Proposition 3.15. *A link diagram D has a non-trivial two coloring iff D has at least two components.*

Proof. Draw the four allowable colorings near a crossing and note that the two underpassing arcs receive the same label. Thus in a coloring modulo 2 every component receives the trivial coloring. This proves the forward direction.

If D has more than one component then color one of them with a zero and all of the others with a one. □

Definition 3.16. A link $L \subset \mathbb{R}^3$ is *split* if L is disjoint from the yz plane but meets both half-spaces $\{(x, y, z) \mid x > 0\}$ and $\{(x, y, z) \mid x < 0\}$. The diagram of a split link is called a split diagram.

A link is *splittable* if it is isotopic to a split link. Again, we use the same language for diagrams.

Proposition 3.17. *If D is splittable then D has non-trivial colorings in every modulus greater than one.*

Proof. Fix n. Perform Reidemeister moves to split the diagram. Now color the components in the positive half-space with a zero. Color the components in the negative half-space with a one. We are now done, by Theorem 3.11. □

Corollary 3.18. *The Borromean rings are not splittable.*

Proof. The Borromean rings cannot be non-trivially colored modulo three. □

Exercise 3.19. Show that the standard diagram of the Borromean rings cannot be 3–colored.

Figure 16: A four coloring of the Borromean rings.

Exercise 3.20. As in the examples above, find non-trivial colorings of the Whitehead link. Careful: you cannot divide by two in the ring $\mathbb{Z}_{2m}$.

4 Shadows and checkerboards

As we have seen above, a coloring of a diagram may be thought of as a collection of variables (one for each arc of the diagram). Every crossing now gives a relation among these variables. We'll call this the *crossing equation*. The examples you computed above all suggest that there is a dependence among the crossing equations. This is indeed the case: the proof is somewhat lengthy!

Definition 4.1. Suppose that D is a diagram. If we forget the crossing information we are left with a *shadow*: a smoothly immersed loop in the plane where are self-intersections are transverse, and without triple points. The *edges* of the shadow are the arcs remaining when we remove the self-intersection points.

We say that a diagram is *connected* if its shadow is connected.

Definition 4.2. Suppose that S is a shadow. A component of $\mathbb{R}^2 \setminus S$ is called a *region* of S. Two regions are *adjacent* if they intersect along one or more edges of S.

We will need the following observation:

Lemma 4.3. *If $x, y \in R$ then x and y can be connected by a polygonal path.* □

Exercise 4.4. Draw a shadow where at least two of the regions are not disks.

Exercise 4.5. Suppose that the diagram D has shadow S. Suppose that S has a at least two regions which are not disks. Show that D is splittable.

A *checkerboard coloring* of a shadow is a coloring of every region of S to be black or white so that white regions are only adjacent to black regions and, conversely, black regions are only adjacent to white ones.

Lemma 4.6. *Any shadow S admits a checkerboard coloring.*

The proof that follows is somewhat technical. Can you find a simpler proof? Perhaps if the shadow is very simple, such as a union of circles, say?

Proof of Lemma 4.6. Isotope S slightly to put S in general position. For every region R of S we define a *parity* $e(R)$ as follows: pick a point $x \in R$ and pick a ray L emanating from x, which is transverse to S. Now define $e(R)$ to be the parity of $L \cap S$.

We must check that $e(R)$ is well-defined. First suppose that L' is another ray emanating from $x \in R$, again transverse to S. Then there is a family $\{L_t \mid t \in [0, 1]\}$ so that each L_t is a ray and $L_0 = L$, $L_1 = L'$. For only finitely many t the ray L_t is not transverse. Then for small ϵ we find that $L_{x-\epsilon}$ and $L_{x+\epsilon}$ are transverse and differ by passing a crossing or tangency of S. Thus

$$|L_{x-\epsilon} \cap S| = |L_{x+\epsilon} \cap S| \mod 2.$$

Now suppose that we choose $y \in R$ instead of x, when defining the parity. Since R is path-connected we may connect x and y by a polygonal path with vertices $x_0 = x, x_1, \ldots, x_n = y$. We may assume that $n \geq 2$ and move the points x_i (for $0 < i < n$) slightly so that the lines through $\{x_{i-1}, x_i\}$ and through $\{x_i, x_{i+1}\}$ are transverse to S. Since the line segment $[x_i, x_{i+1}]$ is disjoint from S it follows that x_i and x_{i+1} both yield the same parity. This proves that $e(R)$ is well-defined.

Finally, suppose that R and R' are adjacent regions and $x \in R$ and $x' \in R'$ are points. Then we may connect x and x' by a polygonal path which meets S exactly once. Similar to the argument above, $e(R)$ and $e(R')$ have are opposite parities. □

Definition 4.7. A shadow S is *reduced* if every self-intersection is adjacent to four distinct regions.

Proposition 4.8. *Suppose that L is a link. Then (perhaps after an isotopy) L has a reduced diagram.*

To prove Proposition 4.8 we will need the *Jordan curve theorem*:

Theorem 4.9 (Schönflies). *Suppose that $C \in \mathbb{R}^2$ is a loop without self-intersections. Then there is a homeomorphism $h\colon \mathbb{R}^2 \to \mathbb{R}^2$ throwing C onto the unit circle S^1.* □

This can be restated as: every *simple* loop in the plane bounds a disk. The Jordan curve theorem is a basic tool in topology and is "obviously true." That doesn't mean that it is easy to prove, however...

Proof of Proposition 4.8. Suppose that D is a nonreduced diagram for L. Then there is a crossing of $c \in D$ which meets one region twice. Since R is path-connected (Lemma 4.3) there is a simple loop P in the plane so that $P \cap S = c$ and $P \smallsetminus c \subset R$. That is, P is essentially contained in R and separates D into two pieces D' and D'', where D' is the piece contained in the disk bounded by P (Theorem 4.9). There is an isotopy of L which flips D' over, leaving D'' unchanged. This gives a new diagram, isotopic to the old one, with one fewer crossing. Repeat this process until arriving at a reduced diagram. □

Definition 4.10. For every diagram there is a *dual graph* G_D embedded in the plane. To construct G_D let S be the shadow. For every region of S there is a vertex of G_S. For every edge of S there is an dual edge of G_D connecting the two adjacent regions.

When D is reduced and connected the dual graph is a decomposition of the plane into quadrilaterals (and one punctured quadrilateral, where the puncture is the point at infinity).

Now suppose that D is a connected and reduced diagram. Choose a checkerboard coloring of the diagram. Orient the edges of G_D so that all edges point from white to black in the checkerboard coloring. Finally orient all quadrilaterals of $\mathbb{R}^2 \smallsetminus G_D$ using the counterclockwise convention.

Now, the crossings of D and the quadrilaterals of G_D are in one-to-one correspondence. Each quadrilateral gives an equation as follows: add the variables corresponding to the sides of the quadrilateral with a plus sign if the orientation of the quadrilateral and the side agree and with a minus sign if the orientation of the quadrilateral and the side disagree.

So for every crossing we have either the equation

$$-a + c - b + c = 0$$

or

$$+a - c + b - c = 0$$

and we notice that the first is the crossing equation and the second is its negative.

Proposition 4.11. *With the above choice of signs, the sum of the crossing equations is zero.*

Proof. Quadrilaterals adjacent across an edge each contribute a copy of the corresponding variable, but with opposite sign. □

To sum up: Lemma 3.10 shows that one of the variables may be ignored (set to be zero). Proposition 4.11 similarly shows that one of the crossing equations may be ignored. We will use these facts to compute the *determinant* of a diagram — this will be the order of the *coloring group*.

5 Determinants of diagrams

Suppose that $D = \sqcup C_i$ is a diagram. We say that C_i is a *overcrossing component* if

- every crossing between C_i and the rest of D has C_i above and
- C_i does not cross itself.

It follows that the diagram is splittable and that C_i is an unknot. (Convince yourself that this is correct!)

For the rest of this section we will assume that all diagrams are connected, reduced, and have no overcrossing component.

Lemma 5.1. *With the assumption above, the number of arcs of D equals the number of crossings of D.*

Proof. Orient the diagram. Since there are no overcrossing components, every arc $\alpha \subset D$ meets a crossing at each end (however, these crossings may be equal. Eg the Hopf link). Associate α with the crossing at the head of α, according to the orientation. This is the desired bijection. □

Now define A_+ to be the matrix of crossing equations. We have one row for every crossing and one column for every arc of D. It follows that A_+ is a square $|D| \times |D|$ matrix. By Proposition 4.11 each column of the matrix sums to zero. By the definition of the crossing equations each row sums to zero. (This fact is essentially the same as Lemma 3.10.)

Exercise 5.2. Compute A_+ for the twist knots, T_k.

We now define the matrix A by deleting any one column and any one row from A_+. Let $d = |D| - 1$. To find a coloring modulo n is now equivalent to solving the linear equation $Ax = nb$ where

$$x = (x_1, \ldots, x_d), \quad b = (b_1, \ldots, b_d)$$

are column vectors in $\mathbb{Z}^d$.

Case (i) Suppose that $\det(A) = 0$. Then there is a non-zero solution over $\mathbb{Q}$ to the equation $A \cdot y = 0$. If $y_i = p_i/q_i$ then set $Q = \prod q_i$. Taking $z = Qy$ gives a solution of $A \cdot z = 0$ over $\mathbb{Z}$. Take $z' = z/\gcd(z_i)$. Then $z' \neq 0 \mod n$ for any $n > 1$. Setting $z'_0 = 0$ gives a non-trivial coloring modulo n for any $n > 1$.

To sum up: when $\det(A) = 0$ then the diagram D can be non-trivially colored in every modulus.

Case (ii) Suppose that $\det(A) \neq 0$. Then Cramer's rule tells us that there is a unique solution over $\mathbb{Q}$, namely

$$x_k = \frac{|A_1 \ldots nb \ldots A_d|}{|A_1 \ldots A_k \ldots A_d|},$$

where A_k is the k^{th} column of A and nb replaces A_k in the numerator. So we find that

$$x_k = \frac{n \cdot |A_1 \ldots b \ldots A_d|}{\det(A)}.$$

Thus when $\det(A)|n$ we have solutions over $\mathbb{Z}$. As a special case, take $n = |\det(A)|$ and then

$$x_k = |A_1 \ldots b \ldots A_d|$$

for $1 \leq k \leq d$. Setting $x_0 = 0$ now solves the coloring equations. However this solution is not completely satifactory. It is conceivable that, for any choice of b, the coloring is trivial modulo n.

To sum up: when $|\det(A)| = 1$ then all colorings, in every modulus, are trivial. When $|\det(A)| > 1$ then perhaps we may find non-trivial solutions modulo $|\det(A)|$ using Cramer's rule and perhaps we may not; we are assured, however, that there are some moduli where only trivial solutions are possible.

Exercise 5.3. Show that the absolute value $|\det(A)|$ is independent of the choice of row and column deleted from A^+.

Exercise 5.4. [Harder] Show that the Smith normal form of A is independent of the choice of row and column deleted from A^+. (See Lemma 6.2 below, also.)

Definition 5.5. The *determinant* of a link L is $\det(L) = |\det(A)|$ where A is the matrix defined above. We shall see later in the course that this is an invariant of the isotopy class of L.

Exercise 5.6. In the Case (ii) above the choice of b is not specified. Setting $b = (1, 0, \ldots, 0)$ we find that $x_k = (-1)^{k+1}\text{Minor}_{1,k}(A)$ and also that $\det(A) = A^1 \cdot x$ where A^1 is the first row of A. Use this to find a coloring modulo $\det(K)$ of the knot 6_3.

Exercise 5.7. Check that if the diagram is *alternating* (every overcrossing arc goes over exactly one crossing) then the variables may be ordered so that the matrix A^+ has twos along the diagonal.

Exercise 5.8. Compute the determinant of the 5_1 knot (the cinefoil) and the twist knots. Notice that the first two twist knots are the trefoil and the figure eight.

6 Smith normal form and the coloring group

6.1 Smith normal form and Cramer's rule

We now find a method, more general than the one involving Cramer's rule, for finding non-trivial colorings.

Theorem 6.1. *The link L has a non-trivial coloring modulo $n > 1$ if and only if* $\gcd(n, \det(L)) > 1$.

We need to recall Smith normal form:

Lemma 6.2. *Given a $d \times d$ integer matrix A, as above, there are matrices R, C, B so that R, C are isomorphisms of $\mathbb{Z}^d$, B is diagonal, and $B = RAC$.*

It follows that $\det(A) = \det(B)$.

Proof of Lemma 6.2. Use row (and column) operations of the form

Type 1. $r_i \mapsto r_j$ and $r_j \mapsto -r_i$, and

Type 2. $r_i \mapsto r_i + ar_j$ for $i \neq j$ and $a \in \mathbb{Z}$.

Notice that operating on rows is equivalent to multiplying A on the left by a signed permutation matrix (1) or an elementary matrix (2). Operating on columns is equivalent to multipying on the right. Also, such matrices are isomorphisms of $\mathbb{Z}^d$.

Exercise 6.3. Show that type 1 operations can be obtained from three operations of type 2.

Let a_{ij} be a non-zero entry of A of smallest absolute value. Using type 1 operations move a_{ij} into the upper-left corner of A.

$$A \mapsto \begin{bmatrix} a_{ij} & * \\ * & * \end{bmatrix}$$

Now use type 2 operations to add or subtract row one (column one) from the other rows (columns) to decrease the absolute value of all other entries in the first row and column. Repeat the above steps until all entries in the first row and column, except for perhaps the upper-left entry, are zero. Note that this process must finish, as the absolute value of the upper-left entry only decreases.