Edge flipping in the complete graph

Steve Butler[1,*]

Dept. of Mathematics, Iowa State University, Ames, IA 50011, USA

Fan Chung

Dept. of Mathematics and Dept. of Computer Science and Engineering, UC San Diego, La Jolla, CA 92093, USA

Jay Cummings

Dept. of Mathematics, UC San Diego, La Jolla, CA 92093, USA

Ron Graham

Dept. of Computer Science and Engineering, UC San Diego, La Jolla, CA 92093, USA

Abstract

We consider the following random process on the complete graph: repeatedly draw edges (with replacement) and with probability p assign the vertices of the edge blue and with probability $1-p$ assign the vertices of the edge red. This is a random walk on a state space of red/blue colorings of the complete graph and so has a stationary distribution. We derive this stationary distribution as well as answer some related questions.

Keywords: edge flipping, stationary distribution, asymptotic analysis

1. Introduction

In a previous paper, Chung and Graham [2] considered the following "edge flipping" process on a connected graph G (originally suggested to them by Persi Diaconis, see also [1]). Beginning with the graph in some arbitrary coloring, repeatedly select an edge (with replacement) at random and color both of its vertices blue with probability p and red with probability $q := 1 - p$. This creates a random walk on all possible red/blue colorings of the graph and has a unique stationary distribution. Chung and Graham were able to determine the

*Corresponding author

Email addresses: `butler@iastate.edu` (Steve Butler), `fan@ucsd.edu` (Fan Chung), `jjcummings@ucsd.edu` (Jay Cummings), `graham@ucsd.edu` (Ron Graham)

[1]Partially supported by an NSA Young Investigator Grant.

stationary distributions for paths and cycles as well as obtain some asymptotic results related to these graphs.

We remark that finding the stationary distribution is difficult since the state space of this random walk generally is exponential in the number of vertices in the graph. As a result, direct analysis can only be carried out for small graphs. The goal of this note is to show how to find the stationary distribution of this process for the complete graph K_n. In Figure 1 we illustrate the state space of this process for K_3 (where for simplicity we use symmetry to reduce from eight possible colorings down to the four (light) blue/(dark) red colorings shown).

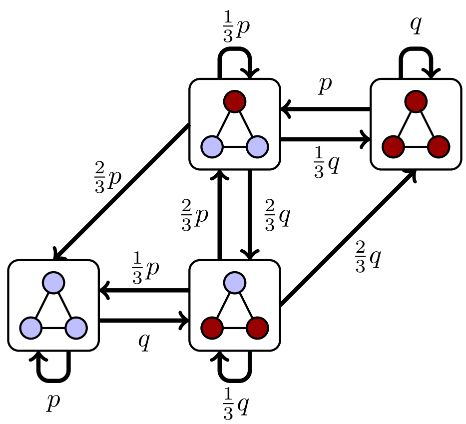

Figure 1: The edge flipping process on K_3

Using the figure it is a straightforward exercise to verify that the stationary distribution for the edge flipping process on K_3 satisfies the following:

$$\begin{aligned}\mathbb{P}(3\text{ blue; }0\text{ red}) &= p^2\\ \mathbb{P}(2\text{ blue; }1\text{ red}) &= pq\\ \mathbb{P}(1\text{ blue; }2\text{ red}) &= pq\\ \mathbb{P}(0\text{ blue; }3\text{ red}) &= q^2\end{aligned}$$

The goal of this note is to establish the following general result.

Theorem 1. *Let $b + r = n$. The stationary distribution for the edge flipping process on the complete graph K_n satisfies the following:*

$$\mathbb{P}(b\text{ blue; }r\text{ red}) = \frac{2^n p^b q^r}{\binom{2n-2}{n-2}} \sum_j \sum_k \binom{n-1}{b-2j,\, r-2k,\, j+k-1,\, j,\, k} (4p)^{-j} (4q)^{-k}.$$

To find the probability that b specific vertices are blue and the remaining r are red, divide by $\binom{n}{b} = \binom{n}{r}$.

Here we use $\binom{n}{i_1,i_2,\ldots,i_k}$ with $n = i_1+i_2+\cdots+i_k$ to represent the multinomial coefficient, i.e., we have

$$\binom{n}{i_1,i_2,\ldots,i_k} = \frac{n!}{i_1!i_2!\cdots i_k!},$$

with the convention that if any of the terms are negative, then the value is 0.

Since flipping an edge can only change the color of at most two vertices, a direct proof of Theorem 1 can be carried out by verifying

$$\sum_b \mathbb{P}(b \text{ blue}; \, n-b \text{ red}) = 1$$

and that for each $b + r = n$

$$\begin{aligned}\binom{n}{2}\mathbb{P}(b \text{ blue}; \, r \text{ red}) &= p\binom{r+2}{2}\mathbb{P}(b-2 \text{ blue}; \, r+2 \text{ red}) \\ + p(b-1)(r+1)\mathbb{P}(b-1 \text{ blue}; \, r+1 \text{ red}) &+ \left(p\binom{b}{2} + q\binom{r}{2}\right)\mathbb{P}(b \text{ blue}; \, r \text{ red}) \\ + q(b+1)(r-1)\mathbb{P}(b+1 \text{ blue}; \, r-1 \text{ red}) &+ q\binom{b+2}{2}\mathbb{P}(b+2 \text{ blue}; \, r-2 \text{ red}).\end{aligned}$$

We will take a different approach which will give more insight into the process and introduce more tools that might be useful for working with other graphs.

We will proceed by first going into more details on the edge flipping process in Section 2. Then in Section 3 we will focus on establishing the probability that k specified labeled vertices are blue and use this to find the probability that all the vertices are blue. In Section 4 we use what is known about the all-blue probability to establish Theorem 1. In Section 5 we look at some of the asymptotics as n goes to ∞. Finally in Section 6 we give some concluding remarks including further directions of research.

Throughout the paper we will always use p to denote the probability a selected edge will have its vertices be colored blue, and q to denote the probability a selected edge will have its vertices be colored red.

2. Equivalent interpretations to edge flipping

The goal of this section is to take a closer look at the edge flipping process, and in particular look at different ways to analyze what is going on. We will assume that the graphs we are working with are connected. We start with the original interpretation of edge flipping.

> **Random walk interpretation of edge flipping:** Take a graph G with edges $\{e_1, e_2, \ldots, e_m\}$, and some initial random red/blue vertex coloring of G. Randomly choose edges (with replacement) and change the color of the vertices of the selected edge to blue with probability p and to red with probability q. Continue this process indefinitely.

Our first observation is that the edge flipping process is memoryless, i.e., a vertex v is only affected by the last edge drawn that was incident to v. This suggests that we should focus only on the last time that a particular edge was selected, and leads us to the following interpretation.

> **Reduced interpretation of edge flipping:** Take a graph G and a deck of $|E(G)|$ cards, i.e., one card for each edge. Randomly shuffle the deck and then deal out the cards one at a time. For each card, change the vertices of the indicated edge to blue with probability p and to red with probability q.

Note that since each edge of G will be considered, the initial red/blue vertex coloring of G is not important.

Proposition 1. *Given a fixed vertex coloring c of a graph G, the probability of being at c in the stationary distribution for the random walk interpretation is the same as the probability of realizing c by the reduced interpretation.*

Proof. The probability of being at c in the random walk can be found by looking at the probability that we are at c after N steps as $N \to \infty$. For large N almost every list of N edges will contain each edge at least once (i.e., the contribution from lists not containing all edges $\to 0$).

So now let us look at a fixed list $\mathcal{L}$ of N edges from G in which each edge appears at least once in the list. Given any permutation π of $\{e_1, \ldots, e_m\}$, define $\pi(\mathcal{L})$ to be the list that applies π to each member of $\mathcal{L}$. Since at each stage we randomly chose an edge, the probability that a random list of length N is $\mathcal{L}$ is the same as the probability that it is $\pi(\mathcal{L})$; indeed, both probabilities are $1/m^N$.

Moreover, the locations of the final edge appearances in $\mathcal{L}$ are the same as in $\pi(\mathcal{L})$, i.e., if e_i's final appearance is in position j of $\mathcal{L}$, then $\pi(e_i)$'s final appearance in $\pi(\mathcal{L})$ is in position j. This shows that, given two orderings $\mathcal{O}_1$ and $\mathcal{O}_2$ of $E(G)$, the number of length-N edge sequences in which the final appearances of each edge occurs in the order $\mathcal{O}_1$ is the same as the number that appear in the order $\mathcal{O}_2$. And hence, the probability of the ordering $\mathcal{O}_1$ is the same as the probability of the ordering $\mathcal{O}_2$.

Since these orderings are the only thing that determine the final coloring, and each ordering is equally likely, it suffices to look over all possible orderings $\mathcal{O}$ of the m edges and determine the resulting final coloring, i.e., the reduced interpretation of edge flipping. □

One immediate consequence of this is that there are only finitely many possible orderings of the edges and each occurs with probability some monomial in p and q. Therefore the probability of realizing a particular coloring c is a polynomial function of p and q. We note that the probability for a coloring c of a connected graph G is 0 if and only if the graph is bipartite with $V = A \cup B$ and all vertices in A are one color and vertices in B are the other color.

In the reduced interpretation we only considered each edge once. In fact, in the process of coloring the edges even fewer of the cards have an impact on the final coloring. This is because cards that occurred near the start of the deck are likely to have both of its vertices recolored by some other edges later in the process. This suggests we focus only on edges which will impact the final coloring, and leads to our next interpretation.

> **Reversed interpretation of edge flipping:** Take a graph G and a deck of $|E(G)|$ cards, i.e., one card for each edge, and start with no coloring on the vertices of G. Randomly shuffle the deck and then deal out the cards one at a time. For each card, if one (or both) of the vertices of the corresponding edge is uncolored, then with probability p color the uncolored vertex (or vertices) blue, and with probability q color the uncolored vertex (or vertices) red. If a vertex is colored already, do not recolor it.

Proposition 2. *Given a vertex coloring c of G, the probability that we end with coloring c is the same for both the reduced interpretation and the reversed interpretation.*

Proof. We could instead color in the reverse order, as follows. Given some ordering $\mathcal{O}$, let $\mathcal{O}'$ be the reverse ordering. Run through the reversed order, coloring edges as before, but now when presented with a vertex that is already colored, instead of recoloring it just leave it as is. This gives the same coloring as before.

We can now view this as taking a deck of m cards, one card for each edge of G, randomly shuffling this deck, and dealing the cards out one at a time. When e_i's card is dealt, locate e_i in G and, if any of its end vertices are uncolored, color them blue with probability p and red with probability q. □

The coloring on the reversed interpretation grows in bits and pieces, i.e., a card for an edge either colors two vertices, one vertex, or no vertex. We can approach the process through understanding this evolving coloring, which gives our last interpretation.

> **Constructive interpretation of edge flipping:** Take a graph G and a deck of $|E(G)|$ cards, i.e., one card for each edge. Start with an uncolored graph on the vertices of G but with no edges. Randomly shuffle the deck and deal out the cards one at a time. Each time an edge comes up, insert the edge into the graph. If at the time of insertion of the edge, one (or both) of the vertices of the edge is uncolored, then with probability p color the uncolored vertex (or vertices) blue and with probability q color the uncolored vertex (or vertices) red.

Since this works in the same manner as the reversed interpretation we have the following result.

Proposition 3. *The probability that we end with coloring c for the reversed interpretation is*

$$\frac{1}{m!}\sum_{\mathcal{O}} p^{s(\mathcal{O})}q^{t(\mathcal{O})}$$

where the sum is taken over all orderings, $\mathcal{O}$, of $E(G)$ which can yield the final coloring of c in the constructive interpretation (note in particular that the set of orderings is dependent on c). Further, for a given ordering $s(\mathcal{O})$ and $t(\mathcal{O})$ are the number of cards in the ordering which colored at least one final vertex blue and red, respectively.

When applying the constructive process on the graph G, if we disregard the edges which do not color a vertex, then the growing sequence of graphs induced by the resulting collection of edges will form a forest with no isolated vertices. In particular, each tree will have one edge which colored both its vertices and the remaining edges on the tree colored precisely one vertex. We can conclude that k is the number of trees in the forest if and only if there were exactly $n-k$ edges which contributed to the final coloring.

We now consider the special case of the all-blue coloring of the graph (so each card which colors at least one vertex was chosen to color blue with probability p). We have that the coefficient of p^{n-k} is the proportion of all orderings where $n-k$ edges contributed to the final coloring, i.e., the final graph when ignoring edges which did not color has k trees.

Let $F_G(k)$ denote the number of orderings of the edges of G so that the final graph when ignoring edges which did not color has k trees. When $G = K_n$ we simply write $F_n(k)$. Then we have

$$\mathbb{P}(c \text{ is all blue}) = \sum_k \frac{F_G(k)}{m!}p^{n-k}.$$

A similar analysis can be done when we are not using the all-blue coloring, and we will return to this in a later section.

3. Probability k specified vertices are colored blue

In this section we will look at the probability that a specific set of k labeled vertices are colored blue in the stationary distribution. While this is an interesting question in its own right, we will mainly use this to establish the all-blue case for Theorem 1.

Before we begin, we will need to introduce the idea of a *restricted* coloring. Let G be a graph and c be a coloring of a subgraph H of G. Define $\mathbb{P}_G(c)$ to be the probability of realizing the coloring c on G, i.e., the probability that the coloring when restricted to the vertices of H agree with c (the vertices $V(G) \setminus V(H)$ are allowed to be any color). Alternatively, $\mathbb{P}_G(c)$ is the sum of the probabilities in the stationary distribution of all final colorings which agree with c on H.

Lemma 1. *Let G be a graph and c be a coloring of a subgraph H of G. Let G' be the graph obtained by removing all the edges from G which have neither endpoint in H. If G' has no isolated vertices, then*

$$\mathbb{P}_G(c) = \mathbb{P}_{G'}(c).$$

Proof. We will use the reverse interpretation of edge flipping. Let T be the deleted edges with $m = |E(G)|$ and $t = |T|$. Let L be any list of the edges of G which gives the coloring c. Notice that the edges from T do not color any of the vertices of $V(H)$ whose colors we demand match c. Therefore removing them from L still leaves a list L' of the edges of G' which color H exactly as before (note that every vertex of G still gets a color since G' has no isolated vertices).

Moreover, for a fixed list L' of the edges of G' giving the coloring c, it is easy to see that the number of lists of $E(G)$ which reduce to L' is precisely $\binom{m}{t}t!$, i.e., we pick the location of the t edges from m slots as well as the ordering of the t edges. Note that this quantity is independent of our choice of L'. Likewise, given a list of $E(G')$ that gives a coloring different than c, there are again precisely $\binom{m}{t}t!$ lists of $E(G)$ which reduce to the chosen list.

Since every list of G can be reduced we conclude that the proportion of lists of $E(G)$ giving the coloring c is the same as the proportion of lists of $E(G')$ giving the coloring c. □

So to determine the probability that k specified vertices are blue we can work on a "simpler" graph. In particular we get the following recurrence.

Theorem 2. *Let $G = K_n$ and let $2 \le k \le n$. Let $Q_n(k)$ denote the probability that k* specified *vertices are blue (regardless of the coloring on the remaining $n - k$ vertices). Then*

$$Q_n(k) = \frac{(2n-2k)p}{2n-k-1}Q_n(k-1) + \frac{(k-1)p}{2n-k-1}Q_n(k-2)$$

with $Q_n(0) = 1$ and $Q_n(1) = p$.

Proof. We will use the reversed interpretation of the problem. Observe that the initial condition $Q_n(0) = 1$ holds since any coloring works (i.e., there is no restriction). Also $Q_n(1) = p$ since in any ordering of the edges, the first edge to contain the specified vertex will determine its color, and with probability p that card will color the vertex blue.

By Lemma 1, we may instead consider the graph G' obtained by removing all edges disjoint from our specified vertices, $\{v_1, \dots, v_k\}$. Note that G' is the graph $K_k \vee \big((n-k)K_1\big)$, i.e., the join of a clique on k vertices with $n-k$ isolated vertices. We now use this to establish the recurrence.

Consider a list L of $E(G')$, and let e_1 be the first edge in L. The first case is that e_1 has exactly one vertex in the clique on $\{v_1, v_2, \dots, v_k\}$. Say $e_1 = \{v_i, u\}$ where $u \notin \{v_1, \dots, v_k\}$. This case occurs $k(n-k)/\big(\binom{k}{2}+k(n-k)\big) = (2n-2k)/(2n-k-1)$ proportion of the time, and in order for v_i to be colored blue we must have that the vertices of e_1 were chosen to be blue, which happens with probability p.

Assuming the above occurs, we claim that the probability that the remainder of the process produces a legal coloring is $Q_n(k-1)$. Consider the current state of v_i. Vertex v_i is now the correct color and so for the rest of the list it does not matter what colors its incident edges are given; this is exactly the coloring property of the vertices $V(G) \setminus \{v_1, \ldots, v_k\}$. Moreover, all edges from v_i to $V(G) \setminus \{v_1, \ldots, v_k\}$ will not affect whether the coloring c occurs, and so (essentially by another application of Lemma 1) we may disregard these edges. Therefore by moving v_i out of the clique (shrinking the clique size by 1), we see that we have an analogous problem with probability $Q_n(k-1)$.

The second case is that e_1 has both vertices in the clique on $\{v_1, \ldots, v_k\}$. This occurs $\binom{k}{2}/\big(\binom{k}{2}+k(n-k)\big) = (k-1)/(2n-k-1)$ proportion of the time and with probability p the vertices of the edge will be colored blue, as required. Given this, the only edges containing either of these vertices that matter are the ones between one of them and another vertex in $\{v_1, \ldots, v_k\}$. Therefore, as before, we may move these vertices out of the clique and with probability $Q_n(k-2)$ the remaining process will produce a legal coloring. This gives the recurrence. ∎

By examining small cases an explicit solution to this recurrence was found.

Theorem 3. *Let $Q_n(k)$ denote the probability that k* specified *vertices are blue (regardless of the coloring on the remaining $n-k$ vertices). Then*

$$Q_n(k) = \frac{\sum_j (n-k+j)^{\overline{\lfloor k/2\rfloor - j}}\, 2^{\lfloor k/2\rfloor - 2j} p^{k-j} \binom{k}{2j} \frac{(2j)!}{j!}}{\prod_{\ell=1}^{\lfloor k/2\rfloor}(2n-2\ell-1)},$$

where $m^{\overline{k}} = m(m+1)\cdots\big(m+(k-1)\big)$ denotes the rising factorial.

Proof. For $k=0$ and 1, the denominator will be an empty product which by convention is 1 and the sum in the numerator will be nonzero only for $j=0$, which gives 1 and p respectively. This establishes the base cases.

It now suffices to verify the recurrence, and for this we will find it useful to treat these expressions as polynomials in p, i.e.,

$$Q_n(k) = \sum_j c_{n,k}(j) p^{k-j}$$

where

$$c_{n,k}(j) = \frac{(n-k+j)^{\overline{\lfloor k/2\rfloor - j}}\, 2^{\lfloor k/2\rfloor - 2j} \binom{k}{2j} \frac{(2j)!}{j!}}{\prod_{\ell=1}^{\lfloor k/2\rfloor}(2n-2\ell-1)}.$$

Note that the $\binom{k}{2j}$ term indicates this will only be nonzero when $0 \le j \le \lfloor k/2\rfloor$.

The recurrence from Theorem 2 translates into the following recurrence on the coefficients:

$$c_{n,k}(j) = \frac{2n-2k}{2n-k-1} c_{n,k-1}(j) + \frac{k-1}{2n-k-1} c_{n,k-2}(j-1). \tag{1}$$

So it suffices to verify (1). Since the expression for the coefficient involves the term $\lfloor k/2 \rfloor$, it is useful to separate the verification of (1) into two cases depending on the parity of k, i.e., $\lfloor (k-1)/2 \rfloor = \lfloor k/2 \rfloor$ when k is odd and $\lfloor (k-1)/2 \rfloor = \lfloor k/2 \rfloor - 1$ when k is even. Now substituting in the expressions for $c_{n,k}(j)$ and carry out routine simplifications for both cases establishes the results (we leave the details to the interested reader). □

The important case for us is when $k = n$, which is equivalent to having all of the vertices colored blue. In particular by evaluating $Q_n(k)$ at $k = n$ we have the following.

Corollary 1. *The probability that the complete graph is colored all-blue under the edge flipping process is given by*

$$\sum_{j=1}^{\lfloor n/2 \rfloor} \frac{\binom{2j-1}{j}\binom{n-1}{2j-1}2^{n-2j}}{\binom{2n-2}{n}} p^{n-j}.$$

The proof proceeds by taking the definition of $Q_n(n)$ and then applying routine simplifications. Since this involves the expression $\lfloor n/2 \rfloor$, it is useful to break into two cases, n is even and n is odd. We again leave the details to the interested reader.

Recall that $F_n(k)$ denotes the number of orderings of the edges of K_n in the constructive interpretation of edge flipping so that the final graph (when ignoring edges which did not color) has k trees. We also have

$$\mathbb{P}(c \text{ is all blue}) = \sum_k \frac{F_n(k)}{m!} p^{n-k}.$$

Combining this with the previous corollary now gives us the following.

Corollary 2. $F_n(k) = \dfrac{\binom{2k-1}{k}\binom{n-1}{2k-1}2^{n-2k}\binom{n}{2}!}{\binom{2n-2}{n}}$.

It is possible to give a direct combinatorial proof in the case $k = 1$. Finding a general combinatorial proof remains an open problem.

Proof of $k = 1$ case. This is equivalent to showing

$$\frac{F_n(k)}{\binom{n}{2}!} = \frac{2^{n-2}}{C_{n-1}},$$

where C_{n-1} is the $(n-1)$-st Catalan number. On the left hand side is the probability that an ordering of the edges results in a single tree, so we now work to show that the right hand side gives the same probability.

The first card will always color two vertices, and so if at the end there is a single tree, then it must be the case that every subsequent card which colors will only color one vertex. In particular, if there are currently t vertices in the tree, then the probability that the next edge which colors will color precisely one vertex is

$$\frac{t(n-t)}{t(n-t)+\binom{n-t}{2}} = \frac{2t}{n+t-1}.$$

This must hold for $t = 2, \ldots, n-1$ and therefore the probability that the ordering of the edges results in a single tree is

$$\prod_{t=2}^{n-1} \frac{2t}{n+t-1} = \frac{2^{n-2}\prod_{t=2}^{n-1} t}{\prod_{t=2}^{n-1}(n+t-1)} = \frac{2^{n-2}(n-1)!}{(2n-2)!/n!} = \frac{2^{n-2}}{C_{n-1}}. \qquad \square$$

4. Probability of an arbitrary coloring

Once we have the probability that b specified vertices are colored blue, then by use of an inclusion-exclusion argument we can find the probability that b specified vertices are blue and the remainder are red. However, this will result in an alternating sum, and be entirely in terms of p, when we want a sum with non-negative terms which is in terms of p and q. Therefore we will take a slightly different approach by using the constructive interpretation of edge flipping.

Theorem 4. *For the edge flipping process on the complete graph $G = K_n$, let c be a coloring which assigns red to r* specified *vertices with the remainder being blue. Then the probability of having the coloring c is given by*

$$\sum_{k=1}^{\lfloor n/2\rfloor}\sum_{s=0}^{\lfloor r/2\rfloor} \frac{\binom{n-1}{n-2k-r+2s,\,r-2s,\,k-1,\,k-s,\,s}}{\binom{2n-2}{n-2,\,n-r,\,r}} 2^{n-2k}p^{n-k-r+s}q^{r-s}.$$

We note the result does not depend on which vertices have been specified red and blue, and so if we only care that r of the vertices are red and the remainder are blue then we multiply the probability in Theorem 4 by $\binom{n}{r}$. Further, making the substitutions $n = b+r$, $-k+s = -j'$ and $s = k'$ and then simplifying shows that this result is equivalent to Theorem 1.

Before we begin the proof of Theorem 4 we point out that when $r = 0$ then s is forced to be 0 and this reduces the expression to the one given in Corollary 1. So the result is true in this case.

Proof. We use the constructive interpretation of edge flipping. Let $\mathcal{F}_n(k)$ be the set of orderings of the edges of K_n so that the final graph when ignoring edges which will not color has k trees ($1 \le k \le \lfloor n/2\rfloor$). Note that $F_n(k) = |\mathcal{F}_n(k)|$. For each tree there will be a unique edge which colors both of its vertices, and the remaining edges in the tree color one of its vertices. In particular there will be k edges which color 2 vertices, $n-2k$ edges which color one vertex, and the remaining edges will not color.

Consider an arbitrary $S \in \mathcal{F}_n(k)$ and let $\mathcal{T}_S$ be the set of k edges which color two vertices. Clearly these edges color an even number of vertices red, so suppose $2s$ is that number ($0 \le s \le \lfloor r/2 \rfloor$). We will now count the number of such orderings which will color s of these special edges red and have the proper *number* of final red and blue vertices.

Define an *r-set* to be a set of r vertices from G. We first count the total number of r-sets for which exactly $2s$ of its members are paired up by s edges from $\mathcal{T}_S$, and the remaining $r-2s$ vertices are disjoint from the edges of $\mathcal{T}_S$; call this property of an r-set *property-s*. This is easy, as there are $\binom{k}{s}$ ways to pick the s edges from $\mathcal{T}_S$ containing our first $2s$ vertices, and $\binom{n-2k}{r-2s}$ ways to choose the remaining. Summing over all $S \in \mathcal{F}_n(k)$ gives a total of $\binom{n-2k}{r-2s}\binom{k}{s}F_n(k)$ distinct occurrences of property-s, among all builds in $\mathcal{F}_n(k)$.

Since G is symmetric between r-sets, and the set $\mathcal{F}_n(k)$ is symmetric within $V(G)$ in the sense that any build from this collection can be translated to another by simply permuting the vertex set, it is clear that any two r-sets will have property-s the same number of times among the entire collection. Therefore this total must be evenly distributed among all $\binom{n}{r}$ of the r-sets. In particular, the unique r-set of red vertices in our coloring $c$must occur precisely

$$\frac{\binom{n-2k}{r-2s}\binom{k}{s}F_n(k)}{\binom{n}{r}}$$

times.

Given such an $S \in \mathcal{F}_n(k)$, there is a $1/\binom{n}{2}!$ probability that it will appear, and if it does there is a $p^{n-k-r+s}q^{r-s}$ probability that it will be colored in the unique way giving the coloring c. In particular, the s edges from $\mathcal{T}_S$ that pair up $2s$ red vertices must be red, and the remaining $r-s$ red vertices will each be colored by whatever is the first edge in S to contain it. Since these vertices are not part of $\mathcal{T}_S$, precisely one edge is needed for each. Therefore there are $r-s$ edges which will color the red vertices, and since precisely $n-k$ edges contribute to the coloring, there must be $n-k-(r-s)$ edges corresponding to coloring the vertices of an edge blue. Summing over all choices of s and k gives a probability of

$$P_G(c) = \sum_{k=1}^{\lfloor n/2 \rfloor} \sum_{s=0}^{\lfloor r/2 \rfloor} \frac{\binom{n-2k}{r-2s}\binom{k}{s}F_n(k)}{\binom{n}{r}\binom{n}{2}!} p^{n-k-r+s}q^{r-s}.$$

of obtaining the coloring c. Applying Corollary 2 and collapsing the resulting binomial coefficients to multinomial coefficients gives the result. □

5. Asymptotic results

In this section we will consider some asymptotic results related to Theorem 4, in particular focusing on how likely a particular coloring will occur. The approach we will give is to first express results in terms of a multivariate generating function and then apply known asymptotic tools to estimate the coefficients of these functions.

Generating functions

We begin with the following expression for the probability that b vertices are blue and r vertices are red:

$$\frac{1}{\binom{2n-2}{n-2}}\sum_{k\geq 1}\sum_{s\geq 0}\binom{n-1}{n-2k-r+2s,\, r-2s,\, k-1,\, k-s,\, s}2^{n-2k}p^{n-k-r+s}q^{r-s}, \qquad (2)$$

where $n = b + r$ and we have multiplied by $\binom{n}{r}$ and simplified the term in front since we don't care which r vertices are red. Define

$$g(n,r) = \sum_{k\geq 1,s}\binom{n-1}{n-2k-r+2s,\, r-2s,\, k-1,\, k-s,\, s}2^{n-2k}p^{n-k-r+s}q^{r-s} \qquad (3)$$

$$= \sum_{k\geq 1,s}\binom{n-1}{n-2k-r+2s,\, r-2s,\, k-1,\, k-s,\, s}(2p)^n\left(\frac{1}{4p}\right)^k\left(\frac{q}{p}\right)^r\left(\frac{p}{q}\right)^s.$$

Next, we define the generating function

$$\begin{aligned}
F(x,y,z,w) &= \sum_{n,k\geq 1,r,s}\binom{n-1}{n-2k-r+2s,\, r-2s,\, k-1,\, k-s,\, s}x^n y^k z^r w^s \\
&= \sum_{n,k\geq 1,s}\binom{n-1}{n-2k,\, k-1,\, k-s,\, s}(1+z)^{n-2k}x^n y^k w^s z^{2s} \\
&= \sum_{n,k\geq 1,s}\binom{n-1}{n-2k,\, k-1,\, k-s,\, s}(x(1+z))^n\left(\frac{y}{(1+z)^2}\right)^k (z^2w)^s \\
&= \sum_{n,k\geq 1}\binom{n-1}{n-2k,\, k-1,\, k}(1+z^2w)^k(x(1+z))^n\left(\frac{y}{(1+z)^2}\right)^k \\
&= \sum_{n,k\geq 1}\binom{n-1}{n-2k,\, k-1,\, k}(x(1+z))^n\left(\frac{y(1+z^2w)}{(1+z)^2}\right)^k \\
&= \sum_{n,k\geq 1}\frac{1}{2}\binom{n-1}{2k-1}\binom{2k}{k}(x(1+z))^n\left(\frac{y(1+z^2w)}{(1+z)^2}\right)^k \\
&= \sum_{k\geq 1}\frac{1}{2}\binom{2k}{k}(x(1+z))\left(\frac{y(1+z^2w)}{(1+z)^2}\right)^k\sum_{N}\binom{N}{2k-1}(x(1+z))^N \\
&\qquad \text{(where we set } N = n-1\text{)} \\
&= \sum_{k\geq 1}\frac{1}{2}\binom{2k}{k}(x(1+z))\left(\frac{y(1+z^2w)}{(1+z)^2}\right)^k\frac{(x(1+z))^{2k-1}}{(1-x(1+z))^{2k}} \\
&= \frac{1}{2}\sum_{k\geq 1}\binom{2k}{k}\left(\frac{x^2y(1+z^2w)}{(1-x(1+z))^2}\right)^k \\
&= \frac{1}{2}\left(\frac{1}{\sqrt{1-\frac{4x^2y(1+z^2w)}{(1-x(1+z))^2}}}-1\right)
\end{aligned}$$

$$= \frac{1}{2}\left(\frac{(1-x(1+z))}{\sqrt{(1-x(1+z))^2-4x^2y(1+z^2w)}}-1\right)$$

Finally, we consider the generating function:

$$G(X,Y) = \sum_{n\geq 1,r} g(n,r)X^nY^r. \tag{4}$$

From (2) and (3), we have

$$\begin{aligned} G(X,Y) &= F\left(2pX, \frac{1}{4p}, \frac{q}{p}Y, \frac{p}{q}\right) \\ &= \frac{1}{2}\left(\frac{1-2pX(1+\frac{q}{p}Y)}{\sqrt{(1-2pX(1+\frac{q}{p}Y))^2-4(2pX)^2\frac{1}{4p}(1+\frac{q^2}{p^2}Y^2\frac{p}{q})}}-1\right) \\ &= \frac{1}{2}\left(\frac{1-2X(p+qY)}{\sqrt{(1-2X(p+qY))^2-4X^2(p+qY^2)}}-1\right) \end{aligned} \tag{5}$$

Asymptotics for the all blue coloring

We consider the special case that $r=0$, i.e., all the vertices of K_n are blue. The corresponding generating function is given by substituting $Y=0$ in (5).

$$G(X,0) = \sum_n g(n,0)X^n = \frac{1}{2}\left(\frac{1-2pX}{\sqrt{(1-2pX)^2-4pX^2}}-1\right) \tag{6}$$

To determine the asymptotic behavior of $g(n,0)$ as $n\to\infty$, we use the following result (see [5]).

Theorem 5 (Darboux [3])**.** *Suppose that $f(z)$ is analytic for $|z|<r$, $r>0$, and has only algebraic singularities on $|z|=r$. Let a be the minimum of $\mathbf{Re}(\alpha)$ for the terms of the form $(1-z/w)^\alpha h(z)$ at the singularities of $f(z)$ on $|z|=r$, and let w_j, α_j and $h_j(z)$ be the w, α, and $h(z)$ for those terms of the form $(1-z/w)^\alpha h(z)$ for which $\mathbf{Re}(\alpha)=a$. Then, as $n\to\infty$,*

$$[z^n]f(z) = \sum_j \frac{h_j(w_j)n^{-\alpha_j-1}}{\Gamma(-\alpha_j)w_j^n}. \tag{7}$$

Here, $[z^n]f(z)$ denotes the coefficient of z^n in the series expansion for $f(z)$. To apply this to our situation, we take

$$\begin{aligned} f(X) &= \frac{1}{\sqrt{1-2(p+\sqrt{p})X}}, \\ h(X) &= \frac{1-2pX}{\sqrt{1-2(p-\sqrt{p})X}}, \\ w &= \frac{1}{2(p+\sqrt{p})}, \end{aligned}$$

$$\alpha = a = -\frac{1}{2}.$$

Plugging these values into (7), using $\Gamma(\frac{1}{2}) = \sqrt{\pi}$ and simplifying gives

$$g(n,0) = \frac{1}{2}[X^n]f(X) = \frac{1}{2\sqrt{2\pi n}\sqrt{1+\sqrt{p}}}(2(p+\sqrt{p}))^n + o\left(\frac{(2(p+\sqrt{p}))^n}{\sqrt{n}}\right) \quad (8)$$

where the extra factor of $\frac{1}{2}$ comes from the (easy-to-forget) factor of $\frac{1}{2}$ in (6).

Now, to get the asymptotic value of the probability $\mathbb{P}(n \text{ blue; } 0 \text{ red})$, we must divide by $\binom{2n-2}{n-2}$ which by Stirling's formula is asymptotic to $\frac{2^{2n-2}}{\sqrt{\pi n}}$. Putting this together with (8) gives the final result.

Theorem 6. *For $0 \le p \le 1$, the probability that all the vertices of K_n are blue is*

$$\mathbb{P}(n \textit{ blue; } 0 \textit{ red}) = \sqrt{\frac{2}{1+\sqrt{p}}}\left(\frac{p+\sqrt{p}}{2}\right)^n (1+o(1)) \quad \textit{as } n \to \infty. \quad (9)$$

Setting $p = \frac{1}{2}$ we obtain,

Corollary 3. *With $p = \frac{1}{2}$, the probability that all the vertices of K_n are blue is*

$$\mathbb{P}(n \textit{ blue; } 0 \textit{ red}) = (4 - 2\sqrt{2})\left(\frac{1+\sqrt{2}}{4}\right)^n (1+o(1)) \quad \textit{as } n \to \infty. \quad (10)$$

Asymptotics for a fixed proportion of the vertices blue

We now look at the problem of estimating $\mathbb{P}(b \text{ blue; } r \text{ red})$ along a ray where the ratios $\frac{b}{r}$ and $\frac{r}{b}$ are both bounded away from 0. For this case we will rely on some relatively recent tools which obtain asymptotic estimates for multivariate generating functions (e.g., see [6]), as opposed to the univariate generating functions which we had in the previous section.

Consider a general generating function G of the form

$$G(x,y) = \frac{F(x,y)}{(H(x,y))^\beta} = \sum_{n,r} c_{n,r} x^n y^r$$

where H is analytic and β is positive. The growth rates of the coefficients $c_{n,r}$ are determined by the solutions of $H(x,y) = 0$. For the "directional" asymptotics $r/n \sim \lambda$, with $0 < \lambda < 1$, the growth rate is determined by solving the following system of two equations:

$$H(x,y) = 0,$$
$$\frac{yH_y}{xH_x} = \lambda.$$

The solutions for these equations are called *critical points.*

We need the following result, which is a special case of a more general result.

Theorem 7 (Greenwood [4]). *Let H be an analytic function with a single smooth strictly minimal critical point (x_0, y_0), where x_0 and y_0 are real and positive. Suppose H has only real coefficients in its power series expansion about the origin. Assume $H(0,0) > 0$, and consider $H^{-\beta}$ for β a real positive number with the standard branch chosen along the negative real axis, so that $H(0,0)^\beta > 0$. Let $\lambda = \frac{r+O(1)}{n}$ be fixed, $0 < \lambda < 1$, as $n, r \to \infty$. Define the following quantities:*

$$\theta_1 = \frac{H_y(x_0, y_0)}{H_x(x_0, y_0)} = \frac{\lambda x_0}{y_0},$$
$$\theta_2 = \frac{1}{2H_x}(\theta_1^2 H_{xx} - 2\theta_1 H_{xy} + H_{yy}) \text{ evaluated at the point } (x_0, y_0)$$
$$\theta_3 = \frac{1}{\sqrt{\frac{2\theta_2}{x_0} + \frac{\theta_1^2}{x_0^2} + \frac{\lambda}{y_0^2}}},$$

In the definition of θ_3, the term underneath the square root is always positive, and the positive square root should be taken. Assume that $H_x(x_0, y_0)$ and $\frac{2\theta_2}{x_0} + \frac{\theta_1^2}{x_0^2} + \frac{\lambda}{y_0^2}$ are nonzero. Then the following expression holds as $n, r \to \infty$ with $\lambda = \frac{r+O(1)}{n}$:

$$c_{n,r} \sim \frac{\theta_3 (H_x(x_0, y_0) x_0)^{-\beta} F(x_0, y_0) n^{\beta - 3/2}}{\Gamma(\beta)\sqrt{2\pi}}.$$

Using the above theorem, we have the following result for general p.

Theorem 8. *For $0 < p < 1$, with $p + q = 1$, we have*

$$\mathbb{P}(pn \text{ blue; } qn \text{ red}) = \frac{1}{\sqrt{3pq\pi n}} + o\left(\frac{1}{\sqrt{n}}\right). \tag{11}$$

Proof. We start with the generating function in (5). Our goal is to estimate $g(n, \lambda n)$ for $\lambda = 1 - p$. (As one would expect, when $\lambda \neq 1 - p$, then the probability that there are just λn red vertices goes to 0 exponentially rapidly in n; see the example after the proof).

The generating function G can be written (replacing X by x and Y by y) as

$$G(x, y) = \frac{F(x, y)}{(H(x, y))^{1/2}} - \frac{1}{2}$$

where

$$H(x, y) = (1 - 2x(p + qy))^2 - 4x^2(p + qy^2),$$
$$F(x, y) = \frac{1 - 2x(p + qy)}{2}.$$

For the directional asymptotics $r/n \sim \lambda$, the growth rate is determined by solutions (x_0, y_0) of the following system of two equations:

$$H(x, y) = (1 - 2x(p + qy))^2 - 4x^2(p + qy^2) = 0,$$

$$\frac{yH_y}{xH_x} = \lambda.$$

The unique solution satisfying $0 < x_0 \leq 1/2$ and y_0 positive is $x_0 = 1/4$ and $y_0 = 1$. Then in our case (where $\beta = \frac{1}{2}$ and $\lambda = 1 - p$),

$$g(n, \lambda n) \sim C(n) x_0^{-n} y_0^{-\lambda n} = C(n)\, 4^n$$

where $C(n)$ is determined by the following values:

$$\begin{aligned}
\theta_1 &= \frac{\lambda x_0}{y_0} = \frac{1-p}{4}, \\
\theta_2 &= \frac{1}{2H_x}(\theta_1^2 H_{xx} - 2\theta_1 H_{xy} + H_{yy}) \text{ evaluated at } (x_0, y_0) = (1/4, 1) \\
&= \frac{(5p-4)(1-p)}{16}, \\
\theta_3 &= \frac{1}{\sqrt{\frac{2\theta_2}{x_0} + \frac{\theta_1^2}{x_0^2} + \frac{\lambda}{y_0^2}}} = \sqrt{\frac{2}{3p(1-p)}}, \\
C(n) &= \frac{\theta_3 (H_x(x_0, y_0) x_0)^{-\beta} F(x_0, y_0) n^{\beta - 3/2}}{\Gamma(\beta)\sqrt{2\pi}} = \frac{1}{4\pi n \sqrt{3pq}} 4^n.
\end{aligned}$$

Therefore we can estimate the probability of pn blue nodes and qn red nodes by

$$\mathbb{P}(pn \text{ blue}; qn \text{ red}) = \frac{g(n, (1-p)n)}{\binom{2n-2}{n}} \sim \frac{C(n) 4^n}{\frac{4^n}{4\sqrt{\pi n}}} = \frac{1}{\sqrt{3pq\pi n}}$$

as desired. □

We note that when λ differs from $1 - p$, the solution of the two equations has x_0' strictly greater than $1/4$ and consequently $\mathbb{P}((1-\lambda)n \text{ blue}; \lambda n \text{ red})$ will be $O((\frac{1}{4x_0'})^n)$, i.e., it will go to 0 exponentially rapidly in n. For example, with $p = \frac{1}{2} = q$ and $\lambda = \frac{1}{3}$, we find that the two equations have a unique solution with $y_0' = 0.62741\ldots$ being the positive root of $4y^3 + 8y^2 - 5y - 1$ and $x_0' = \frac{10}{9} y_0'^2 + 2y_0' - \frac{25}{18} = 0.303313\ldots$.

6. Conclusion

We have been able to establish the stationary distribution for the edge flipping process on the complete graph. There are many directions to take this problem and we mention a few of them here.

- One natural question is to ask the rate of convergence of the random walk interpretation of the edge flipping process, i.e., how quickly do we converge to the stationary distribution. One way to determine this is to use the spectrum of the probability transition matrix on the state graph.

The spectrum, and in particular how closely the non-trivial eigenvalues cluster around 0, give a bound on the rate of convergence to the stationary distribution. Our work does not show how to establish the spectrum. We remark that the spectrum for the path and cycle were determined in the analysis carried out in the earlier paper of Chung and Graham [2].

- We can view the edge flipping process on the complete graph in the following way: Pick two vertices at random and with probability p color them both blue and with probability $q := 1-p$ color them both red. Instead of picking two vertices at a time we could pick k vertices at a time and then with probability p make them all blue and with probability q make them all red. This is equivalent to carrying out edge flipping on the complete k-uniform hypergraph. We point out that the case $k = 1$ simply states to flip each vertex randomly and in that case the stationary distribution is easily determined, in particular we have:

$$\mathbb{P}(b \text{ blue}; r \text{ red}) = \binom{b+r}{b} p^b q^r.$$

- We can also increase the number of colors, for instance blue, red, and yellow. So that an edge changes to blue with probability p, to red with probability q and to yellow with probability $r := 1-p-q$. We point out that the results of coloring k specified vertices blue given in Section 3 still hold since we can combine the two other colors together (making orange?). So that the work in establishing the general case comes in the bootstrapping given in Section 4.

- The analysis can be carried out for any graph or families of graphs. The most natural next family to consider would be the complete bipartite graphs. For the special case $K_{1,n}$ the analysis is again straightforward to carry out. Namely the probability that the "center" vertex is blue and b of the remaining are blue and $r = n - b$ are red is

$$\binom{b+r}{b} \frac{b}{b+r} p^b q^r.$$

  Similarly the probability that the center vertex is red and b of the remaining are blue and $r = n - b$ are red is

$$\binom{b+r}{b} \frac{r}{b+r} p^b q^r.$$

  (It is easy to check that the probabilities sum to 1.)

  In addition to cycles, paths, and complete graphs it is possible to implement a program to directly determine the stationary distribution for any *small* graph. The study of these small graphs might offer some insight into what is happening for larger graphs and warrant further exploration. For example, one can show that for the edge flipping process on the Petersen graph, P, that

$$\mathbb{P}(P \text{ is all blue}) = \frac{326p^9 + 4352p^8 + 10923p^7 + 4744p^6 + 130p^5}{20475}.$$

We look forward to seeing more progress in this area.

Acknowledgment. The authors thank the referee for their close reading of the paper and many useful suggestions and corrections. Steve Butler's work was conducted while visiting the Institute for Mathematics and its Applications.

References

[1] C. Athanasiatis, P. Diaconis, Functions of random walks on hyperplane arrangements, *Adv. in Applied Math.* **45** (2010), 410–437.

[2] F. Chung and R. Graham, Edge flipping in graphs, *Adv. in Applied Math.* **48** (2012), 37–63.

[3] G. Darboux, Memoire sur l'approximation des fonctions de très-grande nombres, et sur une classe étendu de développements en série, *J. Math. Pures Appl.* **4**, (1878), 5–56, 377–416.

[4] T. Greenwood, Asymptotics of the coefficients of bivariate analytic functions with algebraic singularites, (preprint) (2014).

[5] A. M. Odlyzko, Asymptotic enumeration methods, in Handbook of Combinatorics, R. Graham, L. Lovász, M. Grötschel, eds. Elsevier, 1995, pp. 1063–1229.

[6] R. Pemantle and M. Wilson, Analytic combinatorics in several variables, Cambridge Studies in Advanced Mathematics, 140. Cambridge Univ. Pr., Cambridge, 2013, xiv+380 pp.