

EXPERIMENTS WITH THE MARKOFF SURFACE

MATTHEW DE COURCY-IRELAND AND SEUNGJAE LEE

ABSTRACT. We confirm, for the primes up to 3000, the conjecture of Bourgain, Gamburd, and Sarnak on strong approximation for the Markoff surface $x^2 + y^2 + z^2 = 3xyz$ modulo primes. For primes congruent to 3 modulo 4, we find data suggesting that some natural graphs constructed from this equation are asymptotically Ramanujan. For primes congruent to 1 modulo 4, the data suggests a weaker spectral gap. In both cases, there is close agreement with the Kesten-McKay law for the density of states for random 3-regular graphs. We also study the connectedness of other level sets $x^2 + y^2 + z^2 - 3xyz = k$. In the degenerate case of the Cayley cubic, we give a complete description of the orbits.

1. INTRODUCTION

The *Markoff equation* or *Markoff surface*

$$(1.1) \quad x^2 + y^2 + z^2 = 3xyz$$

is preserved by the operations

$$(1.2) \quad m_1 : x \mapsto 3yz - x, m_2 : y \mapsto 3xz - y, m_3 : z \mapsto 3xy - z.$$

Indeed, (1.1) is a cubic equation overall, but only quadratic in each individual variable, and these moves amount to switching to the other root of the quadratic. They are called *Markoff moves* or *Vieta operations*. Markoff proved in 1880 [16] that any solution to (1.1) in nonnegative integers except $(0, 0, 0)$ can be reached by a sequence of Vieta operations and transpositions. This means that all solutions to equation (1.1) can be found quickly by navigating from the root $(1, 1, 1)$ using the moves m_1, m_2, m_3 . This can be represented graphically as a tree with a vertex for each solution, and edges between solutions that are connected by one of the moves m_1, m_2, m_3 .

For the Markoff equation over a prime field \mathbb{F}_p , it is no longer guaranteed that all solutions can be found by these moves. Does every solution mod p lift to a solution over the integers? If so, then the same sequence of Vieta moves used to reach the lift will reach its image mod p because the moves are polynomial operations in (x, y, z) . Over a finite field, instead of the Markoff tree we have a *Markoff graph* with cycles. To streamline the graphs, we found it convenient to use the following operations known as *Dehn twists* instead of the Markoff moves m_1, m_2, m_3 . The three operations D_1, D_2, D_3 are

$$(1.3) \quad D_1 : (x, y, z) \mapsto (3yz - x, z, y)$$

$$(1.4) \quad D_2 : (x, y, z) \mapsto (x, z, 3xz - y)$$

$$(1.5) \quad D_3 : (x, y, z) \mapsto (x, 3xy - z, y)$$

Date: December 18, 2018.

With either choice of generators, one never has parallel edges because $(0, 0, 0)$ is the only solution to $m_j m_k(x, y, z) = (x, y, z)$, and likewise the only solution to $D_j D_k^{-1}(x, y, z) = (x, y, z)$ for $j \neq k$. Whereas each Markoff move has on the order of p fixed points, the Dehn twists have only a bounded number. Indeed, D_2 and D_3 have no fixed points except $(0, 0, 0)$ solving (1.1). If $D_1(x, y, z) = (x, y, z)$, then $z = y$ and $2x = 3y^2$. Substituting into Markoff's equation gives $9y^4/4 + 2y^2 = 9y^4/2$. So the only solutions are $(0, 0, 0)$ and, if 8 is a square mod p , $(4, \pm\sqrt{8}, \pm\sqrt{8})/3$. For each point $(x, y, z) \neq (0, 0, 0)$ on the Markoff surface over \mathbb{F}_p , we take an edge between (x, y, z) and each of its images under D_1, D_2, D_3 . This defines a 3-regular graph with at most two loops for each prime p , often no loops at all, and in any case no parallel edges. Note that $D_j = \tau_{23} \circ m_j$, where τ_{23} is the transposition exchanging the second and third coordinates.

Note that if $x^2 + y^2 + z^2 = 3xyz$, then $(X, Y, Z) = (x, y, z)/3$ solves

$$(1.6) \quad X^2 + Y^2 + Z^2 = XYZ.$$

Over the integers, the factor $3xyz$ in (1.1) guarantees that the base solution is $(1, 1, 1)$ instead of $(3, 3, 3)$. Over \mathbb{F}_p with $p > 3$, it can be more convenient to use version (1.6) of the Markoff surface. The corresponding Markoff moves are $x \mapsto yz - x, y \mapsto xz - y, z \mapsto xy - z$. We will denote these also by m_1, m_2, m_3 as it will be clear from context whether we are using (1.1) or (1.6).

The connectedness of these graphs for all p is the question of whether *strong approximation* holds for the Markoff surface. Baragar was the first to conjecture that this connectedness does hold for all p and verified it for $p \leq 179$ (see p. 124 of [3]). The present paper extends this to $p < 3000$ and suggests that the graphs are not only connected but even form an expander family as p grows. In Section 3, we present numerical evidence that the Markoff graphs have a spectral gap. For $p \equiv 3 \pmod{4}$, the gap is almost as large as possible, while for $p \equiv 1 \pmod{4}$ it is somewhat weaker (Figure 3.1). We also discuss the bulk distribution of eigenvalues. The data suggest that this converges to the Kesten-McKay law (Figure 3.2), which has recently been confirmed theoretically by Magee and de Courcy-Ireland [8].

In addition to the numerical evidence, there are compelling theoretical reasons to believe in strong approximation. Bourgain-Gamburd-Sarnak proved that it holds unless $p^2 - 1$ has many prime factors, which happens only for rare values of p [5]. The strong approximation conjecture is equivalent to a certain group action being transitive, namely the action of Vieta moves and coordinate permutations on solutions (x, y, z) modulo p to equation (1.1), up to double sign changes $(x, y, z) \mapsto (\sigma_1 x, \sigma_2 y, \sigma_3 z)$ with each $\sigma_j = \pm 1$ and $\sigma_1 \sigma_2 \sigma_3 = 1$. Meiri and Puder show that, if $p \equiv 1 \pmod{4}$ and $p^2 - 1$ is not very smooth, so that the analysis of Bourgain-Gamburd-Sarnak shows the action to be transitive, then the resulting permutation group is either the alternating group or the symmetric group on the set of blocks (modulo sign change) [19]. They prove this for $p \equiv 3 \pmod{4}$ as well, but this requires an additional hypothesis about p . Carmon shows in the appendix to [19] that this assumption holds except for another sparse sequence of primes. It had been conjectured around the same time by Cerbu-Gunther-Magee-Peilen that the group is alternating for $p \equiv 3 \pmod{16}$ and symmetric otherwise [7]. This phenomenon of being fully transitive is a kindred spirit to expansion: Not only is the graph/action connected/transitive, but it is very robustly connected so that there are many ways to go from one point to any other.

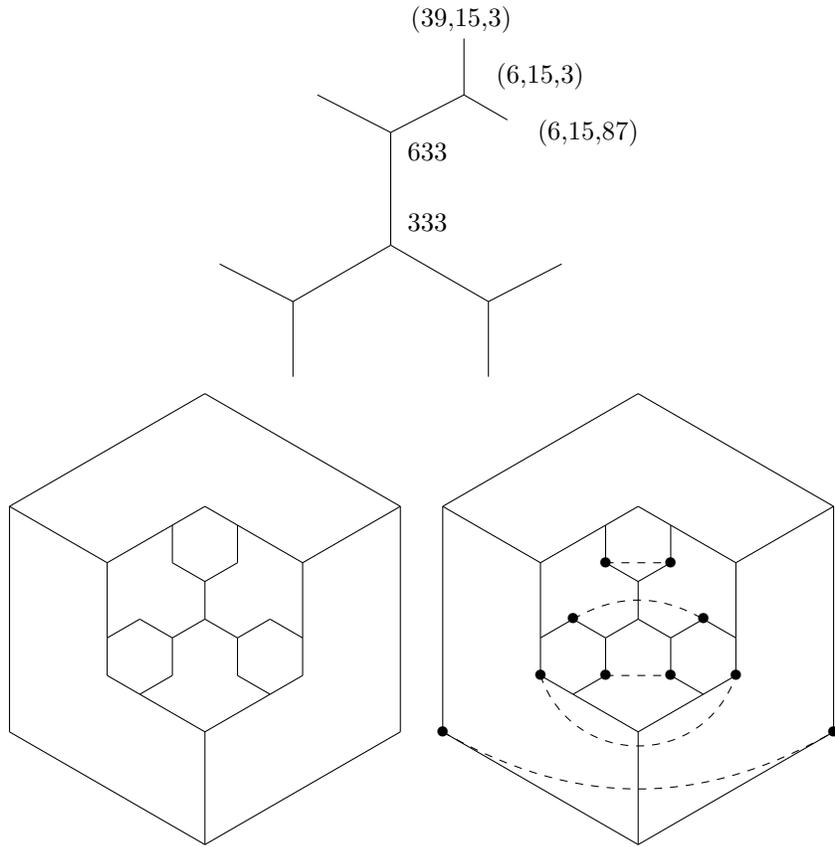


FIGURE 1.1. Top: Part of the tree of solutions to $x^2 + y^2 + z^2 = xyz$ in positive integers. Bottom left: The 28 solutions mod 7 connected by Markoff moves m_1, m_2, m_3 . Bottom right: The same solutions connected by D_1, D_2, D_3 where $D_j = \tau_{23} \circ m_j$.

For example, we consider $p = 7$ in detail. Using the form $x^2 + y^2 + z^2 = xyz$, we start from the base solution $(3, 3, 3)$. The Markoff moves lead to $(6, 3, 3)$, $(3, 6, 3)$, and $(3, 3, 6)$. At the second level, one finds the permutations of $(1, 6, 3) = m_1(3, 6, 3)$ because $6 \times 3 - 3 \equiv 1 \pmod{7}$. In characteristic 0, we would have $(15, 6, 3)$ instead of $(1, 6, 3)$, and instead of a loop $m_3(1, 6, 3) = (1, 6, 3)$ we would simply have $m_3(15, 6, 3) = (15, 6, 87)$. At the third level come $(3, 1, 4) = m_3(3, 1, 6)$ and its six permutations. At the fourth level the relation $(1, 1, 4) = m_1(3, 1, 4) = m_2(1, 3, 4)$ and its permutations lead to three cycles of length 6. The remaining Markoff move leads to, for instance, $(3, 4, 4) = m_2(3, 1, 4)$ (and two other permutations). Three cycles of length 8 are formed between $(3, 3, 3)$ and the permutations of $(3, 4, 4)$. At the fifth level, one obtains three points of the form $(4, 4, 6) = m_3(4, 4, 3)$. At the sixth level, one has all the solutions, with three points of the form $(4, 6, 6)$ completing a final cycle of length 6 together with the points $(4, 4, 6)$. This procedure is shown graphically in Figure 1.1.

The Markoff graph mod 7 has 12 loops, at the points $(4, 6, 6)$, $(1, 1, 4)$, $(1, 6, 3)$, and their permutations. If we use the generators D_1, D_2, D_3 instead of the Markoff moves, then $(4, 6, 6)$ and $(4, 1, 1)$ will still be fixed by D_1 but the other points will be joined pairwise. Note that 6 and 1 are the two square roots of $8 \equiv 1 \pmod{7}$. Also, the Dehn-neighbours of $(3, 3, 6)$ are the Markoff-neighbours of $(3, 6, 3)$. Thus using $D_1, D_2, D_3 \pmod{7}$ has reflected the graph left-to-right around $(3, 3, 3)$ and created four extra edges. As another example, the graph constructed from D_1, D_2, D_3 for $p = 11$ is shown in Figure 1.2. It has no loops, since 2 is not a square modulo 11.

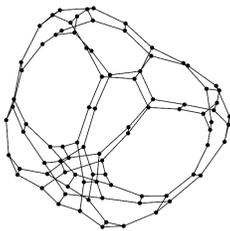


FIGURE 1.2. Markoff graph for \mathbb{F}_{11}

The same Vieta operations act on Markoff-like equations $x^2 + y^2 + z^2 - 3xyz = k$ for other values of k besides 0. The resulting graphs need not be connected, and the structure of the connected components depends on arithmetic relations between p and k . For example, if k is a square modulo p , then there will be a component of size 6 containing $(0, 0, \pm\sqrt{k})$ and its permutations. In Section 4, we give more examples and discuss some patterns in the component sizes (Table 4.1).

$k \setminus p$	5	7	11	13
0	1 40	1 28	1 88	1 208
1	4 6 16	6 16	6 160	6 112
2	36	4 6 16 24	144	196
3	16	64	6 160	6 216
4	6	64	6 160	6 112
5		36	6 72	144
6		64	40 60	128 16
7			144	144
8			40 60	196
9			4 6 16 48 48	6 216
10			16 128	6 112
11				196
12				4 6 16 48 96

TABLE 1.1. Each entry lists the sizes of the orbits of (x, y, z) satisfying $x^2 + y^2 + z^2 - 3xyz = k \pmod{p}$ under the group generated by Dehn twists, permutations, and double sign changes. The prime p runs horizontally and the level k runs horizontally. Each column can be extended periodically since k is taken modulo p .

Most conspicuously, for each $p > 3$, there is a particular value $k \equiv 4/9 \pmod{p}$ such that the graph has a large number of orbits compared to other level sets. In the

normalization $x^2 + y^2 + z^2 = xyz + k$ instead of $3xyz$, this special value is simply $k = 4$. This degenerate level set is called the *Cayley cubic*:

$$(1.7) \quad x^2 + y^2 + z^2 = xyz + 4.$$

In Section 5, we discuss this case in detail and prove the following theorem.

Theorem 1.1. *For each prime $p \geq 5$, the orbits of the action of Markoff moves and permutations on solutions of (1.7) modulo p are in bijection with those divisors of $p^2 - 1$ that are multiples of $p + 1$ or $p - 1$. The number of orbits is*

$$(1.8) \quad e_2 \prod_{q \in Q^-} (e_q + 1) + 2 \prod_{q \in Q^+} (e_q + 1) - 2.$$

where Q^+ and Q^- are disjoint sets of odd primes dividing $p^2 - 1$ such that the prime factorizations of $p + 1$ and $p - 1$ are

$$p + \varepsilon = 2 \prod_{q \in Q^+} q^{e_q}, \quad p - \varepsilon = 2^{e_2 - 1} \prod_{q \in Q^-} q^{e_q}.$$

We write $\varepsilon = (-1)^{\frac{p-1}{2}}$. Given such a divisor $\prod_q q^{f_q} \neq p^2 - 1, \frac{p^2-1}{2}$, the size of the corresponding orbit is

$$(1.9) \quad \frac{1}{2} \prod_{q: f_q < e_q} (q^2 - 1) q^{2(e_q - f_q - 1)}$$

whereas $p^2 - 1$ and $\frac{p^2-1}{2}$ correspond to orbits of twice this size.

We illustrate the notation of Theorem 1.1 for $p = 5, 7, 11$ in Table 1.2, and give some larger examples in Section 5. Some of the orbits described in Theorem 1.1 merge when we include the further symmetries $(x, y, z) \mapsto (\sigma_1 x, \sigma_2 y, \sigma_3 z)$ with signs obeying $\sigma_1 \sigma_2 \sigma_3 = 1$. In terms of divisors, we show in Section 5 that the effect is to join the orbits of t and $2t$ when the power of 2 dividing t is just one less than the power dividing $p^2 - 1$. Thus we have the following corollary, also illustrated in Table 1.2.

Corollary 1.2. *The number of orbits in the Cayley cubic, including the effect of sign changes, is*

$$(1.10) \quad (e_2 - 1) \prod_{q \in Q^-} (e_q + 1) + \prod_{q \in Q^+} (e_q + 1) - 1.$$

Another consequence of Theorem 1.1 is that the number of orbits is small compared to p .

Corollary 1.3. *The number of orbits for the Cayley cubic modulo p is at most the number of divisors of $p^2 - 1$. Hence for any $\delta > 0$ there is a number $A_\delta > 0$ such that there are at most $A_\delta p^\delta$ orbits.*

On the other hand, (1.8) shows that the number of orbits is at least e_2 . Let p be congruent to $\pm 1 \pmod{2^k}$ with $k \geq 2$. Dirichlet's theorem guarantees that there are infinitely many such primes for each k . The first one is at most 2^{kL} where L is Linnik's constant. For such primes,

$$e_2 = k + 1 > \frac{1}{L} \log_2 p$$

so Theorem 1.1 implies that along an infinite subsequence beginning $p = 5, 7, 17, 23, \dots$ of primes congruent to $\pm 1 \pmod{2^k}$ for $k = 1, 2, 3, \dots$, the number of orbits grows at least logarithmically. See Linnik’s original articles [13], [14] proving that there is a finite such L , and [22] for a recent numerical value $L \leq 5$ obtained by Xylouris. Assuming the Riemann Hypothesis, one could take L arbitrarily close to 2. If $p = 2^l - 1$ is a Mersenne prime, then

$$e_2 = l + 1 = \log_2(p + 1) + 1$$

and it follows that the Cayley cubic modulo p has more than $\log_2 p$ orbits in these cases. Over all primes up to a given magnitude, the average number of orbits is of order $\log p$, along the same lines as Titchmarsh’s divisor problem [21].

It may be surprising that the factors of $p^2 - 1$ prove decisive for the orbit structure modulo p . A preliminary change of variable naturally leads one to work in the extension \mathbb{F}_{p^2} rather than \mathbb{F}_p , and $p^2 - 1$ is the order of the multiplicative group $\mathbb{F}_{p^2}^\times$. The special feature of (5.1) is that the action linearizes, and our approach in Section 5 is to determine the orbits explicitly by “linear algebra mod $p^2 - 1$ ”.

p	ε	$p^2 - 1$	Q^+	Q^-	e_2	Orbit sizes	\pm -orbit sizes
5	1	24	{3}	\emptyset	3	1 3 4 6 12	4 6 16
7	-1	48	{3}	\emptyset	4	1 3 4 6 12 24	4 6 16 24
11	-1	120	{5}	{3}	3	1 3 4 6 12 12 36 48	4 6 16 48 48

TABLE 1.2. Examples of the notation in Theorem 1.1, showing the sizes of the orbits under permutations and Markoff moves as well as the sizes of the orbits after including double sign changes.

We recommend the book by Aigner [1] and the article by Bombieri [4] for more information about the Markoff equation over \mathbb{Z} and its many interconnections with other subjects. It is known roughly how many solutions there are subject to a given bound on the coordinates. Zagier proved that the number of solutions with x, y , and z less than T is asymptotic to $C \log(T)^2$, with $C = 0.1807\dots$ given by an explicit infinite series [23]. Mirzakhani gave a proof making use of the Markoff surface’s beautiful relation to trace identities and geometry [20]. We review Fricke’s trace identity in Section 5. Because of the connection between the Markoff surface and trace identities, the connectedness (or not) of different level sets is related to group-theoretic conjectures of McCullough and Wanderley [18]. We begin more modestly with a formula (Proposition 2.1) for the number of solutions mod p to the Markoff equation (1.1) and equations of a similar form.

2. NUMBER OF SOLUTIONS

For a finite field \mathbb{F}_p , simply counting the triples (x, y, z) with no regard for whether they solve (1.1) or not shows that the Markoff equation has at most p^3 solutions, and one expects approximately p^2 solutions since a single equation is imposed. Using the Legendre symbol to detect how many roots a quadratic equation has, we give an exact formula for the number of solutions to (1.1) modulo any prime $p \geq 5$. This is the number of vertices in the graph, which is useful to know in advance when we come to enumerate the connected components.

Proposition 2.1. *The number of solutions to the equation $x^2 + y^2 + z^2 - axyz = k$ with $a \neq 0$ in a finite field \mathbb{F}_p is*

$$(2.1) \quad N(p, k, a) = p^2 + \left(3 + \left(\frac{k}{p}\right)\right) \left(\frac{a^2k - 4}{p}\right)p + 1.$$

In particular, for the Markoff equation and its level sets $x^2 + y^2 + z^2 - 3xyz - k = 0$, the number of solutions mod p is

$$(2.2) \quad N(p, k) = \begin{cases} p^2 + 3\left(\frac{-1}{p}\right)p + 1 & \text{if } k = 0 \\ p^2 + 1 & \text{if } 9k = 4 \\ p^2 + 4\left(\frac{9k-4}{p}\right)p + 1 & \text{otherwise} \end{cases}$$

The special case of the original Markoff surface – $a = 3$ and $k = 0$ in the notation above – was previously calculated by Baragar in [3] pages 117-122, which in this case also gives the number of solutions over fields with p^k elements rather than just the prime fields. See also the note of Carlitz [6] for the general case.

Under a scaling $(x, y, z) = b \cdot (\xi, \eta, \zeta)$, the equation $x^2 + y^2 + z^2 - axyz = k$ transforms into $\xi^2 + \eta^2 + \zeta^2 - ab\xi\eta\zeta = k/b^2$. Thus we can fix a convenient value such as $a = 1$ or $a = 3$ with no loss of generality. The degenerate case where $a^2k = 4$ comes from the *Cayley cubic*, which we will see leads to a distinctive graph structure.

Proof. We sum over all x and y in \mathbb{F}_p , tallying how many z satisfy equation (1.1) for each pair (x, y) . There are 2, 1, or 0 solutions modulo p according to whether the discriminant of the resulting quadratic is a quadratic residue, 0, or a nonresidue. We can therefore express $N(p, k, a)$ using the Legendre symbol:

$$(2.3) \quad N(p, k, a) = \sum_{x \in \mathbb{F}_p} \sum_{y \in \mathbb{F}_p} \left(1 + \left(\frac{(axy)^2 - 4(x^2 + y^2 - k)}{p}\right)\right).$$

Summing the term 1 over all pairs (x, y) yields p^2 , which is the main contribution to $N(p, k, a)$ claimed in Proposition 2.1. Now we evaluate the lower-order contribution from the character sum:

$$S = \sum_{x \in \mathbb{F}_p} \sum_{y \in \mathbb{F}_p} \left(\frac{(axy)^2 - 4(x^2 + y^2 - k)}{p}\right).$$

Fixing x , we sum the Legendre symbol of $y^2((ax)^2 - 4) + 4(k - x^2)$ over y . For two values of x , namely $\pm 2/a$, we have $(ax)^2 - 4 = 0$ and then all p of the inner summands are

$$\left(\frac{4(k - x^2)}{p}\right) = \left(\frac{k - x^2}{p}\right) = \left(\frac{k - 4/a^2}{p}\right) = \left(\frac{a^2k - 4}{p}\right).$$

Thus

$$S = 2\left(\frac{a^2k - 4}{p}\right)p + \sum_{(ax)^2 \neq 4} \left(\frac{(ax)^2 - 4}{p}\right) \sum_y \left(\frac{y^2 + 4(k - x^2)/((ax)^2 - 4)}{p}\right)$$

The inner sum over y is a sum of Legendre symbols of *shifted squares* $y^2 - s$. If $s = 0$, every term is 1, except for the term 0 when $y = 0$. Thus the sum is $p - 1$ in case $k = x^2$. Otherwise, we use the following Lemma:

Lemma 2.2. *For any nonzero $s \in \mathbb{F}_p$,*

$$\sum_{z \in \mathbb{F}_p} \left(\frac{z^2 - s}{p} \right) = -1.$$

Let us finish the calculation of S and then prove Lemma 2.2. By the lemma, the sum over y is -1 , unless $k = x^2$ in which case it is $p - 1$. The contribution from $x^2 = k$ is

$$(p-1) \left(\frac{a^2k-4}{p} \right) \left(1 + \left(\frac{k}{p} \right) \right)$$

which, in particular, is 0 if there are no such x . The remaining contribution to S is the sum

$$\sum_{k \neq x^2 \neq 4/a^2} \left(\frac{(ax)^2 - 4}{p} \right) (-1).$$

The constraint $x^2 \neq 4/a^2$ may be ignored since the summand is 0 in that case. We have another sum over shifted squares $u^2 - s$ with $u = ax$ and $s = 4$, except with solutions to $u^2 = a^2k$ omitted if there are any to begin with. Therefore, by Lemma 2.2 again,

$$\sum_{k \neq x^2 \neq 4/a^2} \left(\frac{a^2x^2 - 4}{p} \right) = -1 - \left(\frac{a^2k-4}{p} \right) \left(1 + \left(\frac{k}{p} \right) \right).$$

Including all the cases, we have

$$\begin{aligned} S &= 2 \left(\frac{a^2k-4}{p} \right) p + \left(1 + \left(\frac{k}{p} \right) \right) \left(\frac{a^2k-4}{p} \right) (p-1) - \left(-1 - \left(\frac{a^2k-4}{p} \right) \left(1 + \left(\frac{k}{p} \right) \right) \right) \\ &= \left(\frac{a^2k-4}{p} \right) \left(3 + \left(\frac{k}{p} \right) \right) p + 1, \end{aligned}$$

which implies that $N(p, k, a) = p^2 + S$ is as claimed. \square

To prove Lemma 2.2, first consider the case that s is a quadratic residue, say $s = t^2$ with $t \neq 0$. Then

$$\begin{aligned} \sum_{z \in \mathbb{F}_p} \left(\frac{z^2 - s}{p} \right) &= \sum_z \left(\frac{z-t}{p} \right) \left(\frac{z+t}{p} \right) \\ &= 0 + \sum_{u \neq 0} \left(\frac{u}{p} \right) \left(\frac{u+2t}{p} \right) \end{aligned}$$

having changed variables to $u = z - t$, which runs over \mathbb{F}_p just as z does but gives no contribution when $u = 0$. The inverse u^{-1} is a quadratic residue precisely when u is, so

$$\left(\frac{u}{p} \right) \left(\frac{u+2t}{p} \right) = \left(\frac{u^{-1}}{p} \right) \left(\frac{u+2t}{p} \right) = \left(\frac{1+2t/u}{p} \right).$$

As u ranges over all non-zero values, $1 + 2t/u$ ranges over all values except 1. The sum of all the Legendre symbols is 0, so when we omit $1 = \left(\frac{1}{p} \right)$, the sum is -1 . This proves the lemma in case s is a quadratic residue. Next observe that the sum in Lemma 2.2 only depends on whether s is a quadratic residue. Thus it is $p - 1$ when

$s = 0, -1$ when s is a quadratic residue, and some value n when s is a quadratic non-residue. On the other hand

$$\sum_s \sum_z \left(\frac{z^2 - s}{p} \right) = \sum_z \sum_s \left(\frac{z^2 - s}{p} \right) = 0$$

which implies that $p - 1 - (p - 1)/2 + n(p - 1)/2 = 0$, so n must also be -1 .

3. CONNECTEDNESS AND SPECTRAL GAP FOR $k = 0$

The Cheeger constant $h(G)$ measures whether there is a bottleneck in the graph G . It is defined as

$$(3.1) \quad h(G) = \min \frac{|\partial A|}{|A|}$$

where A is any nonempty subset of G that has at most half the vertices of G and ∂A is its edge boundary. If $h(G) = 0$, then G is disconnected since there is a subset A with $|\partial A| = 0$, that is, no edges from A to $G \setminus A$. A large value of $h(G)$ means that there are many ways to escape from any given A . This is referred to as *expansion*. For a graph G with V vertices, there are about 2^{V-1} subsets of G containing at most half the vertices of G , so there are an exponential number of candidates for the minimum in (3.1). Therefore, instead of computing the exact value of $h(G)$, it is practical to estimate it. The *Cheeger inequality* states that

$$(3.2) \quad \frac{1}{2}(d - \lambda_2) \leq h(G) \leq \sqrt{2d(d - \lambda_2)}$$

where G is a d -regular graph and λ_2 is the second highest eigenvalue of the adjacency matrix of G after d . It is named after an analogous theorem of Cheeger on manifolds, the theorem for graphs being due to Dodziuk and Alon-Milman in [9], [2]. The difference $d - \lambda_2$ is known as the *spectral gap* of G . Note that d is always an eigenvalue of a d -regular graph because the “all 1’s vector” is an eigenvector when all the rows sum to d .

In particular, (3.2) shows that $h(G)$ and $d - \lambda_2$ converge to 0 or not together. Therefore, the spectral gap is an equally good measure of how well connected a graph is. A lower bound on $d - \lambda_2$ indicates the extent to which the graph is well connected. The advantage of the spectral notion is that λ_2 is easier to compute than $h(G)$. The Markoff equation in \mathbb{F}_p has roughly p^2 solutions, so G is represented by a p^2 -by- p^2 matrix, and λ_2 is its second highest eigenvalue. Although computing the eigenvalues of such an enormous matrix is costly, it is much better than checking the roughly $2^V \approx 2^{p^2}$ subsets A needed to compute $h(G)$ by brute force.

Figure 3.1 shows a striking pattern in the values of λ_2 for primes less than 3000. The black horizontal line marks $2\sqrt{2} = 2.828\dots$, and the magenta line marks 2.875. For Markoff graphs modulo a prime p congruent to 3 modulo 4, the data suggests that λ_2 approaches $2\sqrt{2}$. For prime numbers p congruent to 1 modulo 4, the data suggests that λ_2 approaches a higher value. Thus for primes congruent to 1 modulo 4, the Markoff graphs seem to exhibit weaker expansion compared to primes congruent to 3 modulo 4.

The apparent limit $2\sqrt{2}$ is a familiar number for 3-regular graphs. A d -regular graph is a *Ramanujan graph* if $\lambda_2 \leq 2\sqrt{d-1}$. These graphs are the optimal expanders. See [15] for the first construction of such graphs and more information.

Beyond λ_2 , we computed all of the eigenvalues of these matrices for a smaller range of primes. For comparison, the Kesten-McKay Law specifies the eigenvalue

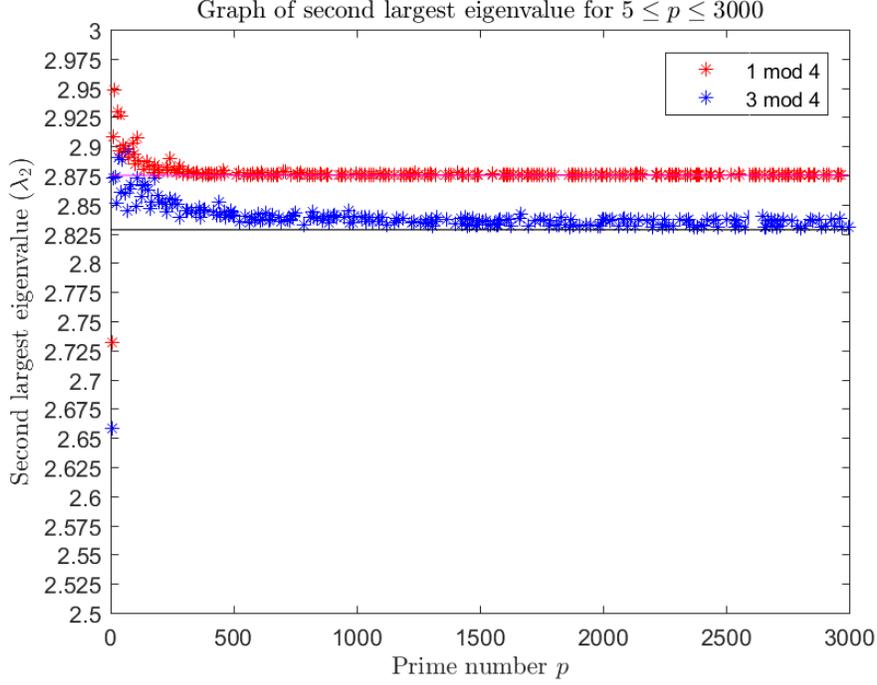


FIGURE 3.1. Graph of λ_2 for $5 \leq p \leq 3000$. The black line marks the Ramanujan case $\lambda_2 = 2\sqrt{2} = 2.828\dots$

distribution of a random d -regular graph [17], [12]. It is given by the probability density function

$$(3.3) \quad \rho_d(\lambda) = \frac{d}{2\pi} \frac{\sqrt{4(d-1) - \lambda^2}}{d^2 - \lambda^2} \mathbb{1}_{[-2\sqrt{d-1}, 2\sqrt{d-1}]}(\lambda)$$

In particular, it is supported on the interval $[-2\sqrt{d-1}, 2\sqrt{d-1}]$. For 3-regular graphs, the distribution is bimodal (the maxima are at $\pm\sqrt{7}$) and supported on the interval $[-2\sqrt{2}, 2\sqrt{2}]$. For both p congruent to 1 modulo 4 and 3 modulo 4 alike, the histogram of eigenvalues follows the Kesten-McKay Law closely (Figure 3.2). This suggests that although λ_2 converges to a higher value for p congruent to 1 modulo 4, this is only because of a vanishing proportion of exceptional eigenvalues above $2\sqrt{2}$. Indeed, Figure 3.3 seems to indicate that the number of eigenvalues above $2\sqrt{2}$ grows only like p out of the total of roughly p^2 eigenvalues. The Kesten-McKay law for Markoff graphs has recently been proved [8], although the resulting bound for the number of exceptional eigenvalues is only $p^2/\log p$ instead of p .

4. THE STRUCTURE OF GRAPHS FROM NON-ZERO k

A variation of the Markoff surface can be created by adding a constant k :

$$(4.1) \quad x^2 + y^2 + z^2 - 3xyz = k.$$

This equation is invariant under the same Vieta moves and permutations of (x, y, z) , so the Dehn twists D_1, D_2, D_3 act on it exactly as in the case $k = 0$. However, for

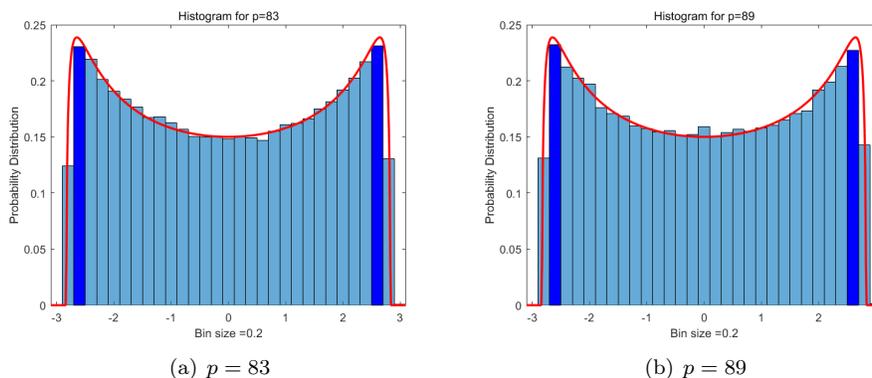


FIGURE 3.2. Histogram of eigenvalues for $p = 83$ and 89

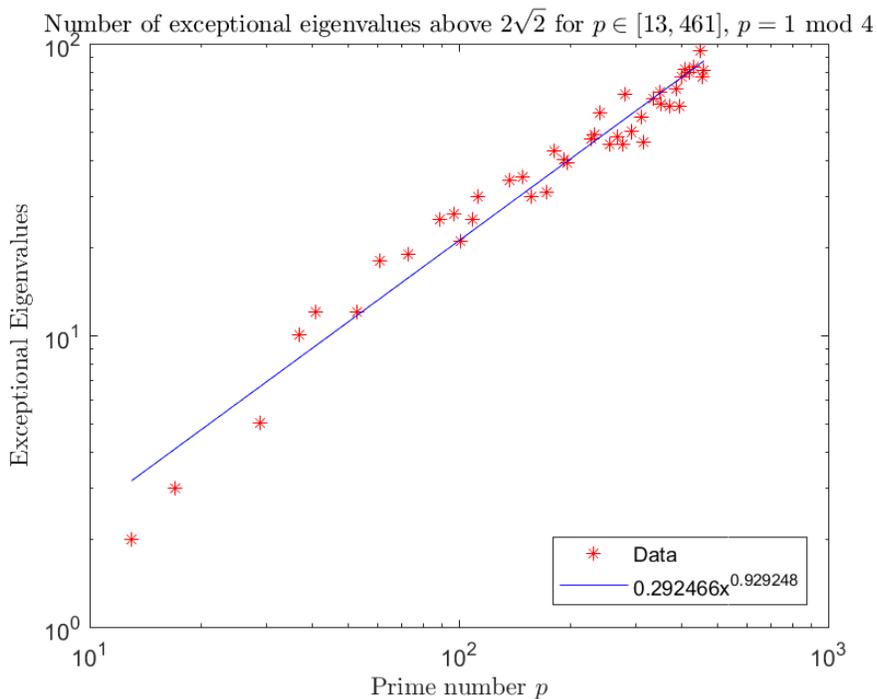


FIGURE 3.3. Log-Log Plot of Eigenvalues above $2\sqrt{2}$ for p congruent to 1 modulo 4

nonzero k , the connectedness of the resulting graph is no longer guaranteed. For example, when $p = 7$ and $k = 2$, there are 10 connected components. Also, the triple $(1, 1, 1)$ is no longer a guaranteed solution. We instead use a brute-force method to find a solution to serve as the root from which to explore using the generators D_j . Eventually, the component of that solution is fully unveiled. If its size agrees with the total number of solutions predicted in Proposition 2.1, then we have finished

constructing the graph. If not, we must use brute force again to find a solution outside this component and continue exploring from there.

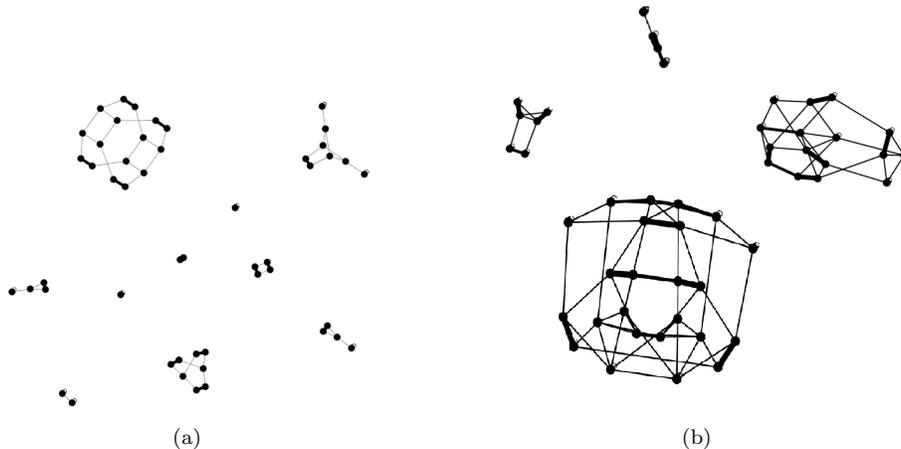


FIGURE 4.1. The solutions to $x^2 + y^2 + z^2 - 3xyz = 2 \pmod{7}$, with edges corresponding to D_1, D_2, D_3 in (a), and then in (b) with additional edges corresponding to the transpositions T_{12} and T_{23} and the sign change N_{12} . The resulting components have sizes 1, 3, 4, 6, 12, 24 in (a) or 4, 6, 16, and 24 in (b).

In an effort to make the Markoff graph connected, we enlarged the generating set by including permutations and double-sign-changes that would connect related vertices. The new operations are

$$\begin{aligned} T_{12} &: (x, y, z) \mapsto (y, x, z) \\ T_{23} &: (x, y, z) \mapsto (x, z, y) \\ N_{12} &: (x, y, z) \mapsto (-x, -y, z) \end{aligned}$$

These operations do indeed join some components together, but for many pairs (p, k) , even the extended graph is still disconnected. Figure 4.1 shows both graphs in the case $p = 7$ and $k = 2$.

Table 4.1 shows many patterns. There are a few component sizes that appear many times in the table. For example, when k is a square modulo p , there is always a size-6 component. This is a result of permutations of $(0, 0, \pm\sqrt{k})$. In particular, the graph is always disconnected when $k = 1$. The size-1 component that appears when $k = 0$ is the $(0, 0, 0)$ component which was disregarded during the discussion in Section 3.

For small primes and $k = 1$, we have just a single component besides the size 6 component containing $(0, 0, 1)$. However, when $p = 41$, there are three components of respective sizes 6, 40, and 1800. This extra component of size 40 stems from the fact that 5 is a square mod 41, so that a finite orbit constructed in characteristic 0 from the golden ratio $(1 + \sqrt{5})/2$ appears. We refer to Dubrovin for these characteristic 0 orbits in the context of the braid group, p.244 of [10]. Modulo other primes for

$k \setminus p$	5	7	11	13	17
0	1 40	1 28	1 88	1 208	1 340
1	4 6 16	6 16	6 160	6 112	6 216
2	36	4 6 16 24	144	196	6 216
3	16	64	6 160	6 216	256
4	6	64	6 160	6 112	6 16 336
5		36	6 72	144	256
6		64	40 60	128 16	36 288
7			144	144	324
8			40 60	196	4 6 16 24 96 144
9			4 6 16 48 48	6 216	6 352
10			16 128	6 112	324
11				196	256
12				4 6 16 48 96	324
13					6 216
14					256
15					6 216
16					6 352

TABLE 4.1. Each entry lists the sizes of the orbits of (x, y, z) satisfying $x^2 + y^2 + z^2 - 3xyz = k \pmod p$ under the group generated by Dehn twists, permutations, and double sign changes. The prime p runs horizontally and the level k runs horizontally. Each column can be extended periodically since k is taken modulo p .

which 5 is a quadratic residue, there is also a component of size 40 but for different levels rather than $k = 1$.

There is one k for each p that generates a Markoff graph with an especially large number of components. In Table 4.1, these pairs (p, k) are $(5, 1), (7, 2), (11, 9), (13, 12) \dots$. These occur whenever

$$(4.2) \quad 9k - 4 \equiv 0 \pmod p.$$

namely $k = (2/3)^2$, with the division by 3 understood modulo $p \geq 5$. For this special value of k , the Markoff equation becomes a form of the Cayley cubic surface. In this surface, the operations that generate the graph linearize, which leads to more components than in other cases.

5. THE CAYLEY CUBIC

The Cayley cubic is a special (degenerate) cubic surface given by

$$(5.1) \quad x^2 + y^2 + z^2 - xyz = 4.$$

This is a special case of the Markoff level set, so the same Dehn twists, permutations, and double sign changes act on its solutions modulo any prime p . The sizes of the resulting components are listed in Table 5.1.

We note that, modulo any prime $p \geq 5$, the Cayley cubic has components of size 4, 6, and 16. These come from particular solutions over \mathbb{Z} . The component of size 4 consists of $(2, 2, 2)$ and its orbit under double sign changes, namely

p	Factors of $p^2 - 1$	Component sizes
5	$2^3 \cdot 3$	4 6 16
7	$2^4 \cdot 3$	4 6 16 24
11	$2^3 \cdot 3 \cdot 5$	4 6 16 48 48
13	$2^3 \cdot 3 \cdot 7$	4 6 16 48 96
17	$2^5 \cdot 3^2$	4 6 16 24 96 144
19	$2^3 \cdot 3^2 \cdot 5$	4 6 16 48 144 144
23	$2^4 \cdot 3 \cdot 11$	4 6 16 24 48 192 240
29	$2^3 \cdot 3 \cdot 5 \cdot 7$	4 6 16 48 96 288 384
31	$2^6 \cdot 3 \cdot 5$	4 6 16 24 48 96 384 384
37	$2^3 \cdot 3^2 \cdot 19$	4 6 16 48 144 432 720
41	$2^4 \cdot 3 \cdot 5 \cdot 7$	4 6 16 24 48 96 144 576 768
43	$2^3 \cdot 3 \cdot 7 \cdot 11$	4 6 16 96 240 720 768
47	$2^5 \cdot 3 \cdot 23$	4 6 16 24 48 96 192 768 1056
53	$2^3 \cdot 3^3 \cdot 13$	4 6 16 144 336 1008 1296
59	$2^3 \cdot 3^3 \cdot 13$	4 6 16 48 48 144 384 1152 1680
61	$2^3 \cdot 3 \cdot 5 \cdot 31$	4 6 16 48 48 144 384 1152 1920
67	$2^3 \cdot 3 \cdot 11 \cdot 17$	4 6 16 240 576 1728 1920
71	$2^4 \cdot 3^2 \cdot 5 \cdot 7$	4 6 16 24 48 48 96 144 192 432 1728 2304
73	$2^4 \cdot 3^2 \cdot 37$	4 6 16 24 48 144 192 432 1728 2736
79	$2^5 \cdot 3 \cdot 5 \cdot 13$	4 6 16 24 48 96 144 336 576 2304 2688
83	$2^3 \cdot 3 \cdot 7 \cdot 41$	4 6 16 48 96 288 768 2304 3360
89	$2^4 \cdot 3^2 \cdot 5 \cdot 11$	4 6 16 24 48 144 240 384 720 2880 3456
97	$2^6 \cdot 3 \cdot 7^2$	4 6 16 24 48 96 96 192 384 768 3072 4704
101	$2^3 \cdot 3 \cdot 5^2 \cdot 17$	4 6 16 48 144 576 1200 3600 4608
103	$2^4 \cdot 3 \cdot 13 \cdot 17$	4 6 16 24 336 576 1008 4032 4608
107	$2^3 \cdot 3^3 \cdot 53$	4 6 16 48 144 432 1296 3888 5616
109	$2^3 \cdot 3^3 \cdot 5 \cdot 11$	4 6 16 48 48 144 240 432 1296 3888 5760
113	$2^5 \cdot 3 \cdot 7 \cdot 19$	4 6 16 24 96 96 288 720 1152 4608 5760
127	$2^8 \cdot 3^2 \cdot 7$	4 6 16 24 96 96 144 384 768 1536 6144 6912
131	$2^3 \cdot 3 \cdot 5 \cdot 11 \cdot 13$	4 6 16 48 48 240 336 720 1920 5760 8064
137	$2^4 \cdot 3 \cdot 17 \cdot 23$	4 6 16 24 576 1056 1728 6912 8448
139	$2^3 \cdot 3 \cdot 5 \cdot 7 \cdot 23$	4 6 16 48 96 144 288 1056 2304 6912 8448
149	$2^3 \cdot 3 \cdot 5^2 \cdot 37$	4 6 16 48 384 1200 2736 8208 9600
151	$2^4 \cdot 3 \cdot 5^2 \cdot 19$	4 6 16 24 48 384 720 1200 2160 8640 9600
157	$2^3 \cdot 3 \cdot 13 \cdot 79$	4 6 16 48 336 1008 2688 8064 12480
163	$2^3 \cdot 3^4 \cdot 41$	4 6 16 144 1296 3360 10080 11664
167	$2^4 \cdot 3 \cdot 7 \cdot 83$	4 6 16 24 48 96 192 288 768 1152 2304 9216 13776
173	$2^3 \cdot 3 \cdot 29 \cdot 43$	4 6 16 1680 3696 11088 13440
179	$2^3 \cdot 3^2 \cdot 5 \cdot 89$	4 6 16 48 48 144 144 384 432 1152 3456 10368 15840
181	$2^3 \cdot 3^2 \cdot 5 \cdot 7 \cdot 13$	4 6 16 48 48 96 144 144 336 384 432 1152 3456 10368 16128
191	$2^7 \cdot 3 \cdot 5 \cdot 19$	4 6 16 24 48 48 96 192 384 720 768 1536 3072 12288 17280
193	$2^7 \cdot 3 \cdot 97$	4 6 16 24 48 96 192 384 768 1536 3072 12288 18816
197	$2^3 \cdot 3^2 \cdot 7^2 \cdot 11$	4 6 16 96 144 240 288 1920 4704 14112 17280
199	$2^4 \cdot 3^2 \cdot 5^2 \cdot 11$	4 6 16 24 48 144 144 240 576 1200 1920 3600 14400 17280

TABLE 5.1. Sizes of components under the action of Vieta moves, coordinate permutations, and double sign changes for the Cayley cubic $x^2 + y^2 + z^2 - xyz = 4 \pmod p$

p	Factors of $p^2 - 1$	Component sizes
5	$2^3 \cdot 3$	1 3 4 6 12
7	$2^4 \cdot 3$	1 3 4 6 12 24
11	$2^3 \cdot 3 \cdot 5$	1 3 4 6 12 12 36 48
13	$2^3 \cdot 3 \cdot 7$	1 3 4 6 12 24 48 72
17	$2^5 \cdot 3^2$	1 3 4 6 12 24 36 96 108
19	$2^3 \cdot 3^2 \cdot 5$	1 3 4 6 12 12 36 36 108 144
23	$2^4 \cdot 3 \cdot 11$	1 3 4 6 12 24 48 60 180 192
29	$2^3 \cdot 3 \cdot 5 \cdot 7$	1 3 4 6 12 12 24 36 72 96 288 288
31	$2^6 \cdot 3 \cdot 5$	1 3 4 6 12 12 24 36 96 96 288 384
37	$2^3 \cdot 3^2 \cdot 19$	1 3 4 6 12 36 48 108 180 432 540
41	$2^4 \cdot 3 \cdot 5 \cdot 7$	1 3 4 6 12 12 24 24 36 72 144 192 576 576
43	$2^3 \cdot 3 \cdot 7 \cdot 11$	1 3 4 6 12 24 60 72 180 192 576 720
47	$2^5 \cdot 3 \cdot 23$	1 3 4 6 12 24 48 96 192 264 768 792
53	$2^3 \cdot 3^3 \cdot 13$	1 3 4 6 12 36 84 108 252 324 972 1008
59	$2^3 \cdot 3 \cdot 5 \cdot 29$	1 3 4 6 12 12 36 48 96 144 288 420 1152 1260
61	$2^3 \cdot 3 \cdot 5 \cdot 31$	1 3 4 6 12 12 36 48 96 144 288 480 1152 1440
67	$2^3 \cdot 3 \cdot 11 \cdot 17$	1 3 4 6 12 60 144 180 432 480 1440 1728
71	$2^4 \cdot 3^2 \cdot 5 \cdot 7$	1 3 4 6 12 12 24 24 36 36 48 72 108 192 432 576 1728 1728
73	$2^4 \cdot 3^2 \cdot 37$	1 3 4 6 12 24 36 48 108 192 432 684 1728 2052
79	$2^5 \cdot 3 \cdot 5 \cdot 13$	1 3 4 6 12 12 24 36 84 96 144 252 576 672 2016 2304
83	$2^3 \cdot 3 \cdot 7 \cdot 41$	1 3 4 6 12 24 48 72 192 288 576 840 2304 2520
89	$2^4 \cdot 3^2 \cdot 5 \cdot 11$	1 3 4 6 12 12 24 36 36 60 96 108 180 288 720 864 2592 2880
97	$2^6 \cdot 3 \cdot 7^2$	1 3 4 6 12 24 24 48 72 96 192 384 768 1176 3072 3528

TABLE 5.2. Sizes of components under the action of Vieta moves, coordinate permutations for the Cayley cubic $x^2 + y^2 + z^2 - xyz = 4 \pmod p$

$(-2, -2, 2)$, $(-2, 2, -2)$, and $(2, -2, -2)$. The orbit of size 6 consists of permutations of $(0, 0, \pm 2)$, which is a special case of the size 6 component that arises from $(0, 0, \pm\sqrt{k})$ whenever k is a square modulo p . The orbit of size 16 consists of $(1, 1, 2)$, its Vieta image $(1, 1, -1)$, and their orbits under permutations and double sign changes. Using only Markoff moves and permutations, without sign changes, one would have $(1 + 3) + 6 + (12 + 4)$ instead of $4 + 6 + 16$: Namely $(2, 2, 2)$ in its own orbit, an orbit of size 3 containing $(2, -2, -2)$, the orbit of size 6 containing $(2, 0, 0)$, an orbit of size 12 containing $(1, 1, 2)$, and another orbit of size 4 containing $(-1, -1, -1)$.

For many primes p , there is a component of size 24, and this also has a simple explanation. If 2 is a square modulo p , then among the solutions to equation (5.1) are $(\sqrt{2}, \sqrt{2}, 0)$ and its Vieta image $(\sqrt{2}, \sqrt{2}, 2)$. Permutations and double sign changes of these then yield a component of size 24. It consists of the vectors $(\varepsilon_1\sqrt{2}, \varepsilon_2\sqrt{2}, 0)$, $(\varepsilon\sqrt{2}, \varepsilon\sqrt{2}, 2)$, $(\varepsilon\sqrt{2}, -\varepsilon\sqrt{2}, -2)$ and their permutations, where $\varepsilon, \varepsilon_1, \varepsilon_2 = \pm 1$ are signs. These can also be reached from one another using Markoff moves instead of sign changes. By the supplement to the law of quadratic reciprocity, 2 is a square if

p	Component sizes
5	1 3 4 6 12
7	1 3 4 6 12 24
11	1 3 4 6 12 12 36 48
13	1 3 4 6 12 24 48 72
17	1 3 4 6 12 24 36 96 108
19	1 3 4 6 12 12 36 36 108 144
23	1 3 4 6 12 24 48 60 180 192
29	1 3 4 6 12 12 24 36 72 96 288 288
31	1 3 4 6 12 12 24 36 96 96 288 384
37	1 3 4 6 12 36 48 108 180 432 540
41	1 3 4 6 12 12 24 24 36 72 144 192 576 576
43	1 3 4 6 12 24 60 72 180 192 576 720
47	1 3 4 6 12 24 48 96 192 264 768 792
53	1 3 4 6 12 36 84 108 252 324 972 1008
59	1 3 4 6 12 12 36 48 96 144 288 420 1152 1260
61	1 3 4 6 12 12 36 48 96 144 288 480 1152 1440
67	1 3 4 6 12 60 144 180 432 480 1440 1728
71	1 3 4 6 12 12 24 24 36 36 48 72 108 192 432 576 1728 1728
73	1 3 4 6 12 24 36 48 108 192 432 684 1728 2052
79	1 3 4 6 12 12 24 36 84 96 144 252 576 672 2016 2304
83	1 3 4 6 12 24 48 72 192 288 576 840 2304 2520
89	1 3 4 6 12 12 24 36 36 60 96 108 180 288 720 864 2592 2880
97	1 3 4 6 12 24 24 48 72 96 192 384 768 1176 3072 3528
101	1 3 4 6 12 12 36 144 144 300 432 900 1152 3456 3600
103	1 3 4 6 12 24 84 144 252 432 1008 1152 3456 4032
107	1 3 4 6 12 36 48 108 324 432 972 1404 3888 4212
109	1 3 4 6 12 12 36 36 48 60 108 180 324 432 972 1440 3888 4320
113	1 3 4 6 12 24 24 72 96 180 288 540 1152 1440 4320 4608
127	1 3 4 6 12 24 24 36 72 96 108 192 384 576 1536 1728 5184 6144
131	1 3 4 6 12 12 36 48 60 84 180 252 480 720 1440 2016 5760 6048
137	1 3 4 6 12 24 144 264 432 792 1728 2112 6336 6912
139	1 3 4 6 12 12 24 36 72 144 264 288 576 792 1728 2112 6336 6912
149	1 3 4 6 12 12 36 96 288 300 684 900 2052 2400 7200 8208
151	1 3 4 6 12 12 24 36 96 180 288 300 540 900 2160 2400 7200 8640
157	1 3 4 6 12 48 84 252 672 1008 2016 3120 8064 9360
163	1 3 4 6 12 36 108 324 840 972 2520 2916 8748 10080
167	1 3 4 6 12 24 24 48 72 192 192 288 576 1152 2304 3444 9216 10332
173	1 3 4 6 12 420 924 1260 2772 3360 10080 11088
179	1 3 4 6 12 12 36 36 48 96 108 144 288 432 864 1152 2592 3960 10368 11880

TABLE 5.3. Sizes of components under the action of Vieta moves, coordinate permutations for the Cayley cubic $x^2 + y^2 + z^2 - xyz = 4 \pmod p$

and only if $p^2 - 1$ is divisible by 16:

$$(5.2) \quad \left(\frac{2}{p}\right) = (-1)^{(p^2-1)/8}.$$

This explains why, in Table 5.1, the primes 7, 17, 23, 31, 41, 47, 71, 73, 79, 89, and 97 are precisely the ones with a component of size 24. It is also a clue that the other component sizes might be explained most directly in terms of $p^2 - 1$ and its factors.

A special feature of the equation $x^2 + y^2 + z^2 = xyz$ is that when we change variables to

$$(5.3) \quad x = \xi + \xi^{-1}, \quad y = \eta + \eta^{-1}$$

the solutions for z are then

$$(5.4) \quad \xi\eta + \frac{1}{\xi\eta}, \quad \xi\eta^{-1} + \frac{1}{\xi\eta^{-1}}.$$

For $x \in \mathbb{F}_p$, there is a solution $\xi \in \mathbb{F}_p^\times$ when $x^2 - 4$ is a square mod p . Otherwise, ξ must be taken from a quadratic extension \mathbb{F}_{p^2} . Thus we let g be a generator of $\mathbb{F}_{p^2}^\times$ and write

$$(5.5) \quad x = g^u + g^{-u}, \quad y = g^v + g^{-v}$$

where the exponents are taken modulo $p^2 - 1$. The solutions for z are then

$$(5.6) \quad g^{u+v} + g^{-u-v}, \quad g^{u-v} + g^{-u+v}.$$

Note that $-u$ and u define the same x , and likewise v is equivalent to $-v$. Hence $(u, v, u-v)$ is equivalent to $(u, -v, u+(-v))$, so that all solutions can be parametrized in the form $(u, v, u+v)$ with the third coordinate equal to the sum of the others. If we had chosen a different generator, say g^w instead of g , the exponents u and v would simply be multiplied by a unit w modulo $p^2 - 1$. We are interested in the “real” solutions to 5.1, that is to say those over \mathbb{F}_p rather than \mathbb{F}_{p^2} . To have $x = g^u + g^{-u}$ lie in \mathbb{F}_p , it is necessary and sufficient that it be fixed by the Galois involution $x \mapsto x^p$. This holds if and only if $pu \equiv \pm u \pmod{p^2 - 1}$. Thus u must be a multiple of $p+1$ or $p-1$. Likewise, the second coordinate v must be a multiple of $p+1$ or $p-1$, perhaps not the same one as u . If both u and v are multiples of the same $p \pm 1$, then $p \pm 1$ also divides the sum $u+v$ and so the third coordinate is also real.

Proposition 5.1. *The Markoff moves, as well as coordinate permutations, act linearly on the coordinates $\begin{bmatrix} u \\ v \end{bmatrix}$. Explicitly, their matrices are given by*

$$(5.7) \quad [m_1] = \begin{bmatrix} 1 & 2 \\ 0 & -1 \end{bmatrix}, \quad [m_2] = \begin{bmatrix} 1 & 0 \\ -2 & -1 \end{bmatrix}, \quad [m_3] = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}.$$

and

$$(5.8) \quad [\tau_{12}] = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, [\tau_{23}] = \begin{bmatrix} -1 & 0 \\ 1 & 1 \end{bmatrix}, [\tau_{13}] = \begin{bmatrix} 1 & 1 \\ 0 & -1 \end{bmatrix}$$

where τ_{ij} is the transposition exchanging i and j . These matrices generate $\mathrm{GL}_2(\mathbb{Z})$ or, modulo $p^2 - 1$, the subgroup of matrices with determinant ± 1 .

Note that these matrices are better interpreted in PGL_2 than GL_2 because the exponents u and v are only defined up to sign. One must change the sign of the entire vector because changing the sign of only one of u, v will not keep the third coordinate $u + v$ equal to the sum of the others.

Proof. First note that m_3 exchanges $(u, v, u + v)$ with $(u, v, u - v)$, or equivalently with $(u, -v, u - v)$. We use the latter form to keep the third coordinate equal to the sum of the first two. The transposition τ_{12} sends $(u, v, u + v)$ to $(v, u, u + v)$. The transposition τ_{23} sends $(u, v, u + v)$ to $(u, u + v, v)$, or equivalently $(-u, u + v, -u + (u + v))$. The transposition τ_{13} sends $(u, v, u + v)$ to $(u + v, v, u)$ or equivalently $(u + v, -v, u)$. In matrix form acting on $\begin{bmatrix} u \\ v \end{bmatrix}$, these operations correspond to

$$[m_3] = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, [\tau_{12}] = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, [\tau_{23}] = \begin{bmatrix} -1 & 0 \\ 1 & 1 \end{bmatrix}, [\tau_{13}] = \begin{bmatrix} 1 & 1 \\ 0 & -1 \end{bmatrix}.$$

Using the relations $m_2 = \tau_{23}m_3\tau_{23}$ and $m_1 = \tau_{13}m_3\tau_{13}$, we then find

$$[m_2] = \begin{bmatrix} 1 & 0 \\ -2 & -1 \end{bmatrix}, [m_1] = \begin{bmatrix} 1 & 2 \\ 0 & -1 \end{bmatrix}.$$

To determine what group these matrices generate, note that multiplying by $[m_3] = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$ changes the sign of the second row or column:

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} [m_3] = \begin{bmatrix} a & -b \\ c & -d \end{bmatrix}, [m_3] \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} a & b \\ -c & -d \end{bmatrix}.$$

Combining this with $\tau_{23} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$, which exchanges two rows or columns, we may also change the sign of the first row or column. This is enough to obtain the standard generators for $\mathrm{SL}_2(\mathbb{Z})$:

$$T = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} = [m_3][\tau_{13}]$$

$$S = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} = [m_3][\tau_{12}]$$

One also has $S = -[\tau_{12}][m_3]$. Hence the group generated by the matrices (5.7) and (5.8) contains $\mathrm{SL}_2(\mathbb{Z})$. Multiplying by any matrix of determinant -1 , for instance τ_{23} , we obtain the other coset of $\mathrm{SL}_2(\mathbb{Z})$ in $\mathrm{GL}_2(\mathbb{Z})$. Hence these matrices generate $\mathrm{GL}_2(\mathbb{Z})$. \square

To obtain simpler graphs, we have previously used $D_1 = \tau_{23} \circ m_1, D_2 = \tau_{23} \circ m_2, D_3 = \tau_{23} \circ m_3$, which do not generate all of $\mathrm{GL}_2(\mathbb{Z})$. But for Table 5.1, we have used the full symmetry of all the Markoff moves, all the transpositions, and also double sign changes. The double sign changes do not act linearly on the exponents (u, v) . Instead, since

$$(5.9) \quad -1 = g^{\frac{p^2-1}{2}} = g^{-\frac{p^2-1}{2}}$$

their effect is to translate one or both of u, v by $(p^2 - 1)/2$. Note that, modulo $p^2 - 1$, the exponent for the third coordinate remains equal to $u + v$: It is translated by $(p^2 - 1)/2$ if only one of u, v is, or by $(p^2 - 1)/2 + (p^2 - 1)/2 = 0$ if both are. We will first determine the orbits under the linear action, and then incorporate these

three translations. The linear action is dictated by matrix arithmetic modulo $p^2 - 1$, which can be understood via the Chinese remainder theorem and the corresponding action modulo prime powers. This is the underlying reason that the factors of $p^2 - 1$ play such an important role in the structure of the Cayley cubic.

5.1. Proof of Theorem 1.1: Number of orbits. Consider the action of matrices with determinant ± 1 on $\mathbb{Z}/q^e \times \mathbb{Z}/q^e$, where q^e is a prime power. Given a vector $\begin{bmatrix} a \\ c \end{bmatrix}$ where at least one of a, c is invertible, either

$$\begin{bmatrix} a & 0 \\ c & a^{-1} \end{bmatrix} \text{ or } \begin{bmatrix} a & -c^{-1} \\ c & 0 \end{bmatrix}$$

will have determinant 1 and send $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ to $\begin{bmatrix} a \\ c \end{bmatrix}$. If neither a nor c is invertible modulo q^e , then they must be divisible by q . Let q^f be the largest power of q dividing both of them. Note that $A(q^f w) = q^f A w$, so that every vector in the orbit of w also has both coordinates divisible by q^f . Conversely, since q^f is the largest power of q dividing both, either w_1/q^f or w_2/q^f is a unit. Thus there is a matrix of determinant 1 taking $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ to w/q^f , which shows that w is in the same orbit as $\begin{bmatrix} q^f \\ 0 \end{bmatrix}$. It follows that $\text{SL}_2(q^e)$ has $e + 1$ orbits on $\mathbb{Z}/q^e \times \mathbb{Z}/q^e$ and a list of representatives is

$$\begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} q \\ 0 \end{bmatrix}, \dots, \begin{bmatrix} q^{e-1} \\ 0 \end{bmatrix}, \begin{bmatrix} q^e \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

The orbits are the same under the group of matrices of determinant ± 1 or even the full group $\text{GL}_2(\mathbb{Z}/q^e)$. Modulo a composite N , two vectors are in the same orbit if and only if their images modulo q^e are in the same orbit for each prime power factor of N . The orbits for the action of $\text{SL}_2(\mathbb{Z}/N)$ on $\mathbb{Z}/N \times \mathbb{Z}/N$ can be found by the Chinese remainder theorem, and likewise for $\text{GL}_2(\mathbb{Z}/N)$ or the subgroup of matrices with determinant $\pm 1 \pmod N$. The orbits are parameterized by all choices of $\{f(q)\}_{q|N}$, where $0 \leq f(q) \leq e_q$ specifies the highest power of q that divides the coordinates of vectors in a given orbit. Equivalently, we may think of the parameter f as a divisor of N , namely $t = \prod q^{f(q)}$, and then the corresponding orbit simply consists of vectors both of whose coordinates are divisible by t . From either perspective, the number of orbits is therefore

$$\prod_{q|N} (e_q + 1)$$

where q ranges over all prime divisors of N and e_q is the highest power of q dividing N .

With $N = p^2 - 1$, all of these orbits are candidates as orbits for the action of permutations and Markoff moves on the Cayley cubic. However, if the coordinates are not divisible by $p + 1$ or $p - 1$, one obtains solutions over the extension \mathbb{F}_{p^2} rather than \mathbb{F}_p . We must discard these orbits. We must also identify (u, v) and $(-u, -v)$ because they define the same solution (x, y, z) , but this does not change the number of orbits because each orbit is already closed under negation.

Note that $p - 1$ and $p + 1$ have no common factor except 2. If $p \equiv 1 \pmod 4$, then $p - 1$ is “highly” divisible by 2 while $p + 1$ is only once divisible by 2. If $p \equiv -1 \pmod 4$, then it is $p + 1$ that contains most of the factors of 2. To avoid considering these

cases separately, let $p + \varepsilon$ be divisible simply by 2 and $p - \varepsilon$ by the remaining factors of 2. Here, the sign is

$$(5.10) \quad \varepsilon = (-1)^{\frac{p-1}{2}} = \left(\frac{-1}{p}\right).$$

Let Q^+ and Q^- be the sets of odd primes dividing $p + \varepsilon$ or $p - \varepsilon$ respectively. These are disjoint. Thus

$$(5.11) \quad p + \varepsilon = 2 \prod_{q \in Q^+} q^{e_q}, \quad p - \varepsilon = 2^{e_2-1} \prod_{q \in Q^-} q^{e_q}.$$

The “real” orbits are the ones with either

- $f(2) \geq 1$ and $f(q) = e_q$ for all $q \in Q^+$, or
- $f(2) \geq e_2 - 1$ and $f(q) = e_q$ for all $q \in Q^-$,

or both. In the first case, $f(q)$ assumes any of e_2 values $1, 2, \dots, e_2$ for $q = 2$, must equal e_q for $q \in Q^+$, and could be any of $0, 1, \dots, e_q$ for $q \in Q^-$. In the second case, $f(2)$ takes only two values $e_2 - 1$ or e_2 , $f(q)$ must equal e_q for $q \in Q^-$, and could be any of $0, 1, \dots, e_q$ for $q \in Q^+$. In case of overlap, both coordinates are divisible by $\text{lcm}(p + \varepsilon, p - \varepsilon) = (p^2 - 1)/2$. This only happens for two orbits, namely $f(2)$ may be e_2 or $e_2 - 1$ but $f(q)$ must equal e_q for all odd q . We subtract 2 to compensate for double-counting these two orbits. The total is

$$(5.12) \quad e_2 \prod_{q \in Q^-} (e_q + 1) + 2 \prod_{q \in Q^+} (e_q + 1) - 2.$$

This is the formula stated in Theorem 1.1. □

5.2. Including double sign changes. Now we incorporate the further symmetry of the Cayley cubic under sign changes of the form $(x, y, z) \mapsto (\varepsilon_1 x, \varepsilon_2 y, \varepsilon_3 z)$ with $\varepsilon_1 \varepsilon_2 \varepsilon_3 = 1$. The double sign change $\sigma_{12} = (x, y, z) \mapsto (-x, -y, z)$ acts on the exponents (u, v) by

$$(5.13) \quad \sigma_{12} : (u, v) \mapsto \left(u + \frac{p^2 - 1}{2}, v + \frac{p^2 - 1}{2}\right)$$

because $-1 = g^{(p^2-1)/2}$. Note that, working modulo $p^2 - 1$, the exponent for $z = g^{u+v} + g^{-u-v}$ remains the sum of the exponents for x and y . The other sign changes are conjugate to this one by transpositions:

$$\begin{aligned} \sigma_{13} &= \tau_{23} \sigma_{12} \tau_{23} \\ \sigma_{23} &= \tau_{13} \sigma_{12} \tau_{13}. \end{aligned}$$

Therefore it is enough to determine how the Markoff+permutation orbits above merge under the action of σ_{12} . For the odd primes q dividing $p^2 - 1$, note that $(p^2 - 1)/2$ remains equally divisible by q , so σ_{12} acts trivially modulo q^e . For $q = 2$, note that $(p^2 - 1)/2$ is only divisible by 2^{e-1} instead of 2^e . Thus the orbits where $f(2) < e_2 - 1$ are not affected, but an orbit with $f(2) = e_2 - 1$ merges with the orbit having $f(2) = e_2$ and the same value $f(q)$ for odd q . The factor of 2 must be removed in the product over Q^+ , because the orbits divisible by $p - \varepsilon$ merge pairwise. The orbits divisible by $p + \varepsilon$ are not affected, unless $f(2) = e_2 - 1$ or e_2 . Effectively, there is one less choice for $f(2)$ so the factor e_2 is replaced by $e_2 - 1$. The two “overlap” orbits divisible by $(p^2 - 1)/2$ have been removed twice in this process, so we must add 1 to compensate. We must therefore subtract 1 instead of

2 compared to the formula above, because now only one orbit is double-counted. The total is then

$$(5.14) \quad (e_2 - 1) \prod_{q \in Q^-} (e_q + 1) + \prod_{q \in Q^+} (e_q + 1) - 1$$

as stated in Corollary 1.2, (1.10). \square

5.3. Proof of Theorem 1.1: Sizes of orbits. Let us determine the size of the orbit of $\begin{bmatrix} t \\ 0 \end{bmatrix}$, where $t = \prod q^{f(q)}$ is a divisor of $N = p^2 - 1$. For any group action, we have the orbit-stabilizer formula

$$(5.15) \quad \#\text{Orb}(t) = \frac{\#G}{\#\text{Stab}(t)}.$$

In the present case, the stabilizer consists of matrices sending $\begin{bmatrix} t \\ 0 \end{bmatrix}$ to itself or alternatively to $\begin{bmatrix} -t \\ 0 \end{bmatrix}$, since these define the same solution to (5.1). The initial group consists of matrices with determinant $\pm 1 \pmod N$, but the orbit structure is the same for $\text{GL}_2(\mathbb{Z}/N)$ and we will make the replacement $G = \text{GL}_2$ to avoid enforcing the determinant condition when we determine $\text{Stab}(t)$ and to simplify the expression for $\#G$. As a base case, $|\text{GL}_2(\mathbb{Z}/q\mathbb{Z})| = (q^2 - 1)(q^2 - q)$ because there are $q^2 - 1$ non-zero choices for the first column, and then $q^2 - q$ vectors not equal to a multiple of the first column. To pass to higher powers of q , we use a version of Hensel's Lemma to lift the matrices from $\text{GL}_2(\mathbb{Z}/q\mathbb{Z})$. Write the matrix whose invertibility is to be determined as $A + qA' + q^2A'' + \dots + q^{e-1}A^{(e-1)}$, where each of the matrices A, A', \dots has entries in $\mathbb{Z}/q\mathbb{Z}$. The condition for another such matrix $B + qB' + q^2B'' + \dots$ to be its inverse is that

$$\begin{aligned} I &= (A + qA' + \dots)(B + qB' + \dots) \\ &= AB + q(A'B + AB') + q^2(A'B' + A''B + AB'') + \dots \end{aligned}$$

Thus we must first of all have $AB = I$, so that A is in $\text{GL}_2(\mathbb{F}_q)$. Then we must have $A'B + AB' = 0$, which can be arranged for any choice of B' by taking $A' = -AB'B^{-1} = -AB'A$. Thus we have q^4 choices at this stage. Then we must have $A'B' + A''B + AB'' = 0$, which determines $A'' = -(AB'' + A'B')B^{-1}$ given any of q^4 choices for B'' . Continuing in this way, we find that

$$|\text{GL}_2(\mathbb{Z}/q^e\mathbb{Z})| = |\text{GL}_2(\mathbb{F}_q)|q^{4(e-1)}.$$

For $N = \prod_{q^e|N} q^e$, it follows that

$$(5.16) \quad |\text{GL}_2(\mathbb{Z}/N\mathbb{Z})| = \prod_{q^e|N} (q^2 - 1)(q^2 - q)q^{4(e-1)}.$$

To determine the stabilizer, we suppose that

$$(5.17) \quad A \begin{bmatrix} t \\ 0 \end{bmatrix} \equiv \pm \begin{bmatrix} t \\ 0 \end{bmatrix} \pmod N.$$

The same congruence holds modulo each prime power q^e (with the same choice of \pm), or equivalently

$$(5.18) \quad A \begin{bmatrix} 1 \\ 0 \end{bmatrix} \equiv \begin{bmatrix} \pm 1 \\ 0 \end{bmatrix} \pmod{q^{e-f}}$$

where f is the highest power of q dividing t . If $e = f$, then $t = 0$ and there is no constraint on A . If $f(q) < e_q$, then the first column of A is constrained. To ensure invertibility, the second diagonal entry must be a unit modulo q^{e-f} . When we lift A to \mathbb{Z}/q^e , it must take the form

$$(5.19) \quad A \in \begin{bmatrix} \pm 1 & \mathbb{Z}/q^{e-f} \\ 0 & \mathbb{Z}/q^{e-f \times} \end{bmatrix} + q^{e-f} M_{2 \times 2}(\mathbb{Z}/q) + \dots + q^{e-1} M_{2 \times 2}(\mathbb{Z}/q).$$

There is only one choice for the first column, since the sign \pm is fixed and common to all factors q^e . There are $\phi(q^{e-f})$ choices for the second diagonal entry, q^{e-f} choices for the other entry in the second column, and q^{4f} ways to lift. It follows that, in the action on $\mathbb{Z}/N \times \mathbb{Z}/N$, the stabilizer has size

$$(5.20) \quad \#\text{Stab}(t) = \prod_{q:f(q)=e_q} |\text{GL}_2(\mathbb{Z}/q^e)| \times \prod_{f(q)<e_q} \phi(q^{e-f})q^{e-f}q^{4f}.$$

In the action on the Cayley cubic, the stabilizer is usually twice this size because $(-t, 0)$ represents the same solution as $(t, 0)$. The exceptional cases are $t = 0$ and $t = (p^2 - 1)/2$, since then $t = -t$. Note that

$$\frac{|\text{GL}_2(\mathbb{Z}/q^e)|}{\phi(q^{e-f})q^{e-f}q^{4f}} = \frac{(q^2 - 1)(q^2 - q)q^{4(e-1)}}{2q^{e-f-1}(q-1)q^{e-3f}} = (q^2 - 1)q^{2e-2f-2}.$$

By the orbit-stabilizer formula, the size of the corresponding orbit is

$$(5.21) \quad \#\text{Orb}(t) = \frac{1}{2} \prod_{f(q)<e_q} (q^2 - 1)q^{2e-2f-2}$$

except without the factor $1/2$ if $t = 0$ or $t = (p^2 - 1)/2$. This completes the proof of Theorem 1.1. \square

5.4. Examples: Sizes of orbits. Note that $p^2 - 1$ is divisible by 8 and by 3 for any odd p , and hence modulo any p there are divisors

$$(5.22) \quad t = 0, \frac{p^2 - 1}{2}, \frac{p^2 - 1}{2}, \frac{p^2 - 1}{3}, \frac{p^2 - 1}{4}, \frac{p^2 - 1}{3}, \frac{p^2 - 1}{6}$$

as well as

$$(5.23) \quad t = \frac{p^2 - 1}{8}, \frac{p^2 - 1}{12}, \frac{p^2 - 1}{24}.$$

We have listed these separately because the divisors in the first list are automatically divisible by $p + 1$ or $p - 1$, while those in the second list may or may not be. We start with

$$g^0 + g^{-0} = 1 + 1 = 2$$

$$g^{\frac{p^2-1}{2}} + g^{-\frac{p^2-1}{2}} = -1 - 1 = -2.$$

We solve for the others using ‘‘bisection’’, that is, substituting previously known values into the relation

$$(5.24) \quad (g^{u/2} + g^{-u/2})^2 = g^u + g^{-u} + 2.$$

For example, $x = g^{(p^2-1)/4} + g^{-(p^2-1)/4}$ solves $x^2 = 0$ so it must be that $x = 0$. We have $(p^2 - 1)/3 \equiv -2(p^2 - 1)/3 \pmod{p^2 - 1}$, so $y = g^{(p^2-1)/3} + g^{-(p^2-1)/3}$ solves

$y^2 = y + 2$. Therefore $y = -1$, since 0 is already spoken for. Then we simply multiply by $-1 = g^{(p^2-1)/2}$ to find

$$g^{\frac{p^2-1}{6}} + g^{-\frac{p^2-1}{6}} = 1$$

or alternatively solve the equation $z^2 = -1 + 2$ using the previous value for -1 . “Bisecting” these values, we find that

$$g^{\frac{p^2-1}{8}} + g^{-\frac{p^2-1}{8}} = \sqrt{2}$$

$$g^{\frac{p^2-1}{12}} + g^{-\frac{p^2-1}{12}} = \sqrt{3}$$

$$g^{\frac{p^2-1}{24}} + g^{-\frac{p^2-1}{24}} = \sqrt{2 + \sqrt{3}}$$

which may or may not lie in \mathbb{F}_p . In any case, we can determine the size of the corresponding orbit in $\mathbb{Z}/N \times \mathbb{Z}/N$ and, if the necessary coordinates lie in \mathbb{F}_p , the Cayley cubic will have an orbit of this size.

The size of the orbit corresponding to a divisor $t = \prod q^{f(q)}$ is

$$(5.25) \quad \frac{1}{2} \prod_{q:f(q) < e_q} (q^2 - 1)q^{2(e-f-1)}$$

or twice that in case $t = 0$ or $t = \frac{p^2-1}{2}$. The easiest case is the orbit of $(u, v) = (0, 0)$, which obviously has size 1. This is the case $t = 0$ and our formula also gives 1, because the product is empty and the factor 1/2 is omitted. This is the orbit of $(x, y, z) = (2, 2, 2)$ in the original coordinates. For $t = (p^2 - 1)/2$, i.e. the orbit of $(-2, 2, -2)$, we again omit the factor 1/2 and find that the orbit has size $(2^2 - 1) \cdot 2^0 = 3$. For $t = (p^2 - 1)/4$, i.e. the orbit of $(0, 2, 0)$, we have $f(2) = e_2 - 2$ so the orbit size is

$$\frac{1}{2}(2^2 - 1) \cdot 2^{2(2-1)} = 6.$$

For $t = (p^2 - 1)/3$, i.e. the orbit of $(-1, 2, -1)$, we have $f(3) = e_3 - 1$ and $f(q) = e_q$ otherwise, so the size of the orbit is

$$\frac{1}{2}(3^2 - 1) \cdot 3^0 = 4.$$

For $t = (p^2 - 1)/6$, i.e. the orbit of $(1, 2, 1)$, we have $f(2) = e_2 - 1$ as well as $f(3) = e_3 - 1$ so the orbit size is

$$\frac{1}{2}(2^2 - 1) \cdot 2^0 \times (3^2 - 1) \cdot 3^0 = 12.$$

This is another way to explain the orbits of size 1, 3, 4, 6, 12 which are present modulo any prime. Recall that double sign changes merge the orbit corresponding to t with the orbit corresponding to $2t$ whenever $f(2) = e_2 - 1$. Thus the orbits of $(p^2 - 1)/2$ and 0 merge, as do the orbits of $(p^2 - 1)/6$ and $(p^2 - 1)/3$. This is why the sizes 4, 6, 16 appear in Table 5.1.

If $p^2 - 1$ is divisible by 16, then we also have the orbit of $(\sqrt{2}, 2, \sqrt{2})$ with $t = (p^2 - 1)/8$. Because $f(2) = e_2 - 3$, the size of this orbit is

$$\frac{1}{2}(2^2 - 1)2^{2(3-1)} = 24.$$

If 3 is a square mod p and we take $t = (p^2 - 1)/12$, then we have an orbit with $f(2) = e_2 - 2$ and $f(3) = e_3 - 1$ and hence of size

$$\frac{1}{2}(2^2 - 1) \cdot 2^{2(2-1)} \times \frac{3^2 - 1}{2} \cdot 3^0 = 48.$$

This occurs when $p \equiv \pm 1 \pmod{12}$, by quadratic reciprocity. If both 2 and 3 are squares, then for $t = (p^2 - 1)/24$ we have $f(2) = e_2 - 3$ and $f(3) = e_3 - 1$, which gives an orbit of size

$$\frac{1}{2}(2^2 - 1) \cdot 2^{2(3-1)} \times (3^2 - 1) = 3 \cdot 2^6 = 192.$$

This component first occurs when $p = 23$. None of these components merge under double sign changes, because $f(2) < e_2 - 1$.

5.5. Examples: Number of orbits. First, consider the case without sign changes. When $p = 29$, we have $p - 1 = 2^2 \cdot 7$ and $p + 1 = 2 \cdot 3 \cdot 5$, so Q^+ consists of 3 and 5 while Q^- consists of 7. The exponent e_2 is 3. The formula (1.8) gives

$$3 \times 2 + 2 \times (2 \times 2) - 2 = 12.$$

For example, when $p = 71$, we have $p - 1 = 2 \cdot 5 \cdot 7$ and $p + 1 = 2^3 \cdot 3^2$, so $Q^+ = \{5, 7\}$ and $Q^- = \{3\}$. The formula gives

$$e_2 \prod_{q \in Q^-} (e_q + 1) + 2 \prod_{q \in Q^+} (e_q + 1) - 2 = 4 \times 3 + 2 \times (2 \times 2) - 2 = 18$$

and indeed there are 18 orbits (of respective sizes 1 3 4 6 12 12 24 24 36 36 48 72 108 192 432 576 1728 1728).

For a first example including sign changes, take $p = 5$. Then Q^- is empty, $Q^+ = \{3\}$, and $e_2 = 3$, so the Cayley cubic splits into $2 + 2 - 1 = 3$ orbits. When $p = 199$, we have $p - 1 = 2^2 \cdot 3^2 \cdot 11$ and $p + 1 = 2^3 \cdot 5^2$, so $Q^0 = \{5\}$, $Q^+ = \{3, 11\}$, and $e_2 = 4$. We have $e_3 = 2 = e_5$ and $e_{11} = 1$, so the number of orbits is $3 \times 3 + 3 \times 2 - 1 = 14$. This explains the number of orbits in Table 5.1.

As a final example, suppose $p = 2l + 1$ is a Sophie Germain prime. Then Q^+ consists only of the prime l and Q^- contains the prime factors of $l + 1$. The number of orbits (including sign changes) is

$$(5.26) \quad \text{ord}_2(l + 1) \prod_{q|(l+1)} (\text{ord}_q(l + 1) + 1).$$

5.6. Finite orbits in characteristic 0. The finite orbits over \mathbb{C} are determined by roots of unity. Whenever the finite field \mathbb{F}_p contains a particular root of unity, the corresponding orbit will appear in the Cayley cubic mod p . Suppose (x, y, z) belongs to a finite orbit of the Cayley cubic over \mathbb{C} . Then some power of the element $\tau_{23} \circ m_3$ must take (x, y, z) to itself. We have

$$\tau_{23} \circ m_3 : (x, y, z) \mapsto (x, y, xy - z) \mapsto (x, xy - z, y).$$

Thus the latter two coordinates are transformed by the matrix $\begin{bmatrix} x & -1 \\ 1 & 0 \end{bmatrix}$, which must have finite order if we are to return to (x, y, z) after finitely many steps. Thus its eigenvalues must be roots of unity. The trace is x and the determinant is 1,

so the eigenvalues are ξ, ξ^{-1} where $x = \xi + \xi^{-1}$. To have $\xi^n = 1$, we must have $x = e^{2\pi ik/n}$ for some integer k . Then

$$x = \xi + \xi^{-1} = 2 \cos \frac{2\pi k}{n}.$$

A similar conclusion for y follows by considering $\tau_{23} \circ m_2$, which acts on (y, z) by $\begin{bmatrix} 0 & 1 \\ -1 & x \end{bmatrix}$. Then using (5.1), we deduce from $x = \xi + \xi^{-1}$ and $y = \eta + \eta^{-1}$ that

$$z = \xi\eta + \frac{1}{\xi\eta}, \text{ or } z = \xi^{-1}\eta + \frac{1}{\xi^{-1}\eta}$$

Thus for (x, y, z) to be part of a finite orbit, it must be of the form

$$2(\cos \alpha, \cos \beta, \cos(\alpha \pm \beta))$$

where α, β are rational multiples of π . Conversely, applying Markoff moves and permutations to such a point will not increase the denominators of the angles α, β , so its orbit will be finite. Dubrovin and Mazzocco do a similar calculation in the context of braid groups in [11], Lemma 1.12.

5.7. Comparison with other levels. All level sets $x^2 + y^2 + z^2 = xyz + k$ have an interpretation by which the same matrices from Proposition 5.1 act. This is given by *Fricke's trace identity*. For 2×2 matrices of determinant 1,

$$(5.27) \quad \text{tr}(A)^2 + \text{tr}(B)^2 + \text{tr}(AB)^2 = \text{tr}(A) \text{tr}(B) \text{tr}(AB) + \text{tr}(ABA^{-1}B^{-1}) + 2.$$

Thus if $\text{tr}(ABA^{-1}B^{-1}) = k - 2$, the vector of traces $(\text{tr } A, \text{tr } B, \text{tr } AB)$ solves the Markoff equation at level k . If $AB = BA$, then $\text{tr}(ABA^{-1}B^{-1}) = 2$ and we have a point on the Cayley cubic. For a matrix of determinant 1, the eigenvalues form a pair ξ, ξ^{-1} inverse to each other, so the trace is

$$(5.28) \quad \text{tr}(A) = \xi + \xi^{-1}$$

and this is the same change of variable from the beginning of this section. If instead $AB = -BA$, then $\text{tr}(ABA^{-1}B^{-1}) = -2$ and we obtain points on the original Markoff surface at level $k = 0$. If $Av = \xi v$, then anticommuting with B forces $A(Bv) = -\xi(Bv)$ so that $-\xi$ is also an eigenvalue. To have $\det(A) = 1$, this implies that $\xi^2 = -1$. Likewise, the eigenvalues of B must be $\pm\sqrt{-1}$. If $p \equiv 1 \pmod{4}$, then the ground field contains a $\sqrt{-1}$ and solutions of this form are helpful in constructing the giant component of Bourgain-Gamburd-Sarnak [5].

Both sides of (5.27) are polynomials in the eight entries of the two matrices, since the determinant being 1 allows one to skip the division by $\det(A), \det(B)$ in computing A^{-1}, B^{-1} . In principle, one can manually verify that they coincide. A more elegant proof is made possible by the Cayley-Hamilton theorem, the cyclic property $\text{tr}(XY) = \text{tr}(YX)$, and the fact that $\text{tr}(X) = \text{tr}(X^{-1})$ for $X \in \text{SL}_2$. See [1], Proposition 4.3, p. 65. The argument is related to why the matrices in Proposition 5.1 give the action of Markoff moves and permutations. For example, $\det A = 1$ implies $B^{-1} = I \text{tr}(B) - B$, by the Cayley-Hamilton theorem (or direct verification). Multiplying by AB and taking traces gives

$$(5.29) \quad \text{tr}(A) = \text{tr}(AB) \text{tr}(B) - \text{tr}(AB^2).$$

Thus $(\text{tr}(A), \text{tr}(B), \text{tr}(AB))$ and $(\text{tr}(AB^2), \text{tr}(B), \text{tr}(AB))$ are related by the Markoff move m_1 . To maintain the convention that the third matrix is the product of the first two, we note that $\text{tr}(B) = \text{tr}(B^{-1})$ and write the move as $A, B \mapsto AB^2, B^{-1}$.

Writing an “abelianized” vector that keeps track of the exponents on A and B but not the order of the product, we may write m_1 as a matrix

$$[m_1] = \begin{bmatrix} 1 & 0 \\ 2 & -1 \end{bmatrix}$$

exactly as in Proposition 5.1. Similar calculations changing the roles of A , B , and AB give the matrices for the other moves m_2 and m_3 . Likewise, the transpositions act by their corresponding matrices. The key difference is that the action is no longer linear.

6. CONCLUSION

We have investigated a family of 3-regular graphs defined from the solutions to (1.1) modulo p for each prime $p \geq 5$. It has already been conjectured that these graphs are connected [3], [5]. On the basis of the data summarized in Figure 3.1, we further conjecture that these graphs are asymptotically Ramanujan for $p \equiv 3 \pmod{4}$. That is, the second largest eigenvalue $\lambda_2(p)$ converges to $2\sqrt{2}$ in this case. For $p \equiv 1 \pmod{4}$, we conjecture that $\lambda_2(p)$ converges to a limit strictly less than 3 and larger than $2\sqrt{2}$, but we do not venture a guess as to its value. It seems that the limit is approximately 2.875... and that there are relatively few eigenvalues above $2\sqrt{2}$. Indeed, Figure 3.3 suggests that the number of exceptional eigenvalues is asymptotic to Cp for a constant $C > 0$. Gathering this data involved computing many eigenvalues instead of only λ_2 , so we considered only an even smaller range of primes. Thus the value of C may not be accurate, but we do conjecture that the exponent p^1 is correct, and in particular that these large eigenvalues comprise a vanishing proportion of the total of roughly p^2 eigenvalues. This means the bulk of the spectrum is supported on $[-2\sqrt{2}, 2\sqrt{2}]$ and we conjectured further that this distribution converges to the Kesten-McKay law. In the meantime, the Kesten-McKay law has now been verified theoretically [8], and Figure 3.2 already shows a good fit even for the small primes $p = 83$ and $p = 89$.

For the level surfaces with $k \neq 0$ in Equation 4.1, connectedness is no longer guaranteed and the more basic question of how many components there are (that is, the multiplicity of $\lambda = 3$) replaces the finer spectral questions above. The most extreme case is when p divides $9k - 4$, and then the components can be understood in terms of a linear action and the Cayley cubic. In general, the component sizes are dictated by arithmetic relations between k and p . The simplest example of this is that there is a component of size 6 whenever k is a square modulo p .

ACKNOWLEDGMENTS

We thank Peter Sarnak for his advice, encouragement, and support over the course of our work. We thank Pedro Henrique Pontes for showing us Lemma 2.2, which provides a simpler way to count solutions than our original proof using Gauss sums. We thank ReMatch, a summer research program at Princeton University, for being a supportive and stimulating research community. Lee was supported by the Bershadsky Family Summer Research Scholars Fund through ReMatch and the Office of Undergraduate Research at Princeton University. de Courcy-Ireland was supported by a PGS D grant from the Natural Sciences and Engineering Research Council of Canada.

REFERENCES

- [1] M Aigner, *Markov's Theorem and 100 Years of the Uniqueness Conjecture: A Mathematical Journey from Irrational Numbers to Perfect Matchings* Springer International Publishing Switzerland (2013)
- [2] N. Alon and V. D. Milman, λ_1 , *isoperimetric inequalities for graphs, and superconcentrators* J. Combin. Theory Ser. B **38**(1):7388, 1985.
- [3] A. Baragar *The Markoff equation and equations of Hurwitz*. Thesis (Ph.D.) Brown University. 1991. MR2686830
- [4] E. Bombieri, *Continued fractions and the Markoff tree*, Expo. Math. **25** (2007) 197-213
- [5] J. Bourgain, A. Gamburd, and P. Sarnak, *Markoff Surfaces and Strong Approximation: 1*, arXiv:1607.01530
- [6] L. Carlitz. *The number of points on certain cubic surfaces over a finite field*. Boll. Un. Mat. Ital. (3) **12** (1957), 1921.
- [7] A. Cerbu, E. Gunther, M. Magee, and L. Peilen. *The cycle structure of a Markoff automorphism over finite fields*. (2016) arXiv:1610.07077
- [8] M. de Courcy-Ireland and M. Magee, *Kesten-McKay law for the Markoff surface mod p* arXiv:1811.00113 [math.NT]
- [9] J. Dodziuk. *Difference equations, isoperimetric inequality and transience of certain random walks*. Trans. AMS **284**(2) (1984) 787794
- [10] B. Dubrovin, *Geometry of 2D Topological Field Theories*, in *Integrable Systems and Quantum Groups*, ed. M. Francaviglia and S. Greco. Lecture Notes in Mathematics no. 1620, Springer (1995)
- [11] B. Dubrovin and M. Mazzocco, *Monodromy of certain Painlevé-VI transcendents and reflection groups*, Invent. math. **141** (2000) 55-147
- [12] H. Kesten, *Symmetric random walks on groups*, Trans. AMS **92** (1959) 336-354
- [13] Yu. V. Linnik, *On the least prime in an arithmetic progression I. The basic theorem*. Rec. Math. (Mat. Sbornik) N.S. **15** (57): 139178. MR 0012111. (1944)
- [14] Yu. V. Linnik, *On the least prime in an arithmetic progression II. The Deuring-Heilbronn phenomenon*. Rec. Math. (Mat. Sbornik) N.S. **15** (57): 347368. MR 0012112. (1944)
- [15] A. Lubotzky, R. Phillips, P. Sarnak, *Ramanujan graphs* Combinatorica **8** (3) (1988) 261-277
- [16] A. Markoff, *Sur les formes quadratiques binaires indéfinies*, Math. Ann. **17** (1880) 379-399
- [17] B. D. McKay, *The expected eigenvalue distribution of a large regular graph*, Linear Algebra and its Applications **40** (1981) 203-216
- [18] D. McCullough and M. Wanderley. *Nielsen equivalence of generating pairs of $SL(2, q)$* . Glasgow Math. J. **55** (2013) 481-509
- [19] C. Meiri and D. Puder, *The Markoff Group of Transformations in Prime and Composite Moduli*, with an appendix by D. Carmon. Duke Math. J. Volume 167, Number 14 (2018), 2679-2720. arXiv:1702.08358 [math.NT]
- [20] M. Mirzakhani, *Counting Mapping Class group orbits on hyperbolic surfaces*. arXiv:1601.03342 [math.GT]
- [21] E. C. Titchmarsh, *A divisor problem*, Rendiconti del Circolo Matematico di Palermo **54**, 414-429 (1930)
- [22] T. Xylouris, *On Linnik's constant*, Acta Arith. **150** (1): 6591. doi:10.4064/aa150-1-4. MR 2825574. (2011)
- [23] D. Zagier, *On the Number of Markoff Numbers Below a Given Bound*, Mathematics of Computation **39**, American Mathematical Society, 1982

DEPARTMENT OF MATHEMATICS, PRINCETON UNIVERSITY, PRINCETON NJ 08544
 E-mail address: mdc4@math.princeton.edu

DEPARTMENT OF MATHEMATICS, PRINCETON UNIVERSITY, PRINCETON NJ 08544
 E-mail address: seungjl@math.princeton.edu