A New Approach and an Exception in the Markov Graph

Devin Vanyo

August 6, 2021

Abstract

Proving the connectivity of the Markov graph mod p remains an open problem. Having provided the problem's necessary background, based largely upon work by Bourgain, Gamburd, and Sarnak, we will present both a new perspective on the problem and an unacknowledged special exception to results previously established.

1 Introduction

All positive integer solutions to the Diophantine equation

$$x_1^2 + x_2^2 + x_3^2 - 3x_1x_2x_3 = 0$$

are known as Markov triples, satisfying the above so-called Markov equation. Any natural number which appears in a Markov triple is known as a Markov number. In an effort to better understand the sequence of Markov numbers, previous work in this area has considered instead the family of congruences

$$x_1^2 + x_2^2 + x_3^2 - 3x_1x_2x_3 \equiv 0 \mod p$$

for all primes p, examining all solutions to these congruences in the corresponding field $\mathbb{F}_p$. In this paper, we will exclusively work with the Markov congruences modulo an arbitrary prime p. Furthermore, any reference to a Markov triple will be a reference to a solution to one such congruence for an understood or arbitrary prime.

A rotation is one of three alike functions:

$$\begin{aligned}
\text{rot}_{x_1}(x_1,x_2,x_3) &= (x_1, x_3, 3x_1x_3 - x_2) \\
\text{rot}_{x_2}(x_1,x_2,x_3) &= (3x_2x_1 - x_3, x_2, x_1) \\
\text{rot}_{x_3}(x_1,x_2,x_3) &= (x_2, 3x_3x_2 - x_1, x_3)
\end{aligned}$$

Rotations preserve Markov triples. Notationally, rot_{x_i} fixes the coordinate in which x_i appears and alters the other two coordinates, producing a new Markov triple. Andrey Markoff showed all Markov triples may be generated by rotations acting on the Markov triple $(1,1,1)$ (see [1] p. 2).

We construct our graph of interest using rotations. The Markov graph $\hat{G}_p$ has vertices being all Markov triples mod p and edges connecting two vertices if a rotation sends one triple to the other. Are $\hat{G}_p$ connected for all primes p? This is the driving question motivating our research.

Rotations invite deeper and more useful tools. For instance, the coordinate x_1 in rot_{x_1} remains fixed, so it need not be represented explicitly in the pre-image or image. Hence, a rotation could be considered as

$$\text{rot}_{x_1}\begin{pmatrix} x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -1 & x \end{pmatrix}\begin{pmatrix} x_2 \\ x_3 \end{pmatrix}$$

where $3x_1 = x$ and clearly

$$\begin{pmatrix} 0 & 1 \\ -1 & x \end{pmatrix} \in \text{SL}_2(\mathbb{F}_p).$$

Thus, each $x \in \mathbb{F}_p$ determines a matrix in $\text{SL}_2(\mathbb{F}_p)$ which, when acting as a linear transformation, may adequately represent a rotation. Call this matrix the rotation matrix of x. Importantly, define the order of any $x \in \mathbb{F}_p$ as the order of the rotation matrix of x in $\text{SL}_2(\mathbb{F}_p)$. Please note: to avoid future confusion, recall that x_1 and x are not the same value (unless they are zero). On the one hand, x_1 appears in tangible Markov triples, whereas on the other, $x = 3x_1$ tends to appear in the context of rotations and rotation matrices so that there is no additional coefficient.

We can classify elements of $\mathbb{F}_p$ into three types via the rotation matrices. Examining the eigenvalues of the rotation matrix produces

$$\lambda = \frac{x \pm \sqrt{x^2-4}}{2}.$$

Hence, we may immediately classify elements $x \in \mathbb{F}_p$ by $(\frac{x^2-4}{p})$, where $(\frac{*}{*})$ is the Legendre symbol:

- if $x \equiv \pm 2 \mod p$, x is *parabolic*,
- if $(\frac{x^2-4}{p}) = 1$, x is *hyperbolic*,
- if $(\frac{x^2-4}{p}) = -1$, x is *elliptic*.

Bourgain, Gamburd, and Sarnak (BGS) demonstrated that each type of element has particular possible orders. Any parabolic element has order either p or $2p$, any hyperbolic element has order dividing $p-1$, and any elliptic element has order dividing $p+1$ [2]. In this paper, we will later demonstrate one previously unknown exception to these BGS rules, which suggests their more careful consideration.

The order of a Markov triple is defined as the maximum order of its three components. A given triple is called maximal if one of its components has maximal order according to its respective type, those maximal orders being either p or $2p$ (to be explained), $p-1$, or $p+1$. These maximal triples, and hence the maximal elements in $\mathbb{F}_p$ which determine them, are important to our study of connectedness. The set of all maximal Markov triples is known as the cage, sometimes denoted $X^*(p)$, and BGS proved that the cage is connected in $\hat{G}_p$, a pivotal result [2].

The BGS approach from this point is to grow this connected component, systematically showing that more and more triples are connected to the cage ([1] organizes this nicely). They break this process into three stages depending upon the order of a given Markov triple. In the "end game,"

they prove that any triple of order $p^{1/2+\delta}$ for a fixed $\delta > 0$ is connected to the cage. This may be done in one step, i.e. one iterative application of a rotation. In the "middle game," BGS aim to show that any triple of order p^{ε} with $\varepsilon > 0$, or in other words of order not independent of p, may be connected to the cage, this time in some finite number of steps, i.e. iterative applications of some finite number of rotations which grow in order. The final stage of growing the cage is the "opening," which is concerned with elements of order uniformly bounded regardless of p at $p < c$ for some constant c.

However, the BGS approach to the "middle game" is via a bounding argument. Their arguments only managed to show connectivity in the "middle game" for truly massive prime moduli, as Fuchs, et al. computationally demonstrated. Therefore, although BGS carved one foothold for the problem, a definitive proof of connectivity for all $\hat{G}_p$ remains open.

This paper's contributions to the problem are twofold. Firstly, we will examine another perspective with which one can view the problem using what we call parabolic cycles. Although yet unyielding, this approach transforms the problem into questions regarding orders in $\mathbb{F}_p$ in a manner both promising and more intuitive to the structure of $\hat{G}_p$. Secondly, as alluded to earlier, in the course of study an exception to the BGS rules concerning the order of elements of $\mathbb{F}_p$ was found. We will demonstrate the proof for this exception and consider its implications.

2 A New Approach to Connectivity

Parabolic elements are particularly interesting because their orders are either p or $2p$ only, and as we shall explain, it makes sense to call both these orders maximal. Hence, any parabolic element is in the cage by definition. Parabolic elements by definition are also particularly easy to find, being just $x \equiv \pm 2 \mod p$. A deeper dive into these elements elicits an alternative perspective to the structure of $\hat{G}_p$.

Firstly, we introduce another piece of terminology used by both BGS and Fuchs, et al. Define $C_i(a)$ to be the set of all Markov triples with ith coordinate equal to a. In practice, we primarily work with $C_1(a)$ since any permutation of a Markov triple is a Markov triple.

One may view $\hat{G}_p$ in terms of these $C_1(a)$ sets. Because rot_a fixes the coordinate a, each orbit of rot_a starting with some triple in $C_1(a)$ must be entirely contained within $C_1(a)$. With this same reasoning, each $C_1(a)$ must be partitioned by disjoint rotational cycles, which "cycle" because rot_a is well-defined, meaning rot_a will follow one, non-intersecting path which does not split, and $C_1(a)$ is finite. Viewing $\langle \text{rot}_a \rangle$ as a group acting on $C_1(a)$, we tend call each of these intra-$C_1(a)$ cycles an *orbit*. In the context of $C_1(a)$, call rot_b where $b \neq a$ and either $(a,b,?) \in C_1(a)$ or $(a,?,b) \in C_1(a)$, a *jumping rotation*, since it "jumps" out of $C_1(a)$ and into some other C_1. See Figure 1.

Recall that $x_1 \equiv \pm\frac{2}{3} \mod p$ are the only parabolic Markov numbers in $\hat{G}_p$. It is a fact [2] that $C_1(\pm\frac{2}{3})$ is nonempty if and only if $p \equiv 1 \mod 4$. Furthermore, explicitly for any $t \in \mathbb{F}_p$,

$$C_1\left(\frac{2}{3}\right) = \left(\frac{2}{3}, t, t \pm \frac{2i}{3}\right)$$
$$C_1\left(-\frac{2}{3}\right) = \left(-\frac{2}{3}, t, -t \pm \frac{2i}{3}\right)$$

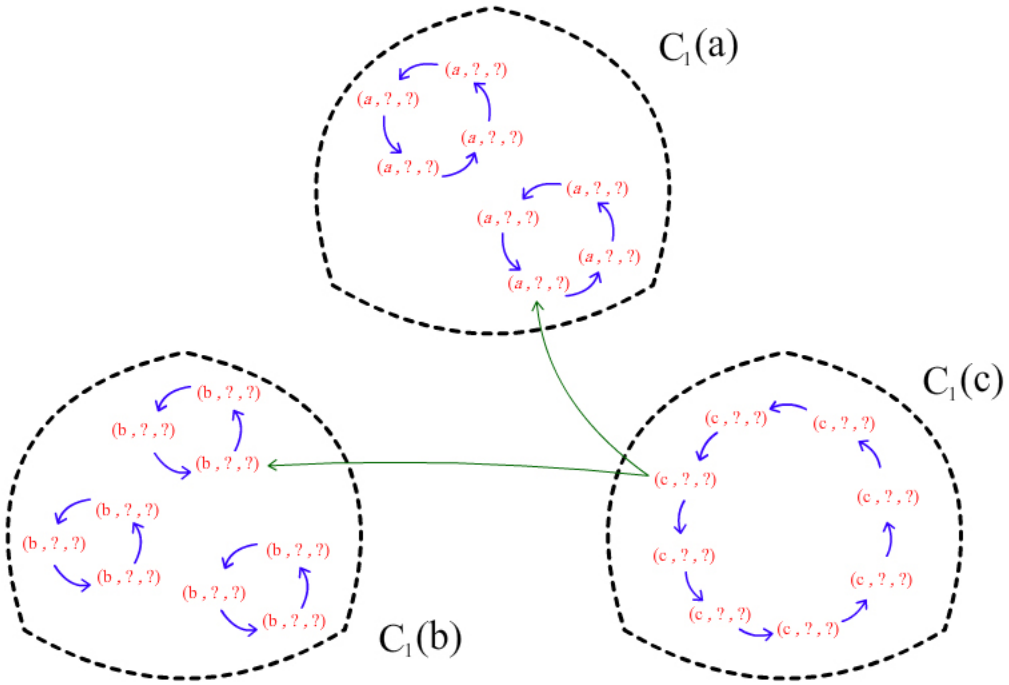

Figure 1: The general structure of $\hat{G}_p$ viewed through a sampling of C_1 sets. Orbits and parabolic cycles (to be discussed) are in blue and jumping rotations are illustrated in green.

and

$$\mathrm{rot}_{2/3}\left(t, t \pm \frac{2i}{3}\right) = \left(t \pm \frac{2i}{3}, t \pm \frac{4i}{3}\right)$$
$$\mathrm{rot}_{-2/3}\left(t, -t \pm \frac{2i}{3}\right) = \left(-t \pm \frac{2i}{3}, -t \mp \frac{4i}{3}\right)$$

where $i^2 \equiv -1 \mod p$ [2]. Note that such an i exists because $p \equiv 1 \mod 4$. The $C_1(\pm\frac{2}{3})$ may be interpreted as disjoint lines. Finally, $x = 2$ has order p because $\mathrm{rot}_{2/3}$ fixes each line in $C_1(\frac{2}{3})$, and $x = -2$ has order $2p$ because $\mathrm{rot}_{-2/3}$ interchanges each line in $C_1(-\frac{2}{3})$ (see [1] p. 10 for an illustration). However, in either case, it is clear that $|C_1(\pm\frac{2}{3})| = 2p$.

Examine when $x_1 = \frac{2}{3}$. Since $\{\frac{2i}{3}, 2(\frac{2i}{3}), \ldots, p(\frac{2i}{3}) = 0\}$ forms a complete residue system mod p, so does $\{t, t+\frac{2i}{3}, t+2(\frac{2i}{3}), \ldots, t+(p-1)(\frac{2i}{3})\}$. Thus, if a triple contains the parabolic Markov number $\frac{2}{3}$, it may be connected via $\mathrm{rot}_{2/3}$ to a parabolic triple containing any coordinate. The same conclusion holds for $x_1 = -\frac{2}{3}$ because $|C_1(-\frac{2}{3})| = 2p$ and 2 has order $2p$, so there is only one orbit, where t as above is arbitrary. This proves the following.

Theorem 2.1. *Let $p \equiv 1 \mod 4$. Then, $C_1(a)$ contains a parabolic triple for every $a \in \mathbb{F}_p$.*

We can be more precise. Remembering all permutations of Markov triples are Markov triples,

Case 1: $\frac{2}{3} = a + \frac{2i}{3}$	$a = \frac{2(1-i)}{3}$	$\left(a, \frac{2}{3}, a+\frac{2i}{3}\right) = \left(a, a+\frac{2i}{3}, \frac{2}{3}\right)$ $\left(a, -\frac{2}{3}, -a-\frac{2i}{3}\right) = \left(a, -a-\frac{2i}{3}, -\frac{2}{3}\right)$
Case 2: $\frac{2}{3} = a - \frac{2i}{3}$	$a = \frac{2(1+i)}{3}$	$\left(a, \frac{2}{3}, a-\frac{2i}{3}\right) = \left(a, a-\frac{2i}{3}, \frac{2}{3}\right)$ $\left(a, -\frac{2}{3}, -a+\frac{2i}{3}\right) = \left(a, -a+\frac{2i}{3}, -\frac{2}{3}\right)$
Case 3: $-\frac{2}{3} = a + \frac{2i}{3}$	$a = \frac{-2(1-i)}{3}$	$\left(a, \frac{2}{3}, a+\frac{2i}{3}\right) = \left(a, -a-\frac{2i}{3}, -\frac{2}{3}\right)$ $\left(a, a+\frac{2i}{3}, \frac{2}{3}\right) = \left(a, -\frac{2}{3}, -a-\frac{2i}{3}\right)$
Case 4: $-\frac{2}{3} = a - \frac{2i}{3}$	$a = \frac{-2(1+i)}{3}$	$\left(a, -\frac{2}{3}, -a+\frac{2i}{3}\right) = \left(a, a-\frac{2i}{3}, \frac{2}{3}\right)$ $\left(a, -a+\frac{2i}{3}, -\frac{2}{3}\right) = \left(a, \frac{2}{3}, a-\frac{2i}{3}\right)$

Table 1: The values of a where parabolic triples are not distinct.

it is clear that when a is not parabolic, $C_1(a)$ nonetheless contains at most 8 parabolic triples:

$$\left(a, \frac{2}{3}, a \pm \frac{2i}{3}\right), \left(a, a \pm \frac{2i}{3}, \frac{2}{3}\right), \left(a, -\frac{2}{3}, -a \pm \frac{2i}{3}\right), \left(a, -a \pm \frac{2i}{3}, -\frac{2}{3}\right)$$

Examining when these solutions might not be distinct, we have cases, as seen in Table 1. One may check that each case is distinct from every other since $p \equiv 1 \mod 4$. This establishes the following result.

Theorem 2.2. *Let $p \equiv 1 \mod 4$ and i be such that $i^2 \equiv -1 \mod p$. If $a = \frac{\pm 2(1 \pm i)}{3}$, there are 6 parabolic triples in $C_1(a)$. For all other $a \in \mathbb{F}_p$, there are 8 parabolic triples in $C_1(a)$.*

It is worth mentioning that this result inspires a slightly sharper previously known corollary.

Corollary 2.3. *If $p \equiv 1 \mod 4$, for every $a \in \mathbb{F}_p$ there exist at least 6 Markov triples containing a.*

From the above discussion, it is clear that parabolic triples are at least evenly distributed throughout the collection of sets $C_1(a)$. We will now examine the rot_a orbits within each $C_1(a)$ which contain parabolic triples. Call these parabolic cycles. Recall that every parabolic triple is maximal and connected to the cage, which implies that every triple in a parabolic cycle is also connected to the cage.

We neglect structurally determining the exact number of parabolic cycles per each $C_1(a)$. The data does not indicate an easily discernible pattern, and to do so is not necessary. Given the structure of rotations, which swap one coordinate and alter the other, it is cursory at least to say each $C_1(a)$ may contain an absolute maximum of 4 parabolic cycles.

Instead, we computed the number of parabolic cycles in total for various prime moduli. We also computed the total number of orbits (contained within the $C_1(a)$'s), of which parabolic cycles are

a subset, as well as the relative numbers of triples contained immediately within parabolic cycles. The results of these computations are summarized in Table 2 and Figure 2 on the next page. It is noteworthy that $C_1(0)$ was not factored into the computations in Table 2, since it has a special size and uniquely predictable behavior; more on this below. In practice simply apply a jumping rotation to any initial element in $C_1(0)$.

The motivation behind examining parabolic cycles is to connect every triple (x_1, x_2, x_3) to some parabolic cycle, if not in $C_1(x_1)$ then in some other $C_1(a)$ after some number of jumping rotations. The statistics in Table 2 indicate that even as parabolic cycles diminish in abundance, the probability that an random Markov triple will be in a parabolic cycle, and thus connected to the cage, remains surprisingly high.

One factor in explaining this fact is that for maximal hyperbolic and elliptic $a \in \mathbb{F}_p$, all of $C_1(a)$ is one orbit, which is a parabolic cycle, since $C_1(a)$ contains some parabolic triples. For a of higher order in general, there are fewer orbits in $C_1(a)$, each having a larger size, so more elements will fall within some parabolic cycle. A greater ratio of elements in $\mathbb{F}_p$ with a relatively large order will indicate a larger ratio of elements within some parabolic cycle, and hence a larger initial connected component.

If relatively few triples are contained within parabolic cycles as p grows, then more jumping rotations should be required before one stumbles into a parabolic cycle within some $C_1(a)$, assuming $\hat{G}_p$ is connected. Therefore, the efficiency of this approach to understanding the structure of $\hat{G}_p$ depends largely upon how the percentages in Table 2 extrapolate out for larger prime moduli or, equivalently, upon the distribution of orders for elements of $\mathbb{F}_p$ given larger p. One limitation to this perspective is it only applies to $p \equiv 1 \mod 4$, for which there exist any parabolic elements. This is the extent of new light this paper has to offer on this alternative perspective.

A difficult next step to developing this approach might also be examining how parabolic cycles relate to non-parabolic cycles. Specifically, one might study jumping rotations off parabolic cycles. How frequently do they land in non-parabolic orbits? As noted above, this is closely related to the distribution of orders for elements of $\mathbb{F}_p$ given larger p.

3 A Note on $C_1(0)$

A minor discovery was made while exploring the above perspective. It relates to the structure of $C_1(0)$, which is unique to all other elements. As a preface, note when $x_1 = 0$, still $x = 3x_1 = 0$.

Lemma 3.1. *Let $p > 2$. If $p \equiv 1 \mod 4$, then the element 0 is hyperbolic. Otherwise, when $p \equiv 3 \mod 4$, 0 is elliptic.*

Proof. $(\frac{x^2-4}{p}) = (\frac{-4}{p}) = 1$ if and only if $(\frac{-1}{p}) = 1$, since $4 = 2^2$ is always a quadratic residue. Equivalently, $p \equiv 1 \mod 4$. □

Previous work by both references implied that if 0 is hyperbolic, $|C_1(0)|$ should be of size $p-1$ like all other hyperbolic elements, and if 0 is elliptic, similarly $|C_1(0)| = p+1$. However, this is not the case. This is the discovered exception to the rule implicitly given in [2] and explicitly explained in [1] that the size of $|C_1(a)|$ may be determined generally by only the element type of a. We reserve more discussion of this following the proof.

Theorem 3.2. *Suppose we exclude the trivial Markov solution $(0,0,0)$. Then,*

Prime Modulus	Parabolic Cycles/Orbits	Triples Within Some Parabolic Cycle
5	100%	100%
13	85%	91.667%
17	82.759%	89.583%
29	62.5%	77.857%
37	70.513%	86.111%
41	61.345%	78.429%
53	61.364%	80.342%
61	58.333%	82%
73	55.385%	81.404%
89	45.161%	72.446%
97	47.923%	77.220%
101	49.102%	75.369%
109	42.857%	72.464%
113	47.342%	75.273%
137	50.348%	76.771%

Table 2: "Parabolic Cycles/Orbits" indicates the percentage of orbits within the $C_1(a)$'s which are parabolic cycles. "Triples Within Some Parabolic Cycle" indicates the percentage of Markov triples which lie immediately in some parabolic cycle.

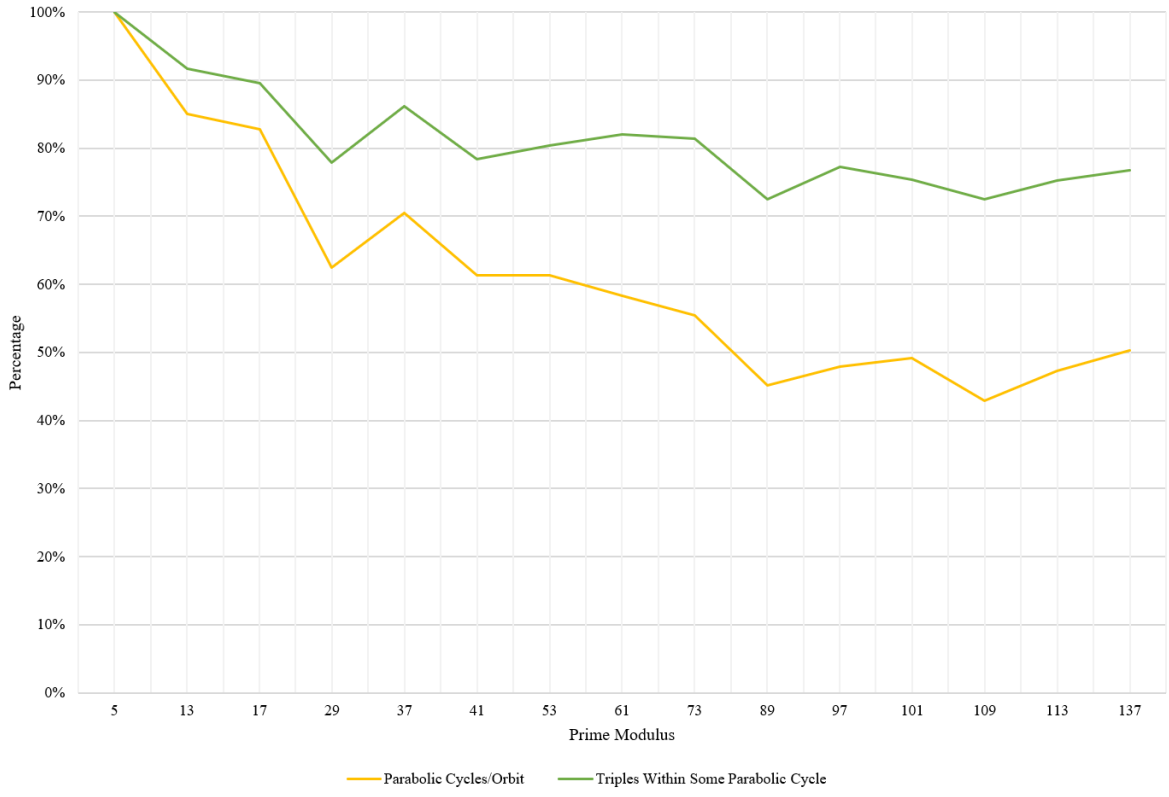

Figure 2: Data from Table 2.

1. *if* $p \equiv 1 \mod 4$, $|C_1(0)| = 2(p-1)$, *and*

2. *if* $p \equiv 3 \mod 4$, $|C_1(0)| = 0$.

Proof. Let $x_1 = 0$. Then, the Markov equation becomes $x_2^2 + x_3^2 \equiv 0 \mod p$, or

$$x_2^2 \equiv -x_3^2 \mod p.$$

The number of solutions to $x_2^2 \equiv -x_3^2 \mod p$, and thus the size of $C_1(0)$, is dependent upon $(\frac{-1}{p})$. This is equivalent to asking the value of $p \mod 4$.

If $p \equiv 3 \mod 4$, then $(\frac{-1}{p}) = -1$. In this case there are no solutions. Hence, excluding the trivial solution $(0,0,0)$, $|C_1(0)| = 0$.

Suppose instead $p \equiv 1 \mod 4$. Then, the above equation has a solution for any $x_3 \in \mathbb{F}_p^*$, since a residue times a residue is always a residue. There are $p-1 = |\mathbb{F}_p^*|$ choices for x_3 and 2 choices for x_2, namely $\pm x_2$. Therefore, there are $2(p-1)$ total solutions, excluding $(0,0,0)$, when $p \equiv 1 \mod p$. □

This exception is theoretically significant. Fix p. For any $a \in \mathbb{F}_p$, the orbits of rot_a partition $C_1(a)$, rot_a being a group action on $C_1(a)$, and each has equal size, namely the order of a. If the order of a equals $|C_1(a)|$, a is defined to be of maximal order, which is significant because it means $C_1(a)$ contains only one large (parabolic) cycle when applying rot_a. The rigid definitions of maximality according to type, defined by [2] and sketched in justification by [1], thus lose potency if $|C_1(a)|$ is not constant for all a of a given type. For parabolic a, that $|C_1(a)|$ is constant has been explicitly and convincingly shown in abstraction [1]. This exception, however, demands a similarly more thorough proof of $|C_1(a)|$ for hyperbolic and elliptic a. Although the complexity and depth of the current sketches [1] require a more sensitive theoretical treatment than could be provided by this present author, experimental data suggests 0 to be the only exception.

Therefore, the above result demands a more thorough and explicit understudy of the fundamental concepts at play within the Markov graph literature. Otherwise, maximality is not currently rigorously defined, at least as appropriately intended.

References

[1] A cryptographic hash function from markoff triples. *Mathematical Cryptology*, 2021.

[2] Jean Bourgain, Alexander Gamburd, and Peter Sarnak. Markoff surfaces and strong approximation: 1. 2021.