b_2\bar{n}_2 + b_3\bar{n}_3$. If $b_0 + b_2 + b_3 \neq 0$, then $\bar{n}_1$ is in the same component as $\bar{n}_0$, $\bar{n}_2$ and $\bar{n}_3$, and thus $\mathrm{G}_{\bar{n}}$ is connected. If $b_0 = b_2 = b_3 = 0$, then $b_1\bar{n}_1 = a_0\bar{n}_0 + a_2\bar{n}_2$, which is clearly different from $c_1\bar{n}_1$, and thus Lemma 4 asserts that $\mathrm{G}_{\bar{n}}$ is connected. □

Lemma 9. *The only $k \in \mathbb{N}$ for which $\mathrm{G}_{k\bar{n}_2}$ is not connected is $k = c'_2$.*

Proof. If $k < c'_2$, then by the minimality of c'_2, $k\bar{n}_2$ has unique expression, whence $\mathrm{G}_{k\bar{n}_2}$ is connected. If $k > c'_2$, then Lemma 8 with $a_0 = 0$ and $a_2 = k$ asserts that $\mathrm{G}_{k\bar{n}_2}$ is connected. Finally, for $k = c'_2$, Lemma 7 ensures that $\mathrm{G}_{k\bar{n}_2}$ is not connected. □

For the rest of the discussion we need to distinguish between $c_2 \geq r_{21} + r_{23}$ and $c_2 < r_{21} + r_{23}$.

1.1. ***The case $c_2 \geq r_{21} + r_{23}$.*** Under the standing hypothesis, we have

$$\begin{aligned} c_1\bar{n}_1 &= \lambda\bar{n}_0 + r_{12}\bar{n}_2 + r_{13}\bar{n}_3, \\ c_2\bar{n}_2 &= \delta\bar{n}_0 + r_{21}\bar{n}_1 + r_{23}\bar{n}_3, \\ \nu\bar{n}_0 + c_3\bar{n}_3 &= r_{31}\bar{n}_1 + r_{32}\bar{n}_2, \end{aligned}$$

and all the coefficients appearing in these equations are nonzero, except eventually δ.

Lemma 10. $\mathsf{Z}(c_2\bar{n}_2) = \{(\delta, r_{21}, 0, r_{23}), (0, 0, c_2, 0)\}$. *In particular, the graph $\mathrm{G}_{c_2\bar{n}_2}$ is not connected.*

Proof. In this setting, $c'_2 = c_2$, and the proof follows from Lemma 7. □

Lemma 11. *Let $a_0, a_2 \in \mathbb{N}$, and let $\bar{n} = a_0\bar{n}_0 + a_2\bar{n}_2$. Assume that $\bar{n} \neq c_2\bar{n}_2$. Then the graph $\mathrm{G}_{\bar{n}}$ is connected.*

Proof. The proof goes as in Lemma 4, except for the case $a_2 > c_2 = c'_2$, for which we use Lemma 8. □

Lemma 12. *Let $a_0, a_3 \in \mathbb{N}$. Assume that $a_0\bar{n}_0+a_3\bar{n}_3 \neq \nu\bar{n}_0+c_3\bar{n}_3$. Then $\mathrm{G}_{a_0\bar{n}_0+a_3\bar{n}_3}$ is connected.*

Proof. Let $\bar{n} = a_0\bar{n}_0 + a_3\bar{n}_3$, and assume to the contrary that $\mathrm{G}_{\bar{n}}$ is not connected. Hence $\bar{n}$ admits at least another expression with support disjoint to the support of $a_0\bar{n}_0 + a_3\bar{n}_3$. This in particular means that $a_0 \neq 0$ by Lemma 6. Hence there exists $a_1, a_2 \in \mathbb{N}$ such that $a_0\bar{n}_0 + a_3\bar{n}_3 = a_1\bar{n}_1 + a_2\bar{n}_2$.

Since $a_0\bar{n}_0+a_3\bar{n}_3 = a_1\bar{n}_1+a_2\bar{n}_2$, we get $a_3n_3 = a_1n_1+a_2n_2$. By the minimality of c_3, we have $a_3 \geq c_3$. If $a_3 = c_3$, since $\mathsf{Z}(c_3n_3) = \{(0, 0, c_3), (r_{31}, r_{32}, 0)\}$, we deduce $a_1 = r_{31}$ and $a_2 = r_{32}$. If follows that $a_0 = \nu$, contradicting $\bar{n} \neq \nu\bar{n}_0 + c_3\bar{n}_3$. Hence $a_3 > c_3$.

If $a_1 \geq c_1$, then $a_0\bar{n}_0+a_3\bar{n}_3 = a_1\bar{n}_1+a_2\bar{n}_2 = (a_1-c_1)\bar{n}_1+(a_2+r_{12})\bar{n}_2+r_{13}\bar{n}_3$. For $a_1 > c_1$ we get that $\mathrm{G}_{\bar{n}}$ is connected. If $a_1 = c_1$, then a_2 cannot be zero, since otherwise $c_1n_1 = a_3n_3$, and c_1n_1 does not admit a factorization of the form $(0, 0, a_3)$. Again, in this setting we obtain that $\mathrm{G}_{\bar{n}}$ is connected, a contradiction.

In the same way we obtain a contradiction if $a_2 \geq c_2$. Hence $a_1 < c_1$ and $a_2 < c_2$. As $a_3n_3 = a_1n_1 + a_2n_2$ and σ is the unique minimal presentation of S, it can be deduced that $(r_{31}, r_{32}) < (a_1, a_2)$ (with the usual partial order; the equality does not hold since otherwise we would obtain $c_3 = a_3$). Hence

$$a_0\bar{n}_0 + a_3\bar{n}_3 = a_1\bar{n}_1 + a_2\bar{n}_2 = \nu\bar{n}_0 + (a_1 - r_{31})\bar{n}_1 + (a_2 - r_{32})\bar{n}_2 + c_3\bar{n}_3.$$

This forces $\mathrm{G}_{\bar{n}}$ to be connected (even if $a_0 = 0$; recall that $\{n_0\}$ is not a connected component), a contradiction. □

Theorem 13. *Let S be a nonsymmetric embedding-dimension-three numerical semigroup, with $c_2 \geq r_{21} + r_{23}$. Let $\bar{n} \in \bar{S}$. The graph $\mathrm{G}_{\bar{n}}$ is not connected if and only if*

$$\bar{n} \in \{c_1\bar{n}_1, c_2\bar{n}_2, \nu\bar{n}_0 + c_3\bar{n}_3\}.$$

Proof. The proof follows from Lemmas 3 to 12. □

Notice also that this result follows as a consequence of Bresinsky's algorithm, since in this setting, as $c_2 \geq r_{21} + r_{23}$, the procedure stops in the first step, and then we only have to homogenize the relations.

Example 14. Let $S = \langle 10, 13, 19\rangle$. The unique minimal presentation for S is

$$\big\{\big((2, 0, 1), (0, 3, 0)\big), \big((7, 0, 0), (0, 1, 3)\big), \big((5, 2, 0), (0, 0, 4)\big)\big\}.$$

In this example, $c_2 = 3 = r_{21} + r_{23}$. The Betti elements of S are 39, 70 and 76, while the Betti elements of $\bar{S}$ are $(3, 39)$, $(7, 76)$ and $(7, 70)$.

Remark 15. Notice that if $c_2 \geq r_{21} + r_{23}$, then, by using Buchberger's criterion (see, for instance, [Cox et al. 2007, Chapter 3]), it is not hard to show that

$$G = \big\{x_1^{c_1} - x_2^{r_{12}}x_3^{r_{13}}, x_2^{c_2} - x_1^{r_{21}}x_3^{r_{23}}, x_1^{r_{31}}x_2^{r_{32}} - x_3^{c_3}\big\}$$

is a reduced Gröbner basis with respect to any total degree ordering. Hence, in view of Theorem 4 in [Cox et al. 2007, Chapter 8], the homogenization of G

$$\{x_1^{c_1} - x_0^{\lambda} x_2^{r_{12}} x_3^{r_{13}}, x_2^{c_2} - x_0^{\delta} x_1^{r_{21}} x_3^{r_{23}}, x_1^{r_{31}} x_2^{r_{32}} - x_0^{\nu} x_3^{c_3}\}$$

would contain a minimal generating set for $I_{\bar{S}}$. None of the elements in this set are redundant, since they correspond to binomials associated to factorizations of different Betti elements of $\bar{S}$ (Lemmas 3, 10 and 5). This gives an alternative proof to Theorem 13 without using Lemmas 4, 6, 9, 8, 11 and 12.

Since all the elements in Betti$(\bar{S})$ have two factorizations, we get the following as a consequence of [García-Sánchez and Ojeda 2010, Corollary 5].

Corollary 16. *Let S be a nonsymmetric embedding-dimension-three numerical semigroup, with $c_2 \geq r_{21} + r_{23}$. Then*

$$\big\{\big((0, c_1, 0, 0), (\lambda, 0, r_{12}, r_{13})\big), \big((0, 0, c_2, 0), (\delta, r_{21}, 0, r_{31})\big), \\ \big((0, 0, 0, c_3), (\nu, r_{31}, r_{32}, 0)\big)\big\}$$

is the unique minimal presentation of $\bar{S}$.

1.2. ***The case $c_2 < r_{21} + r_{23}$.*** Recall that in this setting we have

$$\begin{aligned} c_1\bar{n}_1 &= \lambda\bar{n}_0 + r_{12}\bar{n}_2 + r_{13}\bar{n}_3, \\ \delta\bar{n}_0 + c_2\bar{n}_2 &= r_{21}\bar{n}_1 + r_{23}\bar{n}_3, \\ \nu\bar{n}_0 + c_3\bar{n}_3 &= r_{31}\bar{n}_1 + r_{32}\bar{n}_2. \end{aligned}$$

Lemma 17. $\mathsf{Z}(\delta n_0 + c_2\bar{n}_2) = \{(0, r_{21}, 0, r_{23}), (\delta, 0, c_2, 0)\}$. *In particular, the graph* $\mathrm{G}_{\delta\bar{n}_0 + c_2\bar{n}_2}$ *is not connected.*

Proof. Similar to the proof of Lemma 3. □

Remark 18. Observe that

$$d_2\bar{n}_2 = d_1\bar{n}_1 + d_3\bar{n}_3,$$

with $d_i = (n_j - n_k)/\gcd\{n_3 - n_2, n_2 - n_1\}$, $\{i, k < j\} = \{1, 2, 3\}$. Notice that the set of rational solutions of $\bar{n}_1 x_1 - \bar{n}_2 x_2 + \bar{n}_3 x_3 = 0$ is spanned by (d_1, d_2, d_3). And since $\gcd(d_1, d_2, d_3) = 1$, every integer solution (x_1, x_2, x_2) is a multiple of (d_1, d_2, d_3).

Observe also that

$$\frac{n_3}{\gcd\{n_2, n_3\}} n_2 = \frac{n_2}{\gcd\{n_2, n_3\}} n_3,$$

and thus

$$\frac{n_3}{\gcd\{n_2, n_3\}} \bar{n}_2 = \eta\bar{n}_0 + \frac{n_2}{\gcd\{n_2, n_3\}} \bar{n}_3$$

for some positive integer η. Hence

$$c_2' \le \min\left\{d_2, \frac{n_3}{\gcd\{n_2, n_3\}}\right\}.$$

Lemma 19. *Let $a_0, a_1, a_2, a_3 \in \mathbb{N}$. Assume that*

$$\bar{n} = a_0\bar{n}_0 + a_2\bar{n}_2 = a_1\bar{n}_1 + a_3\bar{n}_3 \notin \{c_2'\bar{n}_2, \delta\bar{n}_0 + c_2\bar{n}_2\}$$

yields a nonconnected graph. Then (a_1, a_2, a_3) belongs to

$$C_2 = \left\{ (x_1, x_2, x_3) \in \mathbb{N}^3 \;\middle|\; \begin{array}{c} n_1x_1 - n_2x_2 + n_3x_3 = 0, \\ x_2 < x_1 + x_3 < x_2 + \delta, \\ 0 < x_1 < r_{21},\ c_3 \le x_3, \\ c_2 < x_2 < c_2' \end{array} \right\}.$$

Moreover,

(1) $(a_1, a_3) \in M_2 :=$ Minimals$_\le\{(x_1, x_3) \mid (x_1, x_2, x_3) \in C_2$ *for some* $x_2 \in \mathbb{N}\}$,

(2) $\mathsf{Z}(\bar{n}) = \{(a_0, 0, a_2, 0), (0, a_1, 0, a_3)\}$.

Proof. If $a_0 = 0$, we know by Lemma 9 that the only nonconnected graph $\mathrm{G}_{a_2\bar{n}_2}$ is $\mathrm{G}_{c_2'\bar{n}_2}$. Hence $a_0 \ne 0$.

From

$$a_0\bar{n}_0 + a_2\bar{n}_2 = a_1\bar{n}_1 + a_3\bar{n}_3,$$

we deduce

$$a_0 + a_2 = a_1 + a_3 \quad \text{and} \quad a_2n_2 = a_1n_1 + a_3n_3.$$

The minimality of c_2 yields $a_2 \ge c_2$. If $c_2 = a_2$, then we get $\delta = a_0$, which is not possible by hypothesis. Hence (a_1, a_2, a_3) is a solution of

$$n_1x_1 - n_2x_2 + n_3x_3 = 0, \quad c_2 < x_2 < x_1 + x_3.$$

If $a_1 \ge c_1$, then $a_0\bar{n}_0 + a_2\bar{n}_2 = a_1\bar{n}_1 + a_3\bar{n}_3 = (a_1 - c_1)\bar{n}_1 + r_{12}\bar{n}_2 + (a_3 + r_{13})\bar{n}_3$. If $a_1 > c_1$, we easily derive that $\mathrm{G}_{\bar{n}}$ is connected. If $a_1 = c_1$, then a_3 cannot be zero, since otherwise $c_1n_1 = a_2n_2$, contradicting that $\mathsf{Z}(c_1n_1) = \{(c_1, 0, 0), (r_{12}, 0, r_{13})\}$. Again, the connectedness of $\mathrm{G}_{\bar{n}}$ follows easily. Hence $a_1 < c_1$.

If $a_1 = 0$, then $a_0 + a_2 = a_3$, and this implies that $a_2 \le a_3$. However, we have $a_2n_2 = a_3n_3 > a_3n_2$, which yields $a_2 > a_3$, a contradiction.

Assume that $a_3 < c_3$. As $a_2n_2 = a_1n_1 + a_3n_3$, and σ is a minimal presentation for S, we can deduce that $r_{21} \le a_1$ and $r_{23} \le a_3$. Note that both equalities cannot hold, since $a_2 \ne c_2$. Hence

$$a_0\bar{n}_0 + a_2\bar{n}_2 = a_1\bar{n}_1 + a_3\bar{n}_3 = (a_1 - r_{21})\bar{n}_1 + (a_3 - r_{23})\bar{n}_3 + \delta a_0 + c_2\bar{n}_2,$$

which leads once more to the connectedness of $\mathrm{G}_{\bar{n}}$. This proves that $a_3 \geq c_3$. As $c_3 = r_{13} + r_{23} > r_{23}$, if $a_1 \geq r_{21}$, then we have

$$a_0\bar{n}_0 + a_2\bar{n}_2 = a_1\bar{n}_1 + a_3\bar{n}_3 = (a_1 - r_{21})\bar{n}_1 + (a_3 - r_{23})\bar{n}_3 + \delta\bar{n}_0 + c_2\bar{n}_2,$$

obtaining once more a connected graph. This shows that $a_1 < r_{21}$.

Hence for the rest of the proof we may assume that $a_0a_1a_2a_3 \neq 0$.

We now focus on (2), which will be used later. If

$$(a'_0, a'_1, a'_2, a'_3) \in \mathsf{Z}(\bar{n}) \setminus \{(a_0, 0, a_2, 0), (0, a_1, 0, a_3)\},$$

then as $\mathrm{G}_{\bar{n}}$ is not connected and $a_0a_1a_2a_3 \neq 0$, either $a'_0 = a'_2 = 0$ or $a'_1 = a'_3 = 0$.

- If $a'_0 = a'_2 = 0$, then $a_0\bar{n}_0 + a_2\bar{n}_2 = a_1\bar{n}_1 + a_3\bar{n}_3 = a'_1\bar{n}'_1 + a'_3\bar{n}'_3$. This in particular means that $(a_1 - a'_1)\bar{n}_1 + (a_3 - a'_3)\bar{n}_3 = 0$. Since $\bar{n}_1$ and $\bar{n}_3$ are linearly independent, $a_1 - a'_1 = 0$ and $a_3 - a'_3 = 0$, that is, $a_1 = a'_1$ and $a_3 = a'_3$, a contradiction.
- The case $a'_1 = a'_3 = 0$ follows analogously, since $\bar{n}_0$ and $\bar{n}_2$ are also linearly independent.

Now, if $a_0 \geq \delta$, as $a_2 > c_2$, we get

$$a_0\bar{n}_0 + a_2\bar{n}_2 = (a_0 - \delta)\bar{n}_0 + (a_2 - c_2)\bar{n}_2 + r_{21}\bar{n}_1 + r_{23}\bar{n}_3 = a_1\bar{n}_1 + a_3\bar{n}_3,$$

obtaining again three different factorizations of $\bar{n}$, a contradiction. Hence $a_0 < \delta$. This also implies that $a_1 + a_3 = a_0 + a_2 < \delta + a_2$.

If $a_2 \geq c'_2$, then

$$a_0\bar{n}_0 + a_2\bar{n}_2 = a_1\bar{n}_1 + a_3\bar{n}_3 = (\gamma + a_0)\bar{n}_0 + r'_{21}\bar{n}_1 + (a_2 - c'_2)\bar{n}_2 + r'_{23}\bar{n}_3,$$

which yields three factorizations of $\bar{n}$, in contradiction with (2).

To prove (1), assume there exists $(b_1, b_2, b_3) \in C_2$ such that $(b_1, b_3) \lneq (a_1, a_2)$. Then $a_0\bar{n}_0 + a_2\bar{n}_2 = a_1\bar{n}_1 + a_3\bar{n}_3 = (a_1 - b_1)\bar{n}_1 + (a_3 - b_3)\bar{n}_3 + a_0\bar{n}_0 + a_2\bar{n}_2$. Thus we get three different expressions of $\bar{n}$, a contradiction. □

Lemma 20. *Let $(a_1, a_3) \in M_2$, and let $\bar{n} = a_1\bar{n}_1 + a_3\bar{n}_3$. Then $\mathrm{G}_{\bar{n}}$ is not connected.*

Proof. As $(a_1, a_3) \in M_2$, there exists positive integers a_0 and a_2 such that $\bar{n} = a_0\bar{n}_0 + a_2\bar{n}_2$, $a_0 < \delta$ and $c_2 < a_2 < c'_2$. Assume to the contrary that $\mathrm{G}_{\bar{n}}$ is connected. Then there exists $(b_0, b_1, b_2, b_3) \in \mathsf{Z}(\bar{n}) \setminus \{(a_0, 0, a_2, 0), (0, a_1, 0, a_3)\}$.

From $a_0\bar{n}_0 + a_2\bar{n}_2 = b_0\bar{n}_0 + b_1\bar{n}_1 + b_2\bar{n}_2 + b_3\bar{n}_3$ we deduce the following.

- As $a_2 < c'_2$, we have $b_0 < a_0$, and consequently $b_0 < \delta$.
- Since $a_0 \neq 0$, we have $b_2 < a_2$. We obtain $b_2 < c'_2$.

Now, from $a_1\bar{n}_1 + a_3\bar{n}_3 = b_0\bar{n}_0 + b_1\bar{n}_1 + b_2\bar{n}_2 + b_3\bar{n}_3$ and Lemma 6, we deduce that $a_1 > b_1$. If $a_3 \geq b_3$, then $(a_1 - b_1)\bar{n}_1 + (a_3 - b_3)\bar{n}_3 = b_0\bar{n}_0 + b_2\bar{n}_2$. Notice that

$0 < a_1 - b_1 \le a_1 < r_{21}$, and that $b_2 \ge c_2$ because $b_2 n_2 = (a_1 - b_1)n_1 + (a_3 - b_3)n_3$, and if $b_2 = c_2$ this forces $a_1 - b_1 = r_{21}$, which is impossible. Hence $c_2 < b_2 < c'_2$. Arguing as in the proof of Lemma 19 we get that $c_3 \le a_2 - b_3$. This means that $(a_1 - b_1, b_2, a_3 - b_3) \in C_2$, but this contradicts $(a_1, b_1) \in M_2$.

Thus $a_3 > b_3$ and $(a_1 - b_1)\bar{n}_1 = b_0\bar{n}_0 + b_2\bar{n}_2 + (b_3 - a_3)\bar{n}_3$. But this contradicts the minimality of c_1, because

$$a_1 - b_1 \le a_1 < r_{21} < c_1 \quad \text{and} \quad (a_1 - b_1)n_1 = b_2 n_2 + (b_3 - a_3)n_3. \qquad \square$$

Lemma 21. *Let $a_0, a_1, a_2, a_3 \in \mathbb{N}$. Assume that*

$$\bar{n} = a_0\bar{n}_0 + a_3\bar{n}_3 = a_1\bar{n}_1 + a_2\bar{n}_2 \notin \{c'_2\bar{n}_2,\ \nu\bar{n}_0 + c_3\bar{n}_3\}$$

yields a nonconnected graph. Then (a_1, a_2, a_3) belongs to

$$C_3 = \left\{ (x_1, x_2, x_3) \in \mathbb{N}^3 \;\middle|\; \begin{array}{c} n_1x_1 + n_2x_2 - n_3x_3 = 0, \\ x_3 < x_1 + x_2 < x_3 + \nu, \\ 0 < x_1 < r_{31},\ c_3 < x_3, \\ c_2 \le x_2 < c'_2 \end{array} \right\}.$$

Moreover,

(1) $(a_1, a_2) \in M_3 := \text{Minimals}_{\le}\{(x_1, x_2) \mid (x_1, x_2, x_3) \in C_3$ *for some* $x_3 \in \mathbb{N}\}$,

(2) $\mathsf{Z}(\bar{n}) = \{(a_0, 0, 0, a_3), (0, a_1, a_2, 0)\}$.

Proof. From Lemma 6, we know that $a_0 \ne 0$. Assume that $a_1 = 0$. Then $a_2\bar{n}_2$ is a nonconnected graph, which according to Lemma 9 means that $a_2 = c'_2$, which is excluded in the hypothesis. Hence a_1 is also not zero. The rest of the proof goes as in Lemma 19. $\square$

Lemma 22. *Let $(a_1, a_2) \in M_3$, and let $\bar{n} = a_1\bar{n}_2 + a_2\bar{n}_2$. Then $\mathrm{G}_{\bar{n}}$ is not connected.*

Proof. According to Lemma 21, there exists positive integers a_0 and a_3 such that $\bar{n} = a_0\bar{n}_0 + a_3\bar{n}_3$, $a_0 < \nu$ and $c_3 < a_3$. We argue as in Lemma 20. Assume that there exists an expression $b_0\bar{n}_0 + b_1\bar{n}_1\bar{+}b_2\bar{n}_2 + b_3\bar{n}_3$ other than $a_0\bar{n}_0 + a_3\bar{n}_3$ and $a_1\bar{n}_1 + a_2\bar{n}_2$. Then $a_1\bar{n}_1 + a_2\bar{n}_2 = b_0\bar{n}_0 + b_1\bar{n}_1 + b_2\bar{n}_2 + b_3\bar{n}_3$. From $a_1 < c_1$, we deduce that $a_2 > b_2$, and from $a_2 < c'_2$ that $a_1 > b_1$. Thus

$$0 \ne (a_1 - b_1)\bar{n}_1 + (a_2 - b_2)\bar{n}_2 = b_0\bar{n}_0 + b_3\bar{n}_3.$$

Hence $b_3n_3 = (a_1 - b_1)n_1 + (a_2 - b_2)n_2$, which implies that $b_3 \ge c_3$, and if $c_3 = b_3$ we would get $a_1 - b_1 = r_{31}$, contradicting that $a_1 < r_{31}$. Therefore $b_3 > c_3$. Also $a_1 - b_1 < r_{31}$, and from this it is not difficult to deduce that $a_2 - b_2$ must be greater than or equal to c_2, since otherwise there will be no way by using the relations in σ to get from $(a_1 - b_1, a_2 - b_2, 0)$ to $(0, 0, b_3)$. Gathering all this information, we obtain that $(a_1 - b_1, a_2 - b_2, b_3) \in C_3$ and $(a_1 - b_1, a_2 - b_2) < (a_1, a_2)$, contradicting $(a_1, a_2) \in M_3$. $\square$

Example 23. Let $S = \langle 11, 18, 21\rangle$. A minimal presentation for S is

$$\big\{\big((3,0,1),(0,3,0)\big),\big((6,1,0),(0,0,4)\big),\big((9,0,0),(0,2,3)\big)\big\}.$$

The Betti elements of S are $\{54, 84, 99\}$, while those of $\bar{S}$ are

$$\big\{(4,54),(7,84),(9,99),(7,126),(7,105)\big\}.$$

In this example C_2 is empty, and $C_3 = \{(3,4,5),(3,8,7),(3,25,23)\}$. The minimality condition imposed to the first two coordinates reduces this set to $\{(3,4,5)\}$.

A minimal presentation for $\bar{S}$ is

$$\begin{aligned}\big\{\big((0,3,0,1),(1,0,3,0)\big),\big((0,6,1,0),(3,0,0,4)\big),\big((0,9,0,0),(4,0,2,3)\big),&\\ \big((1,0,0,6),(0,0,7,0)\big),\big((0,3,4,0),(2,0,0,5)\big)\big\}.&\end{aligned}$$

Notice that this semigroup is no longer generic (in all relations all atoms occur), but it is uniquely presented. The set of integers belonging to C_2 and C_3 can be computed by using [Wolfram Alpha 2013] by simply typing in the search field "find integer solutions to" and then the set of inequalities separated by "and."

Theorem 24. *Let S be a nonsymmetric embedding-dimension-three numerical semigroup, with $c_2 < r_{21} + r_{23}$. Then*

$$\begin{aligned}\mathrm{Betti}(\bar{S}) = \{&c_1\bar{n}_1, \delta\bar{n}_0 + c_2\bar{n}_2, c_2'\bar{n}_2, \nu\bar{n}_0 + c_3\bar{n}_3\}\\ &\cup\{a_1\bar{n}_1 + a_3\bar{n}_3 \mid (a_1,a_3)\in M_2\}\cup\{a_1\bar{n}_1 + a_2\bar{n}_2 \mid (a_1,a_2)\in M_3\}.\end{aligned}$$

Moreover, $\bar{S}$ is uniquely presented.

Proof. If $\bar{n} \in \mathrm{Betti}(\bar{S})$, then at least $\mathsf{Z}(\bar{n})$ has two $\mathcal{R}$-classes. Thus in one of them there are at most two atoms of $\bar{S}$, and neither $\bar{n}_0$ nor $\bar{n}_3$ (Lemma 6) are alone. So we have that the set of atoms involved in one of the $\mathcal{R}$-classes is any of these sets: $\{n_0, n_1\}$, $\{n_0, n_2\}$, $\{n_0, n_3\}$, $\{n_1\}$ and $\{n_2\}$. Lemmas 3 to 9, 17, 19, 20, 21 and 22 cover all possibilities. Moreover, in all cases $\#\mathsf{Z}(\bar{n}) = 2$, and thus according to [García-Sánchez and Ojeda 2010, Corollary 5], $\bar{S}$ is uniquely presented. □

Example 25. Recall that a minimal presentation for $S = \langle 10, 17, 19\rangle$ is

$$\big\{\big((4,1,0),(0,0,3)\big),\big((3,0,2),(0,4,0)\big),\big((7,0,0),(0,3,1)\big)\big\}$$

(Example 2). Moreover, $C_2 = \varnothing$ and $C_3 = \{(1,5,5)\}$. Thus the set of Betti elements of $\bar{S}$ is

$$\begin{aligned}\big\{7\bar{n}_1 = (7,70), \bar{n}_0 + 4\bar{n}_2 = (5,68), 2\bar{n}_0 + 3\bar{n}_3 = (5,57),&\\ 9\bar{n}_2 = (9,153), \bar{n}_0 + 5\bar{n}_3 = (6,95)\big\}.&\end{aligned}$$

Example 26. Let $S = \langle 10, 27, 29\rangle$. In view of Example 1 with $k = 1$, a minimal presentation for S is

$$\bigl\{\bigl((6,1,0),(0,0,3)\bigr),\bigl((5,0,2),(0,4,0)\bigr),\bigl((11,0,0),(0,3,1)\bigr)\bigr\}.$$

Here, $C_2 = \{(3, 14, 12), (4, 9, 7)\}$ and $C_3 = \{(1, 5, 5)\}$. Thus

$$\begin{aligned}\mathrm{Betti}(\bar{S}) = \bigl\{11\bar{n}_1 = (11, 110), 3\bar{n}_0 + 4\bar{n}_2 = (7, 108),&\\ 4\bar{n}_0 + 3\bar{n}_3 = (7, 87), 19\bar{n}_2 = (19, 513),&\\ \bar{n}_0 + 14\bar{n}_2 = (15, 378), 2\bar{n}_0 + 9\bar{n}_2 = (11, 243)\bigr\}.&\end{aligned}$$

Remark 27. The uniqueness of the minimal presentation can be derived in a different way. As a consequence of Bresinsky's algorithm the cardinality of $\mathrm{Betti}(\bar{S})$ equals the cardinality of a minimal presentation for $\bar{S}$ (this is also stated in [Li et al. 2012, Lemma 2.2] without using Bresinsky's procedure; there are no two relations in a minimal presentation corresponding to the same element in $\bar{S}$). Thus for every $b \in \mathrm{Betti}(\bar{S})$, $\mathsf{Z}(b)$ has two $\mathcal{R}$-classes. This does not show that the minimal presentation is unique, because some of these $\mathcal{R}$-classes could have more than one element (see, for instance, [Li et al. 2012, Example 2.5]). However it can be shown that in our setting $\pm(b - b') \notin \bar{S}$ for every $b, b' \in \mathrm{Betti}(\bar{S})$, that is to say, all Betti elements of $\bar{S}$ are Betti-minimal. Hence in view of [García-Sánchez and Ojeda 2010, Proposition 3] every $\mathcal{R}$-class of $\mathsf{Z}(b)$ for every $b \in \mathrm{Betti}(S)$ is a singleton (see also [Charalambous et al. 2007, Theorem 3.4]).

2. The Cohen–Macaulay property

We say that an affine semigroup is Cohen–Macaulay if the semigroup ring $k[S]$ is Cohen–Macaulay. The corollary on page 127 of [Bresinsky 1984] gives a characterization of the Cohen–Macaulay property. Also Remark 2.17 in [Li et al. 2012] offers another characterization of the Cohen–Macaulay property. We will use the test proposed in [Rosales et al. 1998] for affine subsemigroups of $\mathbb{N}^2$ to give an alternative proof of Bresinsky's characterization in our scope (S is not symmetric).

Observe that the (rational) cone spanned by $\{\bar{n}_0, \bar{n}_3\}$ equals the cone spanned by $\bar{S}$. Thus a_1 in [Rosales et al. 1998, Section 1] is n_3. Also μ in [Rosales et al. 1998, Lemma 1.1.3] corresponds with $\mu(s) = \min \mathsf{L}(s)$ for every $s \in S$.

Let G be a reduced Gröbner basis of I_S with respect to any total degree ordering and $(a_1, a_2, a_3) \in \mathsf{Z}(s)$ (observe that G consists also of binomial ideals). For a polynomial $f \in k[x_1, x_2, x_3]$, denote by $\mathrm{NF}_G(f)$ the remainder of the division of f by G. It follows that for $s \in S$ and $(a_1, a_2, a_3) \in \mathsf{Z}(s)$, $\mathrm{NF}_G(x_1^{a_1} x_2^{a_2} x_3^{a_3})$ is a monomial, and if

$$\mathrm{NF}_G(x_1^{a_1} x_2^{a_2} x_3^{a_3}) = x_1^{b_1} x_2^{b_2} x_3^{b_3},$$

then $\mu(s) = b_1 + b_2 + b_3$, the total degree of $\mathrm{NF}_G(x_1^{a_1}x_2^{a_2}x_3^{a_3})$.

Proposition 28. *Let S be a nonsymmetric embedding-dimension-three numerical semigroup. Then $\bar{S}$ is Cohen–Macaulay if and only if $c_2 \geq r_{21} + r_{23}$.*

Proof. Notice that if $c_2 \geq r_{21} + r_{23}$, then by Remark 15,

$$G = \left\{x_1^{c_1} - x_2^{r_{12}}x_3^{r_{13}},\, x_2^{x_2} - x_1^{r_{21}}x_3^{r_{23}},\, x_1^{r_{31}}x_2^{r_{32}} - x_3^{c_3}\right\}$$

is a reduced Gröbner basis with respect to any total degree ordering. Let $B = \mathrm{Ap}(\bar{S}, \bar{n}_0) \cap \mathrm{Ap}(\bar{S}, \bar{n}_3)$. We are going to show that $B = \{(\mu(s), s) \mid s \in \mathrm{Ap}(S, n_3)\}$ and thus by [Rosales et al. 1998, Theorem 1.2], $\bar{S}$ is Cohen–Macaulay (in particular the cardinality of B is n_3 and the Cohen–Macaulayness of $\bar{S}$ also follows from [Li et al. 2012, Theorem 1.2]). It is easy to see that if $(n, s) \in \mathrm{Ap}(\bar{S}, \bar{n}_0)$, then $n = \mu(s)$, and thus the inclusion $\{(\mu(s), s) \mid s \in \mathrm{Ap}(S, n_3)\} \subseteq B$ is clear. Now assume that there exists $(\mu(s), s) \in B$ with $s \notin \mathrm{Ap}(S, n_3)$. Then $s = n_3 + t$ for some $t \in S$ and $(\mu(s) - 1, t) \notin \bar{S}$. It is easy to see that this can only occur if and only if $\mu(t) > \mu(s) - 1$. Let $(b_1, b_2, b_3) \in \mathsf{Z}(t)$ be such that $\mathrm{NF}_G(x_1^{b_1}x_2^{b_2}x_3^{b_3}) = x_1^{b_1}x_2^{b_2}x_3^{b_3}$. Hence

$$\mu(t) = b_1 + b_2 + b_3 \quad \text{and} \quad (b_1, b_2, b_3 + 1) \in \mathsf{Z}(s).$$

As $\mu(t) = b_1 + b_2 + b_3 > \mu(s) - 1$, this means that $\mu(s) < b_1 + b_2 + b_3 + 1$, and consequently

$$\mathrm{NF}_G(x_1^{b_1}x_2^{b_2}x_3^{b_3+1}) \neq x_1^{b_1}x_2^{b_2}x_3^{b_3+1}.$$

This implies that either $x_1^{c_1}$ or $x_2^{c_2}$ or $x_1^{r_{31}}x_2^{r_{32}}$ divide $x_1^{b_1}x_2^{b_2}x_3^{b_3+1}$. As x_3 does not occur in $\{x_1^{c_1}, x_2^{c_2}, x_1^{r_{31}}x_2^{r_{32}}\}$, this means that either $x_1^{c_1}$ or $x_2^{c_2}$ or $x_1^{r_{31}}x_2^{r_{32}}$ divide $x_1^{b_1}x_2^{b_2}x_3^{b_3}$, yielding $\mathrm{NF}_G(x_1^{b_1}x_2^{b_2}x_3^{b_3}) \neq x_1^{b_1}x_2^{b_2}x_3^{b_3}$, a contradiction.

If $c_2 < r_{21} + r_{23}$, then $\mu(c_2n_2) = c_2$ (recall that $\mathsf{Z}(c_2n_2) = \{(0, c_2, 0), (r_{21}, 0, r_{23})\}$). Notice that $r_{21}n_1$ has unique expression, and consequently $r_{21}n_1 \in \mathrm{Ap}(S, n_3)$. Hence

$$c_2 = \mu(c_2n_2) = \mu(r_{21}n_1 + r_{23}n_3) \quad \text{and} \quad \mu(r_{21}n_1) + r_{23}\mu(n_3) = r_{21} + r_{23}.$$

Since $c_2 \neq r_{21} + r_{23}$, Proposition 1.6 in [Rosales et al. 1998] states that $\bar{S}$ cannot be Cohen–Macaulay. □

Corollary 29. *Let S be a nonsymmetric embedding-dimension-three numerical semigroup. Then $\bar{S}$ is Cohen–Macaulay if and only if the cardinality of the minimal presentation of S coincides with the cardinality of the minimal presentation of $\bar{S}$.*

3. The catenary degree of $\bar{S}$

Let $S \subset \mathbb{N}^k$ be an affine semigroup. Let $s \in S$, and let

$$a = (a_1, \dots, a_k), b = (b_1, \dots, b_k) \in \mathsf{Z}(s).$$

The *distance* between a and b is $\mathrm{d}(a, b) = \max\{|a - (a \wedge b)|, |b - (a \wedge b)|\}$, where $a \wedge b = (\min(a_1, b_1), \dots, \min(a_k, b_k))$, the common part to the factorizations a and b. For $N \in \mathbb{N}$, an *N-chain* of factorizations joining a and b is a sequence $a_1, \dots, a_t \in \mathsf{Z}(s)$ such that $\mathrm{d}(a_i, a_{i+1}) \leq N$ for all $i \in \{1, \dots, t-1\}$. The *catenary degree* of s, $\mathsf{c}(s)$, is the minimum N such for any $a, b \in \mathsf{Z}(s)$, there exists an N-chain of factorizations joining a and b. The catenary degree of S is defined as

$$\mathsf{c}(S) = \sup_{s \in S} \mathsf{c}(s).$$

As a consequence of [Chapman et al. 2006, Section 3], this supremum is a maximum and indeed

$$\mathsf{c}(S) = \max_{s \in \mathrm{Betti}(S)} \mathsf{c}(s).$$

If S is a numerical semigroup, as $\bar{S}$ is half-factorial, [García-Sánchez et al. 2013, Theorem 2.3] states that for every $s \in \bar{S}$, there exists $b \in \mathrm{Betti}(\bar{S})$ such that $\mathsf{c}(s) = \mathsf{c}(b)$. Hence in our setting we get the following corollary.

Corollary 30. *Let S be a nonsymmetric embedding-dimension-three numerical semigroup and let $s \in \bar{S}$.*

- *If $c_2 \geq r_{21} + r_{23}$, then $\mathsf{c}(s) \in \{c_1, c_2, \nu + c_3\}$.*
- *If $c_2 < r_{21} + r_{23}$, then*

$$\mathsf{c}(s) \in \{c_1, c_2 + \delta, c_2', \nu + c_3\} \cup \{(x + y) \mid (x, y) \in M_2 \cup M_3\}.$$

The catenary degree of $\bar{S}$ corresponds with the homogeneous catenary degree of S ([García-Sánchez et al. 2013, Proposition 3.5]; the concept of homogeneous catenary degree is introduced in that paper). Hence this result gives a description also of the homogeneous catenary degree of S. Also, the homogeneous catenary degree is a lower bound for the monotone catenary degree [García-Sánchez et al. 2013, Proposition 3.9].

Example 31. We apply the above corollary to the semigroups in Example 1. Recall that $S^k = \langle 10, 17 + 10k, 19 + 10k \rangle$ and that the minimal presentation for S is

$$\big\{\big((7 + 4k, 0, 0), (0, 3, 1)\big), \big((0, 4, 0), (3 + 2k, 0, 2)\big), \big((0, 0, 3), (4 + 2k, 1, 0)\big)\big\}.$$

Hence the catenary degree of S is $\mathsf{c}(S) = 7 + 4k$ (the catenary degree of an element with two factorizations with disjoint support is just the maximum of the lengths of these factorizations). The minimal presentation of $\bar{S}$ is

$$\begin{aligned}\big\{\big((0, 7 + 4k, 0, 0), (3 + 4k, 0, 3, 1)\big), \big((1 + 2k, 0, 4, 0), (0, 3 + 2k, 0, 2)\big),& \\ \big((0, 1, 5, 0), (1, 0, 0, 5)\big)\big\}& \\ \cup \big\{\big((2k + 1 - i, 0, 5i + 4, 0), (0, 3 + 2k - i, 0, 5i + 2)\big) \mid i \in \{0, \dots, 2k + 1\}\big\}&.\end{aligned}$$

Hence $\mathsf{c}(\bar{S}) = 9 + 10k$.

4. The nonsymmetric case

If S is not symmetric, then we know (see, for instance, [Rosales and García-Sánchez 2009, Example 8.23]) that some of the following cases can occur (these also include the possibility that $\{n_1, n_2, n_3\}$ is not a minimal generating system, that is, some of the c_i are equal to one):

(1) $c_1n_1 = c_2n_2 = c_3n_3$,

(2) $c_1n_1 = r_{12}n_2 + r_{13}n_3 \neq c_2n_2 = c_3n_3$ $(r_{12}r_{13} \neq 0)$,

(3) $c_1n_1 = c_2n_2 \neq c_3n_3 = r_{31}n_1 + r_{32}n_2$ $(r_{31}r_{32} \neq 0)$,

(4) $c_1n_1 = c_3n_3 \neq c_2n_2 = r_{21}n_1 + r_{23}n_3$ $(r_{21}r_{23} \neq 0)$ and $c_2 \geq r_{21} + r_{23}$,

(5) $c_1n_1 = c_3n_3 \neq c_2n_2 = r_{21}n_1 + r_{23}n_3$ $(r_{21}r_{23} \neq 0)$ and $c_2 < r_{21} + r_{23}$.

For the cases (1), (2) and (4), Bresinsky's algorithm stops in the first step, and thus both $\bar{S}$ and S have a minimal presentation with two elements.

For (3) and (5), the discussion follows as in the similar case in the nonsymmetric setting.

Observe that the uniqueness of a minimal presentation for $\bar{S}$ is not ensured since S might have more than two minimal presentations.

References

[Bresinsky 1984] H. Bresinsky, "Minimal free resolutions of monomial curves in $\mathbf{P}_k^3$", *Linear Algebra Appl.* **59** (1984), 121–129. MR 85d:14042 Zbl 0542.14022

[Chapman et al. 2006] S. T. Chapman, P. A. García-Sánchez, D. Llena, V. Ponomarenko, and J. C. Rosales, "The catenary and tame degree in finitely generated commutative cancellative monoids", *Manuscripta Math.* **120**:3 (2006), 253–264. MR 2007d:20106 Zbl 1117.20045

[Charalambous et al. 2007] H. Charalambous, A. Katsabekis, and A. Thoma, "Minimal systems of binomial generators and the indispensable complex of a toric ideal", *Proc. Amer. Math. Soc.* **135**:11 (2007), 3443–3451. MR 2009a:13033 Zbl 1127.13018

[Cox et al. 2007] D. Cox, J. Little, and D. O'Shea, *Ideals, varieties, and algorithms: an introduction to computational algebraic geometry and commutative algebra*, 3rd ed., Springer, New York, 2007. MR 2007h:13036

[Delgado et al. 2013] M. Delgado, P. García-Sánchez, and J. Morais, "Numericalsgps: a gap package on numerical semigroups", website, 2013, http://tinyurl.com/numericalsgps.

[García-Sánchez and Ojeda 2010] P. A. García-Sánchez and I. Ojeda, "Uniquely presented finitely generated commutative monoids", *Pacific J. Math.* **248**:1 (2010), 91–105. MR 2011j:20139 Zbl 1208.20052

[García-Sánchez et al. 2013] P. A. García-Sánchez, I. Ojeda, and A. Sánchez-R.-Navarro, "Factorization invariants in half-factorial affine semigroups", *Internat. J. Algebra Comput.* **23**:1 (2013), 111–122. MR 3040805 Zbl 06156066

[Herzog 1970] J. Herzog, "Generators and relations of abelian semigroups and semigroup rings", *Manuscripta Math.* **3** (1970), 175–193. MR 42 #4657 Zbl 0211.33801

[Li et al. 2012] P. Li, D. P. Patil, and L. G. Roberts, "Bases and ideal generators for projective monomial curves", *Comm. Algebra* **40**:1 (2012), 173–191. MR 2876297 Zbl 1238.14020

[Rosales and García-Sánchez 1999] J. C. Rosales and P. A. García-Sánchez, *Finitely generated commutative monoids*, Nova Science Publishers, Commack, NY, 1999. MR 2000d:20074 Zbl 0966.20028

[Rosales and García-Sánchez 2009] J. C. Rosales and P. A. García-Sánchez, *Numerical semigroups*, Developments in Mathematics **20**, Springer, New York, 2009. MR 2010j:20091 Zbl 1220.20047

[Rosales et al. 1998] J. C. Rosales, P. A. García-Sánchez, and J. M. Urbano-Blanco, "On Cohen–Macaulay subsemigroups of $\mathbb{N}^2$", *Comm. Algebra* **26**:8 (1998), 2543–2558. MR 99g:13032 Zbl 0910.20042

[Wolfram Alpha 2013] Wolfram Alpha, website, 2013, http://www.wolframalpha.com.

Received: 2013-02-25 Revised: 2013-05-02 Accepted: 2013-06-01

yumna2009@yahoo.com *Departamento de Álgebra, Facultad de Ciencias, Universidad de Granada, Av. Fuentenueva, s/n, 18071 Granada, Spain*

pedro@ugr.es *Departamento de Álgebra, Facultad de Ciencias, Universidad de Granada, Av. Fuentenueva, s/n, 18071 Granada, Spain*

INVOLVE 7:1 (2014)
dx.doi.org/10.2140/involve.2014.7.97

Effective resistance on graphs and the epidemic quasimetric

Josh Ericson, Pietro Poggi-Corradini and Hainan Zhang

(Communicated by Gaven Martin)

We introduce the epidemic quasimetric on graphs and study its behavior with respect to clustering techniques. In particular we compare its behavior to known objects such as the graph distance, effective resistance, and modulus of curve families.

1. Introduction

This study was initiated by the need to analyze real world data collected in the rural town of Chanute, Kansas.[1] The goal was to study and simulate potential epidemic outbreaks. From the survey, a contact network was constructed representing the sampled population and their potential relationships. Mathematically, this is just a graph where the vertices represent people and the edges represent possible interactions. In this paper, we introduce a new geometric quantity, the epidemic quasimetric, which we study and relate to more classical quantities, such as effective resistance.

One of the simplest geometric object that is used to study finite graphs is the "graph metric". Namely, the graph metric measures the distance between two nodes a and b by computing the minimal number of edges that must be traversed ("hops") to go from a to b.

Epidemics, on the other hand, can be modeled to begin at one node, then spread to all the neighbors, and then to all the neighbors' neighbors, etc. The possible

MSC2010: 05C12, 05C50, 31A15, 91C20, 05C81.
Keywords: effective resistance, finite graphs, modulus of curve families.
Ericson and Zhang completed this project while undergraduate students at Kansas State University. Poggi-Corradini thanks the Center for Engagement and Community Development at Kansas State University, the Department of Electrical and Computer Engineering at Kansas State University, and the I-center in the Department of Mathematics at Kansas State University, for their generous support of this research. This project was supported in part by NSF grant 1201427.

[1] This was a joint project of the first author with Professor C. Scoglio in the Department of Electrical and Computing Engineering and Professor W. Schumm in the Department of Family Studies at Kansas State University.

damage of the epidemic spreads as a circular wave. We use this dynamic to assign to every pair of nodes of a finite graph a number, which we call the *epidemic quasimetric*. To compute the epidemic quasimetric between a and b we expand the range of an epidemic started at a until b is affected and compute the number of edges that became involved in the process. Then we do the same interchanging a and b, and we add the two numbers thus obtained. This is fairly easy to compute numerically and we describe the routine we implemented in Matlab in Section 3.

Part of the inspiration for considering the epidemic quasimetric came from reading [Semmes 1993][2] where a similar quantity is introduced in order to study bi-Lipschitz embeddings of metric spaces. The hope is that the epidemic quasimetric contains geometric information that allows to view the graph under a new light. To partially confirm this intuition, we experimented with the epidemic quasimetric and showed how it can be used to obtain a pretty accurate cut of a classical example graph into "natural" communities; see page 122. In this direction, aside for clustering techniques, the epidemic quasimetric could be also be useful in sparsification techniques.

Our second goal is to compare the notion of epidemic quasimetric to the more classical notion of effective conductance, when the graph is viewed as an electrical network. Effective conductance has also been used in the literature to study graphs from the point of view of community detection and sparsification. Therefore, such a comparison gives us hope that the epidemic quasimetric can also be used effectively to study graphs while being relatively simple to compute numerically.

The paper begins with some preliminaries and notations about graphs; then in Section 3 we define the epidemic quasimetric and state our goal to relate it to effective resistance. Thereafter we review the theory of random walks on finite graphs in Section 4, its connection to electrical networks, and the notion of effective conductance in Section 5. Some references for these sections are [Doyle and Snell 1984; Levin et al. 2009; Grimmett 2010].

Then, in Section 6, we introduce two more concepts drawn from modern geometric function theory — namely, the notions of capacity and modulus of families of curves; see [Ahlfors 1973]. In Theorem 6.3 we show that all of these concepts coincide with the notion of effective conductance (the method of Lagrange Multipliers turned out to be useful in this context). Moreover, in Proposition 6.4 we exploit the definition of modulus to obtain a comparison between modulus and epidemic quasimetric. Thus we get an estimate for the epidemic quasimetric in terms of effective conductance.

Finally, in Section 7 we describe our numerical computations and experiments.

We begin with some preliminaries on elementary graph theory.

[2] Diego Maldonado pointed out that a similar concept had been introduced earlier in [Macías and Segovia 1979].

2. Graphs

Notation and generalities. We will restrict our study to simple, finite, connected graphs. Let $G = (V, E)$ be a *graph* with vertex-set V and edge-set E. We say that G is *simple* if there is at most one undirected edge between any two distinct vertices, and it is *finite* if the vertex set has cardinality $|V| = N \in \mathbb{N}$. In this case, the edge-set E can be thought of as a subset of $\binom{V}{2}$, the set of all unordered pairs from V. Therefore the cardinality of E is $M = |E| \le \binom{N}{2} = N(N-1)/2$.

We say that two vertices x, y are *neighbors* and write $x \sim y$ if $\{x, y\} \in E$. The graph is *connected* if for any two vertices $a, b \in V$ there is a chain of vertices $x_0 = a, x_1, \dots, x_n = b$, so that $x_j \sim x_{j+1}$ for $j = 0, \dots, n-1$. It is known that connected graphs must satisfy $|E| \ge N - 1$ (induction).

Given a subset of vertices $V' \subset V$, we let $E(V') \subset E$ be all the edges of G that connect pairs of vertices in V'. With this notation $G(V') = (V', E(V'))$ is a simple graph which we call the *subgraph induced by V'*. More generally, a *subgraph* of G is a graph $G' = (V', E')$ such that $V' \subset V$ and $E' \subset E(V')$.

The number of edges that are incident at a vertex x is called the *degree* of x and we write $d(x)$. Since every edge is incident at two distinct vertices, it contributes to two degrees. Therefore

$$\sum_{x \in V} d(x) = 2|E|.$$

This identity is sometimes referred to as the *handshake lemma*. It says that instead of counting edges, one can add degrees, that is, switch to $d(x)$ which is a function defined on V.

For instance, the *volume* of a subgraph $H = (V(H), E(H))$ of G can be defined as $|E(H)|$, the number of edges of H; or as half the sum of the H-degrees over the vertices of H:

$$\frac{1}{2} \sum_{x \in V(H)} d_H(x).$$

We say that γ is a *curve* in G if γ is a connected subgraph of G. This is not a very common way of defining curves in graph theory, but it makes sense from the point of view of the function-theory inspired concepts that we will introduce later when we talk about modulus of curve families.

The graph G is *weighted* if there is a weight function $W : E \to [0, +\infty)$ defined on the edges. The unweighted graph is recovered by setting $W_0(e) = 1$, for all $e \in E$.

Given a curve γ in G, it is natural to define its *graph-length* to be the total number of edges in γ:

$$\operatorname{length}_G(\gamma) := \sum_{e \in E(\gamma)} W_0(e). \tag{2-1}$$

The graph distance. A function of two variables $d(x,y) \geq 0$ on a space X is called a *metric* or a *distance*, if it is symmetric, $d(x,y) = d(y,x)$; nondegenerate, $d(x,y) = 0$ if and only if $x = y$; and satisfies the *triangle inequality*, $d(x,y) \leq d(x,z) + d(z,y)$, whenever $x, y, z \in X$.

On a connected graph G the *graph distance* $d_G(x,y)$ is defined as the shortest graph-length of a curve connecting x to y.

$$d_G(x,y) := \min_{\gamma : x \rightsquigarrow y} \operatorname{length}_G(\gamma), \tag{2-2}$$

where $\gamma : x \rightsquigarrow y$ means that γ is a curve connecting x to y. We leave the verification that d_G is a metric to the reader.

The *diameter* of the graph G is

$$\operatorname{diam}(G) = \max_{x,y \in V} d_G(x,y).$$

The *metric ball* centered at a vertex x and of radius r is

$$\mathcal{B}(x,n) := \{y \in V : d_G(x,y) \leq r\}.$$

3. The epidemic quasimetric

Given two nodes x and y, we define the epidemic quasimetric between them to be the size of the part of the graph that would be affected (the potential damage) if an epidemic started at x and reached y, or vice versa. In formulas, we consider all the vertices in

$$\mathcal{B}(x, d_G(x,y)),$$

and form the induced subgraph of G which we call $\Omega(x, d_G(x,y))$. We then compute the volume $|\Omega(x, d_G(x,y))|$ as on page 99, by counting the number of edges. So we define the *epidemic quasimetric* between x and y to be

$$\operatorname{Epidemic}(x,y) := |\Omega(x, d_G(x,y))| + |\Omega(y, d_G(x,y))|. \tag{3-1}$$

The epidemic quasimetric is not a distance in the mathematical sense. For instance, the triangle inequality can fail as badly as possible, as the following example shows.

Example 3.1. In Figure 1,

$$\operatorname{Epidemic}(x_1,x_2) = 3 \quad \text{and} \quad \operatorname{Epidemic}(x_2,x_3) = 4,$$

while

$$\operatorname{Epidemic}(x_1,x_3) = N + 5.$$

So the triangle inequality in this case can be made to fail as badly as needed by letting N increase.

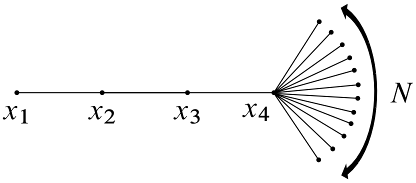


Figure 1. Failure of the triangle inequality.

As described in Section 7, the epidemic quasimetric seems to carry useful geometric information about a graph and in our experiments it appears to behave well with respect to simple clustering algorithms. For the rest of the paper, our intent is to present a comprehensive survey of effective resistance and to answer the following question.

Question 3.2. Is there a connection between the epidemic quasimetric and effective resistance?

We begin in the next two sections by surveying Markov chains and electrical networks. Most of this material can be found in [Levin et al. 2009; Grimmett 2010; Doyle and Snell 1984].

4. Markov chains

A Markov chain is comprised of a finite set S (the state space) and a probability distribution on the state space which can be represented in terms of a *transition matrix* $\{P(i,j)\} = P$ whose entries correspond to the probability of being at state j at time $n+1$ given that you were at state i at time n. P being a transition matrix means that we must have $\sum_j P(i,j) = 1$ (row sums add up to 1.)

This determines a process, that is, a sequence of random variables $X_n = \Omega \to S$, with the property that $\mathbb{P}(X_{n+1} = y \mid X_n = x) = p(x,y)$. Note that $P(x,y)$ is independent of n.

Example 4.1. A random walk on a graph is an example of a Markov chain. Consider a finite, simple graph $G = (V, E)$ with vertices $V = \{1, 2, \dots, N\}$ and edges E. For $x \in V$, d_x indicates the degree or local index of x. We've seen that, by the handshake lemma, $\sum_{x \in V} d_x = 2|E|$.

Define

$$P(x,y) := \begin{cases} 1/d_x & x \sim y, \\ 0 & \text{else.} \end{cases} \tag{4-1}$$

Note that $\sum_{y \in V} P(x,y) = \sum_{y \sim x} 1/d_x = (1/d_x) \sum_{y \sim x} 1 = 1$.

Matrices act on column vectors. Given N states, $\{P(x,y)\} = P$ is an $N \times N$ matrix. Here $N \times 1$ vectors correspond to functions $f : S \to \mathbb{R}$ on the state space.

Then P acts on functions as follows:

$$g(x) = (Pf)(x) = \sum_{y \in S} P(x, y) f(y).$$

Probabilistically,

$$g(x) = \sum_{y \in S} f(y)\mathbb{P}(X_1 = y \mid X_0 = x) = \mathbb{E}_{X_0=x}(f(X_1)),$$

which is the average or expected value of f evaluated on the process at time 1. On a graph:

$$g(x) = \frac{1}{d_x} \sum_{y \sim x} f(y).$$

So g is obtained from f by defining $g(x)$ to be the average of the values of f over all the neighbors of node x.

Definition 4.2. Whenever $x \in V$ and $(Pf)(x) = f(x)$ we say that f is a *harmonic* function on the graph at x.

Example 4.3 (gambler's ruin). Consider six nodes $\{0, 1, 2, 3, 4, 5\}$ representing the dollar amounts held by a gambler. After each bet the gambler either wins or loses a dollar with equal probability. The gambler will walk away whenever his fortune is 0 (ruin), or 5 (predetermined goal). We represent this with a transition matrix P, which is a 6×6 matrix in this case.

We would like to know the probability of reaching 5 before 0, assuming that we start our random walk at node 3. This is an example of a *hitting probability*, what in complex analysis would be called a harmonic measure problem.

Let

$$h(x) = \mathbb{P}(X_n \text{ hits 5 before 0} \mid X_0 = x) = \mathbb{P}_x(X_n \text{ hits 5 before 0}),$$

where $\mathbb{P}_x$ is probability conditioned on $\{X_0 = x\}$.

We want to compute $h(3)$. We call the points in $B = \{0, 5\}$ *boundary points* and those in $I = \{1, 2, 3, 4\}$ *interior points*.

For $x \in I$, we can condition on the first step; this is known as "first-step analysis". For brevity we write $A := \{X_n \text{ hits 5 before 0}\}$:

$$\begin{aligned} h(x) &= \mathbb{P}_x(A) \\ &= \mathbb{P}_x(A \mid X_1 = x-1)\mathbb{P}_x(X_1 = x-1) + \mathbb{P}_x(A \mid X_1 = x+1)\mathbb{P}_x(X_1 = x+1) \\ &= \tfrac{1}{2}h(x-1) + \tfrac{1}{2}h(x+1), \end{aligned}$$

since $P(x, y) = \frac{1}{2}$ if $y = x - 1$ or $x + 1$, and 0 otherwise. The function h is thus harmonic for P at each interior point. Also h has boundary values $h(0) = 0$ and $h(5) = 1$.

Fact 4.4 (maximum principle). *A harmonic function achieves its maximum value M and minimum value m on the boundary.*

Idea: suppose f attains its maximum at an interior point. By harmonicity it must attain it at each neighboring vertex too and this will propagate out as an oil spill all the way to the boundary.

Claim 4.5. *The problem of finding a harmonic function h on I with boundary values $h(0) = 0$ and $h(5) = 1$ has a unique solution.*

Proof. Suppose that h and g are both harmonic at each interior point and that $h(0) = g(0) = 0$, $h(5) = g(5) = 1$. Let $f = h - g$. Then, by linearity, we have that

$$f(x) = \tfrac{1}{2} f(x-1) + \tfrac{1}{2} f(x+1),$$

with $f(0) = 0$ and $f(5) = 0$. We get that $M = \max(f) = 0$ and $m = \min(f) = 0$ and the maximum principle implies that $f \equiv 0$ or $g = h$. □

Remark 4.6. The gambler's ruin example is part of a larger set of problems. Given a subset B of nodes that we will call "boundary" and a function h_B defined only on B, it is always possible to extend this function on the remaining nodes (the *interior points*) so that the extension h is harmonic on the interior points. Claim 4.5, shows that if a solution exists it is unique. However, the existence is obtained by writing down a solution explicitly. For this we must introduce the notion of *stopping times*. Given a walker starting at some interior node x the stopping time τ_B is the first time the walker visits a node in B. Since the walk is a random process, stopping times are random variables. The solution to the boundary-value (or Dirichlet) problem is

$$h(x) := \mathbb{E}_x(h_B(X_{\tau_B})),$$

namely, the expected value of h_B evaluated at the exit point of a walk started at x.

Matrices act on row vectors. On a finite graph a random walker either runs until it hits a given set of boundary points, as in the gambler's ruin example, or it bounces around forever. In the latter case, we can ask what fraction of time does it spend at a node x?

Intuitively we want to define π to be the *stable distribution* if $\sum_{x \in V} \pi(x) = 1$ and $\pi(x)$ denotes the probability of finding the random walker at x in the long run. This however already implies that the long run stabilizes.

Let $\mu_0(x)$ denote the initial distribution for X_0 and similarly $\mu_1(x)$ the distribution of X_1, after one "step". Then, conditioning on the previous location we get:

$$\mu_1(x) = \sum_y \mu_0(y)\mathbb{P}(X_1 = x \mid X_0 = y) = \sum_y \mu_0(y)P(y, x).$$

If we think of $\mu_0(x)$ as a row vector then we can rewrite this in matrix multiplication form as $\mu_1 = \mu_0 P$, $\mu_2 = \mu_1 P = \mu_0 P^2$, and so on. The entries of P^2 will be of the form

$$(P^2)(x, y) = \sum_z P(x, z)P(z, y) = \mathbb{P}(X_2 = y \mid X_0 = x).$$

In general, we have $\mu_n = \mu_0 P^n$, where P^n is the n-th power of the matrix P.

If, as n goes to infinity, $\lim_{n\to\infty} \mu_n = \pi$, then π is a unique fixed-point for P. That's because for large n,

$$\pi \approx \mu_{n+1} = \mu_n P \approx \pi P.$$

We say that π is a *stable distribution* for P if $\pi P = \pi$.

For a Markov chain on a finite state space there are very mild conditions (irreducibility and aperiodicity) that guarantee the existence and uniqueness of a stable distribution π, as well as the convergence $\lim_{n\to\infty} \mu_0 P^n = \pi$ independently of μ_0. In such cases, $\pi(x) > 0$ at every $x \in S$ and $1/\pi(x)$ equals the expected return time to x (in formulas $\mathbb{E}_x(\tau_x^+)$). Namely, π is inversely proportional to the average amount of time it takes for the random walker to find its way back to the starting node. The easier it is to come back, the larger π is, and the larger the proportion of time the random walker spends at x in the long run.

Reversible chains. The chains we will use in the sequel will have one extra property: reversibility.

Definition 4.7. The Markov chain P is *reversible* if there is a distribution π such that

$$\pi(x)P(x, y) = \pi(y)P(y, x) \quad \text{for all } x, y \in V. \tag{4-2}$$

If P is reversible, then the distribution π is stable, as the following shows:

$$\sum_x \pi(x)P(x, y) = \sum_x \pi(y)P(y, x) = \pi(y)\sum_x P(y, x) = \pi(y).$$

The reversibility condition is very closely related to the notion of "symmetric" or "self-adjoint" matrices. In fact, if we let $A(x, y) = \pi(x)P(x, y)$, then (4-2) is $A(x, y) = A(y, x)$. In matrix notation $A = DP$, where D is a diagonal matrix with entries $\pi(x)$. So standard results in linear algebra tell us that A and thus P, are linearizable to a diagonal matrix of real eigenvalues. Moreover, the action of P on row vectors is the action of the adjoint (or transpose) P^T on column vectors and

P^T shares the same eigenvalues as P and A. The difference lies in the eigenvectors (or eigenfunctions), but there are simple formulas involving π to go from one set to the other.

The fact that the rows of P sum to 1 implies that the eigenvalues of P are between -1 and 1. To see this note that

$$|Pf(x)| = \left|\sum_y P(x,y)f(y)\right| \leq (\max_y |f(y)|) \sum_y P(x,y) = \max_y |f(y)|.$$

Suppose that $Pf = \lambda f$. Then, $|\lambda|\,|f(x)| \leq \max_y |f(y)|$ for all x's. Hence, $|\lambda| \leq 1$.

Finally, the fact that the rows of P sum to one also imply straightforwardly that 1 is an eigenvalue of P for constant eigenvectors.

Example 4.8 (random walks on weighted graphs). Let $G=(V,E;W)$ be a weighted graph. The degree of a vertex is usually the number of neighbors of x, but for a weighted graph it can also be defined as

$$\deg x := \sum_{y \backsim x} W(x,y).$$

Then a (weighted) random walk on G is defined to be the Markov chain with state space V and transition matrix

$$P(x,y) = \begin{cases} W(x,y)/\deg x & \text{if } y \backsim x, \\ 0 & \text{otherwise.} \end{cases}$$

In this situation the probability distribution defined at each $x \in V$ by

$$\pi(x) := \frac{\deg x}{Z}, \quad \text{where } Z = \sum_{z \in V} \deg z,$$

will satisfy

$$\pi(x)P(x,y) = \frac{\deg x}{Z}\frac{W(x,y)}{\deg x} = \frac{\deg y}{Z}\frac{W(y,x)}{\deg y} = \pi(y)P(y,x).$$

In particular, P is reversible in this case.

5. Electrical networks and effective resistance

It is often useful when studying graphs and random walks on graphs to think of them as electrical networks. Each edge e then acquires a weight $W(e) = C(e)$ which in this instance plays the role of conductance. One recalls from high-school physics that the electrical current I through a connection is related to the potential difference V via the formulas

$$V = RI \quad \text{and} \quad I = CV,$$

where $R = 1/C$ is the resistance and C the conductance.

If instead of a simple connection one is looking at a series of connections then simple rules allow one to say that the system behaves as if there was only one connection but with an appropriately modified resistance. This "virtual" resistance is what we call the effective resistance between two nodes (see next page for definition).

Flows and currents. In order to make this more precise we need to introduce some notation. Given two nodes a and b we consider a unit current that is allowed to pass through a connected graph with the "source" at $a \in V$ and the "sink" at $b \in V$. In order to effect this passage a certain voltage difference must be applied at a and b, and a corresponding voltage potential will arise at every node x in the graph. Let $v(x)$ denote the voltage at $x \in V$.

The electric current that runs through the graph is an example of a flow. Mathematically, *a flow* is an assignment of a number j_{xy} (representing its intensity) to every directed edge (x, y), that has the property of being *antisymmetric*: $j(x, y) = -j(y, x)$. Given two nodes a and b, we say j is *a flow from a to b* if the following property is satisfied at every node $x \neq a, b$.

Kirchoff's node law. *The flow into a vertex x is equal to the flow out of x. In other words, the divergence* $\operatorname{div}_j(x) := \sum_{y \sim x} j(x, y)$ *is equal to zero at every $x \neq a, b$.*

Moreover, we also demand that $\operatorname{div}_j(a) \geq 0$ (and thus $\operatorname{div}_j(b) \leq 0$), since it's a flow from a to b. The quantity $\operatorname{div}_j(a)$ is the *strength* of the flow j from a to b.

To compute effective resistance we will assume that a unit flow is entering the network at a and exiting at b, so $\operatorname{div}_j(a) = -\operatorname{div}_j(b) = 1$ and $\operatorname{div}_j(x) = 0$ for $x \neq a, b$.

A flow from a to b is furthermore a *current flow*, if it also satisfies Ohm's law below. In this case we use the notation $i(x, y)$.

Ohm's law. *There is a potential function v defined on the nodes such that for every oriented edge (x, y), $R(x, y)i(x, y) = v(y) - v(x)$ where $R(x, y)$ is the resistance of the edge (x, y).*

Whenever a flow can be expressed as the edge difference of a potential function defined on the nodes, as is the case for current, then around any cycle

$$x_1 \sim \cdots \sim x_n \sim x_{n+1} = x_1,$$

the following holds necessarily:

$$\sum_{k=1}^{n} R(x_k, x_{k+1}) i(x_k, x_{k+1}) = 0.$$

This is known as *Kirchoff's potential law* and it turns out to be equivalent to the existence of a potential.

Combining Kirchoff's node law and Ohm's law at $x \neq a, b$, we get

$$0 = \sum_{y \sim x} i(x, y) = \sum_{y \sim x} \frac{v(y) - v(x)}{R(x, y)} = \sum_{y \sim x} C(x, y)v(y) - C(x)v(x),$$

where $C(x, y) = 1/R(x, y)$ is the conductance of the edge (x, y) and

$$C(x) := \sum_{y \sim x} C(x, y) \tag{5-1}$$

is the *local conductance* at node x. In particular, we have

$$v(x) = \frac{1}{C(x)} \sum_{y \sim x} C(x, y)v(y);$$

that is, the potential voltage function v must be harmonic with respect to the graph weighted by local conductances.

That a current flow satisfying all of these laws and requirements exists is a physical fact, however mathematically one has to prove its existence and uniqueness.

Proving uniqueness is easy because flows that satisfy the two Kirchoff laws mentioned above satisfy the superposition principle: given two such flows i_1 and i_2, then $j = i_1 - i_2$ will also satisfy the same laws.

The existence proof is more challenging and its resolution surprising. A flow can actually be constructed explicitly that satisfies all the requirements using the concept of spanning trees. See [Grimmett 2010, Theorem 1.16].

Probabilistically, the resulting current $i(x, y)$ along an edge (x, y) can be shown to equal the expected number of (net) times that the random walker (on the graph weighted by the conductances) crosses from x to y.

Effective resistance. Given two nodes a and b consider a unit current flow entering at a and exiting the network at b. This flow exists and is unique, as explained above, once the Neumann conditions, namely the entering and exiting flow, is fixed at a and b. Moreover, this current determines a unique voltage potential v at each node (up to a constant). The absolute value of the voltage drop between a and b is what we call the *effective resistance between a and b*. In formulas:

$$\mathcal{R}_{\text{eff}}(a, b) := |v(a) - v(b)| = v(b) - v(a). \tag{5-2}$$

Effective resistance and escape probabilities. Recall the definition of local conductance in (5-1). The random walk on the weighted graph is the Markov chain with transition probabilities

$$P(x,y) := \frac{C(x,y)}{C(x)}, \quad \text{whenever } y \sim x. \tag{5-3}$$

Effective resistance, or more appropriately, *effective conductance*,

$$\mathcal{C}_{\text{eff}}(a,b) := \frac{1}{\mathcal{R}_{\text{eff}}(a,b)} \tag{5-4}$$

is related to the probability that a random walker starting at a visits b before returning to a. In symbols, recall the notion of stopping time. Assuming that the random walk is at a at time 0, we write τ_z for the first time it visits node z and τ_a^+ for the first time the walker revisits a after time 1. We are interested in $\mathbb{P}_a(\tau_z < \tau_a^+)$, the probability that the random walk starting at a visits z before returning to a.

Proposition 5.1. *For any $a, b \in V$,*

$$C(a)\mathbb{P}_a(\tau_b < \tau_a^+) = \mathcal{C}_{\text{eff}}(a,b).$$

Proof. The proof is a beautiful application of the maximum principle for harmonic functions (Fact 4.4). Consider $B = \{a,b\}$ to be the boundary and $\Omega = V \setminus \{a,b\}$ the interior. We want to find a harmonic function on Ω which takes the value 0 at a and the value 1 at b. By the maximum principle, the solution to this problem is unique. We now produce two such solutions.

The first is

$$h(x) := \mathbb{P}_x(\tau_b < \tau_a^+).$$

Harmonicity can be checked by conditioning on the first step.

The second solution is given by normalizing the voltage function $v(x)$ which is required in order to have one unit of current flow in at a and out at b. So let

$$g(x) := \frac{v(x)-v(a)}{v(b)-v(a)} = \mathcal{C}_{\text{eff}}(a,b)(v(x)-v(a)).$$

By uniqueness, $h = g$.

Now, by conditional probability,

$$\begin{aligned}
\mathbb{P}_a(\tau_b < \tau_a^+) &= \sum_{x \in V} P(a,x))\mathbb{P}_x(\tau_b < \tau_a^+) && \text{(first-step analysis)}\\
&= \sum_{x \sim a} \frac{C(a,x)}{C(a)} \mathcal{C}_{\text{eff}}(a,b)(v(x)-v(a)) && \text{(by (5-3))}\\
&= \frac{\mathcal{C}_{\text{eff}}(a,b)}{C(a)} \sum_{x \sim a} i(a,x) && \text{(Ohm's law)}\\
&= \frac{\mathcal{C}_{\text{eff}}(a,b)}{C(a)} \operatorname{div}_i(a) = \frac{\mathcal{C}_{\text{eff}}(a,b)}{C(a)} && \text{(by Kirchoff's node law)} \quad \square
\end{aligned}$$

Remark 5.2. The probability $p = \mathbb{P}_a(\tau_b < \tau_a^+)$ is also known as the *escape probability*. With probability $1-p$ escape fails and the walker returns to a before visiting b, at which point another identical walker starts out for another attempt. This is akin to flipping a biased coin that has probability p of success (heads). In particular, the number of tosses N required to achieve success is known to be distributed with the geometric distribution:

$$\mathbb{P}(N = k) = p(1-p)^{k-1} \quad \text{for } k = 1, 2, 3, \dots.$$

A calculation using the geometric series shows that the expected number of tosses $\mathbb{E}(N)$ equals $1/p$. In our context N is the *number of visits to a before escaping through b* (where we count $t = 0$ as a visit) and its expectation is known as the *Green's function* of the random walk started at a and stopped at b. More generally, $G_b(a, c)$ is the expected number of visits to c for the walk started at a, before it is stopped at b. It follows from this discussion and Proposition 5.1 that $G_b(a, a) = C(a)\mathcal{R}_{\text{eff}}(a, b)$.

Rayleigh's monotonicity and energy. The *energy* of a flow j is

$$\text{Energy}(j) := \sum_{e \in E} R(e)(j(e))^2.$$

Notice that for each edge $e \in E$, the quantity $j(e)$ is defined up to a sign change. Therefore, the square $(j(e))^2$ is well defined.

Proposition 5.3. *Consider the unit current flow i from a to b and its corresponding potential v defined on the nodes. Given an arbitrary flow k from a to b,*

$$\sum_{e \in E} R(e)i(e)k(e) = (v(b) - v(a)) \operatorname{div}_k(a).$$

Proof. We have

$$\sum_{e \in E} R(e)i(e)k(e) = \frac{1}{2} \sum_{\substack{x \sim y \\ x, y \in V}} R(x, y)i(x, y)k(x, y) = \frac{1}{2} \sum_{\substack{x \sim y \\ x, y \in V}} (v(y) - v(x))k(x, y),$$

by Ohm's law. Since $k(x, y) = 0$ if $x \not\sim y$, this expression can further be transformed into

$$\frac{1}{2} \sum_{y \in V} v(y) \sum_{x \in V} k(x, y) - \frac{1}{2} \sum_{x \in V} v(x) \sum_{y \in V} k(x, y),$$

which, by Kirchoff's node law, equals

$$-v(a) \operatorname{div}_k(a) - v(b) \operatorname{div}_k(b) = (v(b) - v(a)) \operatorname{div}_k(a),$$

since $\operatorname{div}_k(b) = -\operatorname{div}_k(a)$. This concludes the proof. □

In particular, applying Proposition 5.3 to the case when k equals the unit current flow i itself and using (5-2), we find that

$$\text{Energy}(i) = (v(b) - v(a))\,\text{div}_i(a) = \mathcal{R}_{\text{eff}}(a,b). \tag{5-5}$$

Theorem 5.4 (Thomson's principle). *The unique unit flow from a to b that minimizes energy is the unit current flow.*

Proof. Let j be a unit flow from a to b and i the unit current flow from a to b. Then $k := j - i$ is also a flow from a to b, but of strength zero. We have

$$\begin{aligned}\text{Energy}(j) &= \sum_{e\in E} R(e)(i(e)+k(e))^2 \\ &= \text{Energy}(i) + \text{Energy}(k) + 2\sum_{e\in E} R(e)i(e)k(e).\end{aligned}$$

Using Proposition 5.3, we see that the cross-term $\sum_{e\in E} R(e)i(e)k(e)$ equals zero since $\text{div}_k(a) = 0$. Therefore, the energy of j is strictly greater than the energy of i, unless $\text{Energy}(k) = 0$ in which case $k \equiv 0$ and $i \equiv j$. □

Corollary 5.5. *Effective resistance is a* (*not necessarily strictly*) *increasing function of the edge resistances.*

Proof. Let $R(e) \le R'(e)$ for every $e \in E$, and let i and i' be the corresponding unit current flows from a to b. Then

$$\begin{aligned}\mathcal{R}_{\text{eff}} &= \sum_{e\in E} R(e)(i(e))^2 && \text{(by (5-5))} \\ &\le \sum_{e\in E} R(e)(i'(e))^2 && \text{(by Thomson's principle)} \\ &\le \sum_{e\in E} R'(e)(i'(e))^2 && \text{(by assumption)} \\ &= \mathcal{R}'_{\text{eff}} && \text{(by (5-5)).}\end{aligned}$$

□

Example 5.6. A *tree* T is a connected graph with no loops. It follows that given two vertices x and y on a tree, there is a unique *geodesic* (curve of minimal length). In this case the effective resistance and the graph distance coincide:

$$\mathcal{R}_{\text{eff}}(x,y) = d_G(x,y),$$

because no current can flow along edges that are not part of the unique geodesic.

Moreover, in a graph G we always have

$$\mathcal{R}_{\text{eff}}(x,y) \le d_G(x,y). \tag{5-6}$$

To see this, pick a shortest curve γ connecting x to y and note that it will not have any loops, since removing an edge along a loop does not affect the connectedness of γ. Then complete γ to a *spanning tree* T for G, namely a tree on the same N vertices of G which is also a subgraph of G. By monotonicity, Corollary 5.5, we have $\mathcal{R}_{\text{eff}}(x, y; G) \le \mathcal{R}_{\text{eff}}(x, y; T)$, because removing an edge from G is equivalent to setting its resistance to be ∞. On the other hand, since T is a tree, $\mathcal{R}_{\text{eff}}(x, y; T) = d_T(x, y) = d_G(x, y)$.

Proposition 5.7. *Effective resistance is a metric.*

Proof. Symmetry follows by reversing the current flow. Nondegeneracy follows from (5-5) and the existence of a unit current flow. Finally, the triangle inequality is verified by noticing that inputting a unit current flow at x, extracting it at z, then reinputting it at z and extracting at y is the same as just inputting it at x and extracting it at y. □

Computing effective resistance with matrices.

Adjacency matrix. The *adjacency matrix* of a graph is a way to represent which nodes of a graph are adjacent to each other. Given a simple graph $G = (V, E)$, let A be the matrix with entries

$$A(i, j) = \begin{cases} 1 & (i, j) \in E, \\ 0 & \text{otherwise.} \end{cases}$$

(We write $(i, j) \in E$ instead of $\{i, j\} \in E$ by abuse of notation, even though (i, j) is an ordered pair.)

If the graph is weighted with weight $W(i, j)$, then we set

$$A(i, j) = \begin{cases} W(i, j) & (i, j) \in E, \\ 0 & \text{otherwise.} \end{cases}$$

Combinatorial Laplacian. Let G denote a graph with vertex set $V = \{1, 2, \dots, N\}$ and edge set E. Then the *combinatorial Laplacian* L is an $N \times N$ matrix defined by

$$L(i, j) = \begin{cases} d_i & \text{if } i = j, \\ -W(i, j) & \text{if } (i, j) \in E, \\ 0 & \text{otherwise.} \end{cases}$$

where d_i is the degree of the vertex i:

$$d_i := \sum_j A(i, j).$$

Letting D be the diagonal matrix $D = \text{Diag}(d_i)$, we see that $L = D - A$.

Also recalling the transition matrix P for the random walk on G defined in (4-1), we see that

$$L = D(I - P).$$

The square root of the Laplacian. Given an edge e joining two nodes, arbitrarily assign one of these two nodes to be e's *tail* and the other one to be e's *head.* Once this choice is made we will write $e = (x, y)$ to mean that x is the tail and y the head.

Consider the following $M \times N$ matrix, where $M = |E|$ and $N = |V|$:

$$B(e,x) := \begin{cases} 1 & \text{if } x \text{ is } e\text{'s head,} \\ -1 & \text{if } x \text{ is } e\text{'s tail,} \\ 0 & \text{otherwise.} \end{cases} \tag{5-7}$$

We claim that

$$L = B^T WB, \tag{5-8}$$

where W is the $M \times M$ diagonal matrix with entries $W(e)$ (the weight of edge e). To check this, notice that

$$(B^T WB)(x, y) = \sum_e B^T(x,e)W(e)B(e, y) = \sum_e B(e,x)B(e, y)W(e).$$

If $x \neq y$, the only term that survives in the sum is when x and y are neighbors and $e = (x, y)$, then we get $-W(x, y)$. If $x = y$, any edge from x contributes $B(e, x)^2 W(e) = W(e)$, so we get the degree at x.

Quadratic form. A *quadratic form* is a function of the form $Q(x) = x^T Ax$, where A is a symmetric matrix. Since the combinatorial Laplacian is symmetric we can write

$$\begin{aligned} v^T Lv &= \sum_{x,y\in V} v(x)L(x, y)v(y) = v^T B^T WBv = \|W^{1/2}Bv\|_2^2 \\ &= \sum_e W(e)(Bv)(e)^2 = \frac{1}{2} \sum_{(x,y)\in E} W(x, y)(v(x) - v(y))^2. \end{aligned}$$

It follows that the quadratic form associated to the Laplacian is positive semidefinite, and therefore its eigenvalues are nonnegative.

Remark 5.8. If v is the potential resulting from inputting a unit current flow i at a and extracting it at b, then using Ohm's law and (5-5) we see that the quadratic form satisfies:

$$v^T Lv = \frac{1}{2} \sum_{(x,y)\in E} W(x, y)(v(x)-v(y))^2 = \sum_e R(e)i(e)^2 = \text{Energy}(i) = \mathcal{R}_{\text{eff}}(a,b).$$

The kernel of the Laplacian. We claim that the kernel of the Laplacian consists of the constant vectors:

$$\operatorname{Ker} L = \operatorname{Ker}(W^{1/2}B) = \operatorname{Span}\{[1 \cdots 1]^T\}.$$

We use the quadratic form to check this:

$$\begin{aligned} Lv = 0 &\iff v^T L v = 0 \\ &\iff \|W^{1/2}Bv\|_2^2 = 0 \\ &\iff \sum_{x,y} W(x,y)(v(x) - v(y))^2 = 0 \\ &\iff v(x) = v(y) \quad \text{for all } x, y. \end{aligned}$$

In particular, the smallest eigenvalue of the Laplacian is 0. Traditionally, we label the eigenvalues from 0 to $N-1$, as follows:

$$\lambda_0 = 0 < \lambda_1 \leq \cdots \leq \lambda_{N-1}. \tag{5-9}$$

Diagonalizing the Laplacian. Being symmetric, the combinatorial Laplacian L can be diagonalized, that is, there are eigenvectors $u_0, u_1, \ldots, u_{N-1}$ such that

$$L = \sum_{i=1}^{N-1} \lambda_i u_i u_i^T = U \Lambda U^T.$$

Where $U = [u_0\ u_1 \cdots u_{N-1}]$, $u_0 = (1 \cdots 1)^T$ and Λ is the diagonal matrix of eigenvalues,

$$\Lambda = \begin{bmatrix} 0 & \cdots & \cdots & 0 \\ 0 & \lambda_1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_{N-1} \end{bmatrix}.$$

Recall that given a vector u, the matrix uu^T is a rank one matrix with range the line spanned by u.

Since the Laplacian has a nontrivial kernel, it is not invertible. However, we can still define a pseudoinverse, called the *Green operator*:

$$\mathcal{G} = \sum_{i=1}^{N-1} \frac{1}{\lambda_i} u_i u_i^T.$$

Then $\operatorname{Ker} \mathcal{G} = \operatorname{Ker} L$ and

$$L\mathcal{G} = \mathcal{G}L = \sum_{i=1}^{N-1} u_i u_i^T,$$

which is also equal to the projection in $\mathbb{R}^N$ onto $\operatorname{Span}(u_1, \ldots, u_{N-1})$.

Kirchoff's and Ohm's laws revisited. Let I_{ext} denote the current injected at point a and extracted at point b, which can be thought of as an $N \times 1$ vector:

$$[0 \ \cdots \ 0 \ \ -|I_a| \ \ 0 \ \cdots \ 0 \ \ |I_b| \ \ 0 \ \cdots \ 0]^T.$$

The current $i(e)$ for each edge e can be written as an $M \times 1$ vector, since there are M edges. Since each edge e is assigned a head and a tail, then $i(e)$ will either be positive or negative depending on whether the current flows from tail to head or vice versa. Then Kirchoff's node law can be written in matrix form as

$$B^T i = I_{\text{ext}}, \tag{5-10}$$

where B is the square root of the Laplacian defined in (5-7).

To see this check that at every vertex x,

$$I_{\text{ext}}(x) = \sum_{e \in E} B^T(x, e) i(e) = \sum_{e \in E} B(e, x) i(e).$$

The only terms that survive in this sum correspond to edges e with either their head or tail at x. Let's first assume that each edge was oriented in such a way that the current always flows from tail to head. Then $B(e, x) = 1$ for every edge with current flowing in at x and $B(e, x) = -1$ for every edge with current flowing out of x (and $i(e)$ is always positive with this choice of orientation). So the sum we get is exactly the divergence at x. If for some reason an edge is oriented against the current flow then both $B(e, x)$ and $i(e)$ change sign so their product does not.

Ohm's law says that the resulting voltage v on the network satisfies

$$i(x, y) = \frac{v(y) - v(x)}{R(x, y)} = W(x, y)(v(y) - v(x)).$$

In matrix form Ohm's law can be written as:

$$i = WBv, \tag{5-11}$$

as the following computation shows:

$$(WBv)(e) = \sum_{z \in V} W(e) B(e, z) v(z) = W(e)(v(y) - v(x)),$$

where $e = (x, y)$, since y is the head and x the tail of e.

Combining (5-10) and (5-11) with (5-8), we see that

$$I_{\text{ext}} = B^T W B v = L v.$$

We interpret this as an inhomogeneous problem to solve in v with I_{ext} given. However, as we've seen L is not generally invertible, unless I_{ext} is perpendicular to the kernel of L, that is, $\text{Span}\{[1 \cdots 1]^T\}$. That's exactly our case, since the current input at a equals the current output at b. Therefore, we can apply Green's operator to both sides and get the potential drop

$$v = \mathcal{G}Lv = \mathcal{G}I_{\text{ext}}.$$

In other words, writing

$$\eta_{ab} := [0 \ \cdots \ 0 \ -1 \ 0 \ \cdots \ 0 \ 1 \ 0 \ \cdots \ 0]^T = I_{\text{ext}},$$

we get that the effective resistance

$$\mathcal{R}_{\text{eff}}(a,b) = v(b) - v(a) = \eta_{ab}^T v = \eta_{ab}^T \mathcal{G} \eta_{ab}.$$

If a and b are neighbors, then η_{ab}^T is a row of B corresponding to the edge $e = (a,b)$, and thus

$$(B\mathcal{G}B^T)(e,e) = \mathcal{R}_{\text{eff}}(e).$$

6. Concepts imported from function and potential theory

Capacity. A function u defined on the vertices induces a *gradient* ρ_u on the edges:

$$\rho_u(e) := |u(x) - u(y)| \quad \text{for } e = \{x,y\}.$$

The *energy* of ρ_u is

$$\mathcal{E}(\rho_u) = \sum_{e \in E} \rho_u(e)^2. \tag{6-1}$$

On a weighted graph we modify this definition as follows:

$$\mathcal{E}(\rho_u) = \sum_{e \in E} \rho_u(e)^2 w(e).$$

Given two nodes a and b in a graph G, we will minimize the energy among all the functions $u : V \to \mathbb{R}$ with $u(a) = 0$ and $u(b) = 1$. We define the *capacity* between a and b to be

$$\operatorname{Cap}(a,b) = \min_{\substack{u(a)=0 \\ u(b)=1}} \mathcal{E}(\rho_u).$$

The function u that attains the minimum is called the *capacitary* function for a and b.

Assuming that each edge has unit resistance, recall that Thomson's principle (Theorem 5.4) and (5-5) imply that the effective resistance between a and b can be computed by minimizing the energy of all unit flows between a and b, and that the minimum is achieved for the unit *current* flow. The electric potential v that gives rise to the unit current flow can be used to interpret the unsigned flow as a "gradient". In fact, by Ohm's law, for $e=\{x,y\}$,

$$|i(e)| = |v(x)-v(y)|.$$

The electric potential v is only defined up to an additive constant, what matters is the drop $|v(a)-v(b)|$ which equals $\mathcal{R}_{\text{eff}}(a,b)$.

So, dividing v by $\mathcal{R}_{\text{eff}}(a,b)$ and shifting by a constant C if necessary, we get a function

$$U := \frac{v}{\mathcal{R}_{\text{eff}}(a,b)} + C, \tag{6-2}$$

such that $U(a)=0$, $U(b)=1$, and whose gradient is

$$\rho_U(e) = \frac{|i(e)|}{\mathcal{R}_{\text{eff}}(a,b)}.$$

Computing the energy of ρ_U and using (5-5) we get an upper bound for $\operatorname{Cap}(a,b)$:

$$\operatorname{Cap}(a,b) \le \frac{1}{\mathcal{R}_{\text{eff}}(a,b)} = \mathcal{C}_{\text{eff}}(a,b).$$

This inequality is in fact an equality, even though the two sides are obtained by minimizing the energy of very different objects: flows that don't necessarily admit potentials, on one hand, and gradients which are obtained from a "potential" function defined on the vertex set, on the other hand.

Proposition 6.1. *We always have*

$$\operatorname{Cap}(a,b) = \mathcal{C}_{\text{eff}}(a,b),$$

and the function U defined in (6-2) *is the capacitary function for a and b.*

Proof. The quadratic form induced by the combinatorial Laplacian of Section 5 can be used to compute the energy of gradients, since

$$u^T L u = \sum_{\{x,y\}\in E} |u(x)-u(y)|^2 = \mathcal{E}(\rho_u).$$

In other words, if we number the nodes $1,\dots,N$, the capacity $\operatorname{Cap}(i,j)$ for $i<j$ is computed by minimizing the quadratic form restricted to the affine subspace of codimension 2:

$$A_{ij} = \bigl\{x=(x_1,\dots,x_N)\in\mathbb{R}^N : x_i=0,\ x_j=1\bigr\}.$$

This can be handled with the method of Lagrange multipliers. Let $g_i(x)=x_i$ for $i=1,\dots,N$ and $f(x)=x^T L x$. Then given $i<j$,

$$\nabla f = \lambda_i \nabla g_i + \lambda_j \nabla g_j, \tag{6-3}$$

for some parameters λ_i and λ_j.

Notice that $\nabla g_i = e_i$, the standard unit vector in the i-th direction. Also $\nabla f(x) = 2Lx$. Let $w = (w_1, \dots, w_n)$ be a solution of (6-3). Then, interpreted as a function, w is harmonic at every node except possibly for nodes i and j. Moreover, $w_i = 0$ and $w_j = 1$. By the maximum principle, there is a unique harmonic function with these boundary values. Therefore the solution to the Lagrange multipliers problem coincides with U from (6-2), the harmonic function obtained by renormalizing the electric potential v arising from the effective resistance problem. □

Modulus of curve families. We relax the notion of gradient and consider arbitrary *densities* $\rho : E \to [0, +\infty)$. The *$\rho$-length* of a curve γ is then

$$\rho\text{-length}(\gamma) = \sum_{e \in E(\gamma)} \rho(e). \tag{6-4}$$

We measure the energy of ρ as done before in (6-1) for gradients.

A *curve family* Γ is a collection of curves γ in a graph G. We say that ρ is *admissible* for the curve family Γ if

$$\rho\text{-length}(\gamma) \geq 1 \quad \text{for all } \gamma \in \Gamma. \tag{6-5}$$

We write $\mathcal{A}$ for the family of all admissible densities for a given curve family Γ.

The *modulus* of Γ is

$$\operatorname{Mod}(\Gamma) = \inf_{\rho \in \mathcal{A}} \mathcal{E}(\rho). \tag{6-6}$$

The advantage of modulus is that any choice of admissible density gives rise to an upper-bound. If the family Γ contains a constant curve, then its modulus is infinite. Otherwise, choosing $\rho \equiv 1$ we see that $\operatorname{Mod}(\Gamma)$ is bounded above by $|E|$, the number of edges.

Given two nodes a and b, let $\operatorname{Mod}(a, b)$ be the modulus of the curve-family consisting of all curves γ that contain both a and b. We call this curve-family $\Gamma(a, b)$.

It turns out that modulus and capacity are closely related concepts.

Proposition 6.2. *We always have*

$$\operatorname{Mod}(a, b) = \operatorname{Cap}(a, b).$$

Proof. Let U be the capacitary function for a and b defined in (6-2), whose gradient ρ_U has energy $\mathcal{E}(\rho_U)$ equal to $\operatorname{Cap}(a, b)$. We first show that ρ_U is an admissible ρ-density for $\Gamma(a, b)$. Let γ be an arbitrary curve from a to b. Then, since γ is connected, it must contain a chain of vertices $x_0 = a$, x_1, ..., $x_m = b$ so that $x_j \sim x_{j+1}$ for $j = 0, \dots, m-1$. By the triangle inequality,

$$\begin{aligned} 1 = |U(a) - U(b)| &= \left| \sum_{j=0}^{m-1} (U(x_{j+1}) - U(x_j)) \right| \le \sum_{j=0}^{m-1} |U(x_{j+1}) - U(x_j)| \\ &= \sum_{j=0}^{m-1} \rho_U(\{x_j, x_{j+1}\}) \le \sum_{e \in E(\gamma)} \rho_U(e). \end{aligned}$$

So ρ_U is admissible for $\Gamma(a, b)$ and

$$\operatorname{Mod}(a, b) \le \mathcal{E}(\rho_U) = \operatorname{Cap}(a, b).$$

Conversely, let ρ be an arbitrary admissible density for $\Gamma(a, b)$. Without loss of generality, we can assume there is a curve $\gamma_0 \in \Gamma(a, b)$ such that ρ-length$(\gamma_0) = 1$, because otherwise we could scale ρ by dividing by the shortest ρ-length and still have an admissible density.

Define a function u on the vertices as the ρ-length of a shortest curve from a:

$$u(x) = \min_{\gamma : a \rightsquigarrow x} \rho\text{-length}(\gamma),$$

Then $u(a) = 0$ (the constant curve has length zero), and $u(b) = 1$ because of the curve γ_0 mentioned above.

Furthermore, for an arbitrary edge $e = \{x, y\}$,

$$u(y) \le u(x) + \rho(e),$$

because the shortest curve from a to x followed by the edge e is a curve from a to y. Therefore, inverting the roles of x and y, we find that the gradient of u satisfies

$$\rho_u(e) = |u(x) - u(y)| \le \rho(e).$$

This in turn implies that $\mathcal{E}(\rho_u) \le \mathcal{E}(\rho)$; in other words

$$\operatorname{Cap}(a, b) \le \mathcal{E}(\rho).$$

Since ρ was an arbitrary admissible density, we can minimize over ρ and get

$$\operatorname{Cap}(a, b) \le \operatorname{Mod}(a, b). \qquad \square$$

Putting Propositions 6.1 and 6.2 together we obtain:

Theorem 6.3. *The three concepts of effective conductance, capacity and modulus coincide*:

$$\mathcal{C}_{\text{eff}}(a, b) = \operatorname{Cap}(a, b) = \operatorname{Mod}(a, b).$$

We can now exploit the definition of modulus as an infimum and obtain a comparison between the epidemic quasimetric and effective conductance.

Proposition 6.4. $d_G(a, b)^2 \operatorname{Mod}(a, b) \le \operatorname{Epidemic}(a, b).$

Corollary 6.5. $$\text{Epidemic}(a,b) \geq d_G(a,b)^2 \mathcal{C}_{\text{eff}}(a,b) = \frac{d_G(a,b)^2}{\mathcal{R}_{\text{eff}}(a,b)}.$$

Proof of Proposition 6.4. Recall some of the notations introduced to define the epidemic quasimetric. We considered $\Omega(a, d_G(a,b))$, the subgraph of G induced by the vertices that are in the ball $\mathcal{B}(a, d_G(a,b))$. Define a ρ-density for $\Gamma(a,b)$ by letting

$$\rho(e) = \begin{cases} 1/d_G(a,b) & \text{if } e \text{ is an edge in} \Omega := \Omega(a, d_G(a,b)) \cup \Omega(b, d_G(a,b)), \\ 0 & \text{otherwise.} \end{cases}$$

We claim that this ρ-density is admissible for $\Gamma(a,b)$. To see this pick an arbitrary curve γ that contains both a and b. By definition of graph distance (2-2), we have $\text{length}_G(\gamma) \geq d_G(a,b)$, and this takes care of curves that stay in Ω. Some curves might actually exit Ω, but if they do then they must exit the ball $\mathcal{B}(a, d_G(a,b))$ and therefore they will again have ρ-length greater than one.

By (6-6) we get

$$\text{Mod}(a,b) \leq \frac{|\Omega|}{d_G(a,b)^2}. \qquad \square$$

7. Explicit and numerical computations

Some closed form expressions. We begin by calculating some of these quantities exactly, for special cases or families of cases.

Example 7.1 (Path graphs). For each $N \in \mathbb{N}$, P_N is the unique graph on N vertices that can be labeled $x_1, \dots, x_N$ so that $x_j \sim x_{j+1}$ for $j = 1, \dots, N-1$.

The diameter of P_N is $N-1$.

Being a tree, effective resistance on P_N is equal to the graph distance:

$$\mathcal{R}_{\text{eff}}(x_i, x_j) = d_G(x_i, x_j) = j - i \quad \text{for } i < j.$$

On the other hand,

$$\text{Epidemic}(x_i, x_j) = \min\{j-i, i-1\} + 2(j-i) + \min\{j-i, N-j\}.$$

To study how much is lost in the inequality of 6.5 we consider the discrepancy

$$\delta(a,b) := \frac{\mathcal{R}_{\text{eff}}(a,b)\,\text{Epidemic}(a,b)}{d_G(a,b)^2}.$$

In the case of the path graph P_N, we have

$$2 \leq \delta(x_i, x_j) = \min\left\{1, \frac{i-1}{j-i}\right\} + 2 + \min\left\{1, \frac{N-j}{j-i}\right\} \leq 4.$$

Example 7.2 (star graph and complete graph). The ratio δ can grow when the diameter of a tree is very small.

For instance, the star S_N is a tree with a vertex x_0 that plays the role of hub and has $N-1$ neighbors. The diameter is 2 and the graph distances are either 1 or 2. On the other hand, the epidemic quasimetric is either N or $2(N-1)$. Therefore δ grows linearly with N.

The complete graph K_N is the graph on N vertices with the maximum number of edges: every vertex has $N-1$ neighbors. Here the diameter is 1, and the graph distance is constant equal to 1. Effective resistance can be computed as follow: fix two nodes x and y and let $\widetilde{V}$ be the remaining nodes in $V \setminus \{x, y\}$. By symmetry the potential induced by a unit current flow between a and b will be constant on $\widetilde{V}$. Hence no current will flow along the K_{N-2} complete graph induced by $\widetilde{V}$. We get that $\widetilde{V}$ can be thought as a single vertex connected by $N-2$ edges to x and y respectively. Using the high-school physics rules for circuits in series and in parallel, we find that

$$\mathcal{R}_{\text{eff}}(x, y) = \frac{2}{N}.$$

The epidemic quasimetric in this case is constant and equal to $N(N-1)$. Hence δ is also constant and equal to $(N-1)/2$. Again we see that δ grows linearly with N.

One might conjecture that the discrepancy δ always grows at most linearly in the number of vertices, but that turns out to be false. Using (5-6), the volume of the complete graph K_N and bounding $d_G(a,b) \geq 1$, we find that $\delta(a,b) \leq N^2$. This quadratic worst behavior can be achieved by letting G be the graph obtained from the complete graph K_N by picking a vertex a and connecting it to a new vertex b by a single edge. In this situation $d_G(a,b) = 1$, and $\mathcal{R}_{\text{eff}}(a,b) = 1$, but $\text{Epidemic}(a,b) = N(N-1)/2 + 2$, so $\delta(a,b) = O(N^2)$. One could say that in this case the epidemic quasimetric captures a feature of the graph G that effective resistance does not see.

Computing the epidemic quasimetric. We used Matlab to study specific examples. We are given a simple (possibly weighted) graph in the form of an adjacency matrix; see page 111. In some cases, the matrices are too large and have to be entered in list form. This consists of two columns of labels from 1 to N, where each row represents an edge in the graph.

We call the adjacency matrix K and normalize it so that all the nonzero entries are equal to 1. The matrix K is zero along the diagonal so we let $B = K + I$. We then take powers B^k of B. An entry in the (i, j) spot of B^k is nonzero if and only if at least one of the terms in the sum

$$\sum_{l_1,\ldots,l_k} B(i, l_1)B(l_1, l_2)\cdots B(l_k, j)$$

is nonzero, that is, if and only if there is at least one chain $(i, l_1), \dots, (l_k, j)$ connecting i and j. Each step (l_s, l_{s+1}) either moves to a neighbor or stays at the given node. We again normalize the entries of B^k to be 1 when nonzero. Let B_k be the normalized version of B^k and define $B_0 = I$.

Next, given a node t we look at the row t in the matrix B_k and find a 1 in the column corresponding to each node that is graph distance less or equal to k from t. We use this row to form a diagonal matrix M and compute the matrix

$$K_{t,k} := MKM.$$

The entries of $K_{t,k}$ are of the form

$$\begin{aligned} K_{t,k}(i,j) &= M(i,i)K(i,j)M(j,j) \\ &= \begin{cases} 1 & \text{if } i,j \in \mathcal{B}(t,k) \text{ and } \{i,j\} \in E, \\ 0 & \text{else.} \end{cases} \end{aligned} \tag{7-1}$$

In other words, we see that $K_{t,k}$ is the adjacency matrix of the subgraph of G induced by the vertices that are in the ball $\mathcal{B}(t,k)$ centered at t of radius k. We now compute the volume by summing all the entries of $K_{t,k}$ and dividing by 2. This is equivalent to first summing all the rows and getting the local degrees and then summing the rows and getting the sum of all the local degrees, which by the Handshake Lemma equals twice the number of edges.

Finally, recall that the graph distance between x and y is the first time y belongs to the ball $\mathcal{B}(x,k)$:

$$d_G(x,y) = \min\{k : B_k(x,y) = 1\}.$$

The way we actually compute the graph distance is as follows. We flip each zero and one in B_k, in practice we take a constant matrix O with entries $O(x,y) = 1$, and let $C_k(x,y) = O(x,y) - B_k(x,y)$. Then we define

$$D := \sum_{k=0}^{N} C_k,$$

which leads to

$$D(x,y) = \sum_{k=0}^{d_G(x,y)-1} 1 = d_G(x,y).$$

Numerical experiments using the epidemic quasimetric in clustering algorithms. The epidemic quasimetric carries more information for pairs of nodes that are close relative to the diameter of the graph.

In order to write the Matlab code and check it against a concrete graph it was really useful to have a very simple object such as the one in Figure 2.

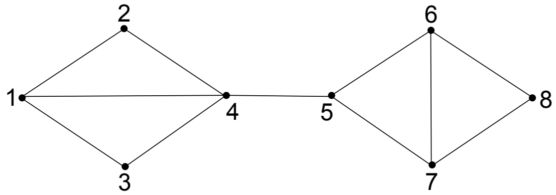


Figure 2. A simple graph.

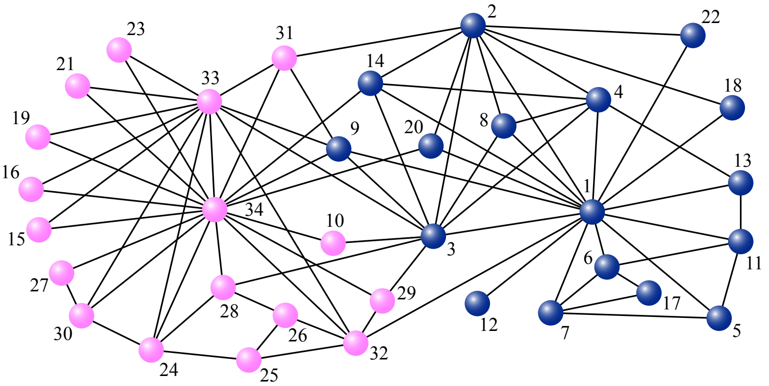


Figure 3. The original split in the Zachary karate club graph.

Once we wrote the algorithm for computing the epidemic quasimetric we tested how it would fare in the case of the classical example graph known as the Zachary karate club network,[3] shown in Figure 3. The nodes represent club members and the edges the friendship relations between them. After an argument between the two leaders, node 1 and 34, the club split into two clubs; thus the coloring of the vertices according to two distinct colors. The splitting is a real life phenomenon and the intuition is that the web of friendships should play a role.

We used the epidemic quasimetric as a measure of “similarity” between nodes in the agglomerative AGNES algorithm (this we performed in R). We obtained the dendrogram in Figure 4. This is an algorithm that begins by putting each vertex in its own class. It then “agglomerates” two classes based on the degree of similarity between them, which is computed by taking averages of all the similarities between elements of the two classes. The algorithm tries to agglomerate the least dissimilar classes first.

Using the two largest clusters we obtained a cut for the example graph that mislabels only three vertices, as can be seen in Figure 5.

[3] This 34-vertex, 78-edge graph refers to data collected by Wayne Zachary [1977] about members of a university karate club.

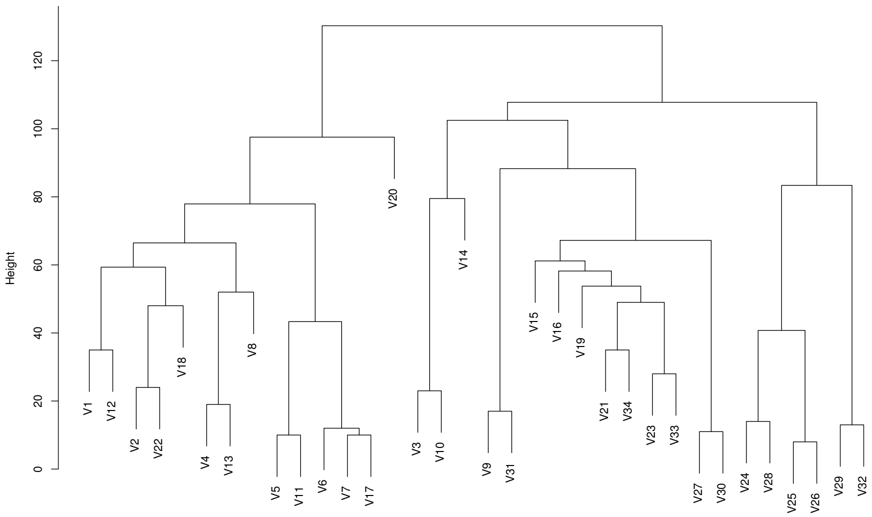


Figure 4. AGNES-produced dendrogram for the karate club graph. The value of the agglomerative coefficient was found to be 0.78.

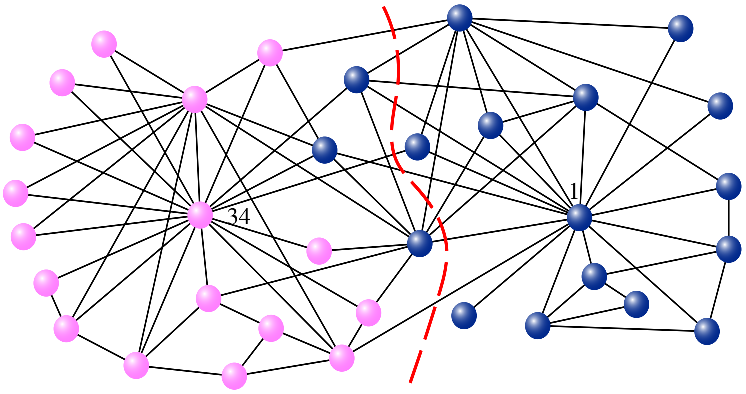


Figure 5. Epidemic cut of the karate club graph.

References

[Ahlfors 1973] L. V. Ahlfors, *Conformal invariants: topics in geometric function theory*, McGraw-Hill Book Co., New York, 1973. MR 50 #10211 Zbl 0272.30012

[Doyle and Snell 1984] P. G. Doyle and J. L. Snell, *Random walks and electric networks*, Carus Mathematical Monographs **22**, Mathematical Association of America, Washington, DC, 1984. MR 89a:94023 Zbl 0583.60065

[Grimmett 2010] G. Grimmett, *Probability on graphs: random processes on graphs and lattices*, Institute of Mathematical Statistics Textbooks **1**, Cambridge University Press, 2010. MR 2011k:60322 Zbl 1228.60003

[Levin et al. 2009] D. A. Levin, Y. Peres, and E. L. Wilmer, *Markov chains and mixing times*, American Mathematical Society, 2009. MR 2010c:60209 Zbl 1160.60001

[Macías and Segovia 1979] R. A. Macías and C. Segovia, "Lipschitz functions on spaces of homogeneous type", *Adv. in Math.* **33**:3 (1979), 257–270. MR 81c:32017a Zbl 0431.46018

[Semmes 1993] S. Semmes, "Bi-Lipschitz mappings and strong A_∞ weights", *Ann. Acad. Sci. Fenn. Ser. A I Math.* **18**:2 (1993), 211–248. MR 95g:30032

[Zachary 1977] W. Zachary, "An information flow model for conflict and fission in small groups", *J. Anthropol. Res.* **33** (1977), 452–473. Data online at http://konect.uni-koblenz.de/networks/ucidata-zachary.

Received: 2013-05-06 Accepted: 2013-09-17

jericson@uoregon.edu *Department of Mathematics, Kansas State University,*
138 Cardwell Hall, Manhattan, KS 66506, United States

Current address: *Department of Mathematics, University of Oregon,*
202 Fenton Hall, Eugene, OR 97403-1222, United States

pietro@math.ksu.edu *Department of Mathematics, Kansas State University,*
138 Cardwell Hall, Manhattan, KS 66506, United States

polarise@ksu.edu *Department of Mathematics, Kansas State University,*
138 Cardwell Hall, Manhattan, KS 66506-2602, United States

Guidelines for Authors

Authors may submit manuscripts in PDF format on-line at the Submission page at the Involve website.

Originality. Submission of a manuscript acknowledges that the manuscript is original and and is not, in whole or in part, published or under consideration for publication elsewhere. It is understood also that the manuscript will not be submitted elsewhere while under consideration for publication in this journal.

Language. Articles in *Involve* are usually in English, but articles written in other languages are welcome.

Required items. A brief abstract of about 150 words or less must be included. It should be self-contained and not make any reference to the bibliography. If the article is not in English, two versions of the abstract must be included, one in the language of the article and one in English. Also required are keywords and subject classifications for the article, and, for each author, postal address, affiliation (if appropriate), and email address.

Format. Authors are encouraged to use LaTeX but submissions in other varieties of TeX, and exceptionally in other formats, are acceptable. Initial uploads should be in PDF format; after the refereeing process we will ask you to submit all source material.

References. Bibliographical references should be complete, including article titles and page ranges. All references in the bibliography should be cited in the text. The use of BibTeX is preferred but not required. Tags will be converted to the house format, however, for submission you may use the format of your choice. Links will be provided to all literature with known web locations and authors are encouraged to provide their own links in addition to those supplied in the editorial process.

Figures. Figures must be of publication quality. After acceptance, you will need to submit the original source files in vector graphics format for all diagrams in your manuscript: vector EPS or vector PDF files are the most useful.

Most drawing and graphing packages (Mathematica, Adobe Illustrator, Corel Draw, MATLAB, etc.) allow the user to save files in one of these formats. Make sure that what you are saving is vector graphics and not a bitmap. If you need help, please write to graphics@msp.org with details about how your graphics were generated.

White space. Forced line breaks or page breaks should not be inserted in the document. There is no point in your trying to optimize line and page breaks in the original manuscript. The manuscript will be reformatted to use the journal's preferred fonts and layout.

Proofs. Page proofs will be made available to authors (or to the designated corresponding author) at a Web site in PDF format. Failure to acknowledge the receipt of proofs or to return corrections within the requested deadline may cause publication to be postponed.

involve

2014 vol. 7 no. 1

1944-4176(2014)7:1;1-8