Probability recurrences on simple graphs in a forest building process

by

Joseph Alameda

A creative component submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Major: Mathematics

Program of Study Committee:
Steve Butler, Major Professor
Bernard Lidický
Michael Young

Iowa State University

Ames, Iowa

2019

DEDICATION

I would like to dedicate this thesis to Scarlitte Ponce who taught me how to add fractions when I first started college. We have come a long way.

TABLE OF CONTENTS

LIST OF FIGURES

ABSTRACT

Consider the following process on a simple graph with no isolated vertices: Randomly order the edges and remove an edge if and only if the edge is incident to two vertices already incident to some preceding edge. This process results in a spanning forest of the graph.

Recurrences are given for the process for multiple families of graphs, and the probability of obtaining k components in the above process is given by a new method for the Fan graph $F_{n-2,2}$. An approach to proving a previously published conjecture is also discussed.

CHAPTER 1. DEFINITIONS AND BACKGROUND

We first introduce the necessary definitions that are used throughout the paper, and we also introduce the forest building process.

1.1 Definitions

Let V be a finite nonempty set. A *graph* $G = (V, E)$ is a pair of sets V and E such that E contains 2 element subsets of V. Elements in V and E are called *vertices* and *edges* respectively, and are often denoted $V(G)$ and $E(G)$. Given vertices u and v in $V(G)$, u and v are said to be *adjacent* if $\{u, v\}$, denoted uv, is in the edge set $E(G)$. Adjacent vertices are often called *neighbors* and the set of neighbors of a vertex v in $V(G)$ is called the *neighborhood* of v. This set will be denoted $N(v)$. An *independent set* is a set of vertices in $V(G)$ that are not adjacent to one another. A *leaf* is a vertex with exactly one neighbor. An *isolated vertex* is a vertex with no neighbor. A graph H is a *subgraph* of graph G if $V(H) \subseteq V(G)$ and $E(H) \subseteq E(G)$. A subgraph H is said to be a *spanning subgraph* of a graph G if $V(H) = V(G)$.

A *tree* T is a graph in which any two vertices is connected by a unique path, and a *forest* F is the disjoint union of trees. A *complete graph*, denoted K_n, is the graph on n vertices with all possible edges, and a *complete bipartite graph*, denoted $K_{m,n}$, is the graph on $m + n$ vertices with an independent set of size m and an independent set of size n with all possible edges between both sets.

Let G be a simple graph with no isolated vertices. Consider the following *forest building process*:

1. Take a random ordering of edges of $E(G)$, say $e_1, e_2, \ldots, e_n$

2. Remove the edge e_j from $E(G)$ if it is incident to two vertices that are incident to e_{j-a} and e_{j-b} for some positive numbers a and b.

We note that since this does not allow cycles, the end result is a forest. This must be a spanning forest of G since every vertex remains in the graph. We illustrate this process on K_4 in Figure 1.1.

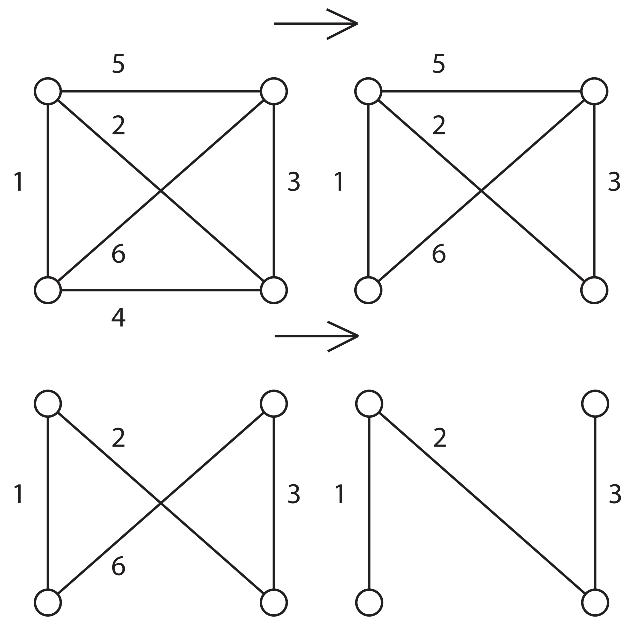

Figure 1.1 Forest Building Process on K_4

1.2 Background

This process was first considered in (2). Let $F(G, k)$ denote the number of edge orderings that result in the forest building process producing k trees. Let $P(G, k)$ denote the probability that a random ordering produces k trees and let $p_G(x) = \sum_k P(G, k)x^k$. The following was shown in (2):

Theorem 1.2.1 (Butler-Chung-Cummings-Graham)**.** *For the complete graph K_n we have*

$$p_{K_n}(x) = \sum_k \frac{\binom{n-1}{n-2k,k,k-1} 2^{n-2k}}{\binom{2n-2}{n}} x^k.$$

The polynomial $p_G(x)$ was also found for the complete bipartite graph in (1).

Theorem 1.2.2 (Berikkyzy-et al.)**.** *For the complete bipartite graph $K_{s,t}$, we have*

$$p_{K_{s,t}}(x) = \sum_k \frac{k(s+t)\binom{s}{k}\binom{t}{k}}{st\binom{s+t}{s}} x^k.$$

In following chapters, we establish a process to build recurrences for multiple graphs, find the polynomial $p_G(x)$ for the Fan graph $F_{n-2,2}$, and also discuss a way to prove a conjecture from (1).

CHAPTER 2. THE GRAPHS $\ell^{(r)}K_n$

We will first construct the graphs $\ell^{(r)}K_n$, and then give a probability recurrence that looks at a certain point in the tree building process.

2.1 Construction of $\ell^{(r)}K_n$

Given a complete graph K_n, we form $\ell^{(r)}K_n$ by attaching r leaves to each vertex. Figure 2.1 shows an example of such a graph.

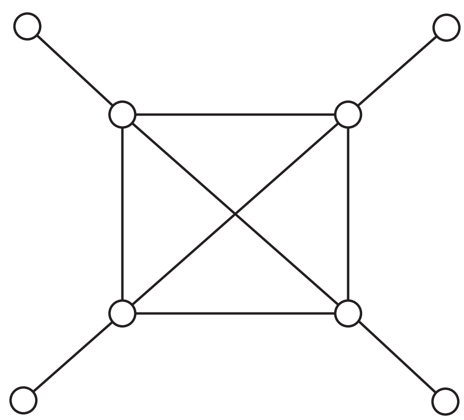

Figure 2.1 The $\ell^{(1)}K_4$ graph

2.2 Probability recurrence for $\ell^{(r)}K_n$

Theorem 2.2.1. *Let $G = \ell^{(r)}K_n$ and let $Q_n(a, \ell)$ be the probability that the forest building process will terminate with k trees given that $n - a$ vertices have already been chosen in the complete graph portion of G, and there are currently $k - \ell$ trees. Then*

$$Q_n(a,\ell) = \frac{a-1}{2n+2r-a-1} Q_n(a-2,\ell-1) + \frac{2(n-a)}{2n+2r-a-1} Q_n(a-1,\ell) + \frac{2r}{2n+2r-a-1} Q_n(a-1,\ell-1)$$

with

$$Q_n(0,\ell) = \begin{cases} 0 & \ell \neq 0 \\ 1 & \ell = 0. \end{cases}$$

Proof. We first observe two things, we only need to look at what changes the state of the current process, and any edges incident to leaves attached to the $n-a$ picked vertices will automatically be picked in the forest building process without changing counts, so we do not need to consider those edges. Notice that there are exactly three situations that can happen given the current state the process is in.

- We add an edge between a vertex in the set of $n-a$ vertices and the set of a vertices. There are $a(n-a)$ edges where this happens and if it does the set of a vertices loses one vertex and no new tree is formed.
- We add an edge between two vertices in the set of a vertices. There are $\binom{a}{2}$ edges where this happens and if it does, the set of a vertices loses two and one new tree is formed.
- We add an edge between a leaf and a vertex in the set of a vertices. There are ra edges where this happens and if it does, the set of a vertices loses one and one new tree is formed.

Therefore there are a total of $a(n-a) + \binom{a}{2} + ra = a\frac{2n+2r-a-1}{2}$ edges. Thus we get the recurrence

$$
\begin{aligned}
Q_n(a,\ell) = {} & \frac{a-1}{2n+2r-a-1} Q_n(a-2,\ell-1) \\
& + \frac{2(n-m)}{2n+2r-a-1} Q_n(a-1,\ell) \\
& + \frac{2r}{2n+2r-a-1} Q_n(a-1,\ell-1).
\end{aligned}
$$

□

Solving this recurrence is an open problem. The graphs that were previously solved were found by examination of values appearing in the recurrence and then guessing and verifying the solution. These cases had small primes showing up in the recurrences which suggested that they came from multiplication of small terms. What makes it difficult to solve this recurrence is the presence of large primes; for example,

$$
p_{\ell^{(1)}K_7}(x) = \frac{232}{3003}x + \frac{5771}{15015}x^2 + \frac{53938}{135135}x^3 + \frac{16871}{135135}x^4 + \frac{89}{6435}x^5 + \frac{1}{1755}x^6 + \frac{1}{135135}x^7
$$

where 16871 is a prime number. In the following section we show that even though we cannot always solve a general recurrence, we can still use the recurrence to solve certain cases that give us other families of graphs.

CHAPTER 3. THE GRAPHS $\mathbf{K_n - K_t}$

We will first construct the graphs $K_n - K_t$, and again give a probability recurrence that looks at a certain point in the tree building process.

3.1 Construction of $\mathbf{K_n - K_t}$

Given the complete graph K_n and the complete graph K_t with $n \geq t$, let $K_n - K_t$ be the complete graph on n vertices with the edges of K_t removed from it. Figure 3.1 gives an example of this graph.

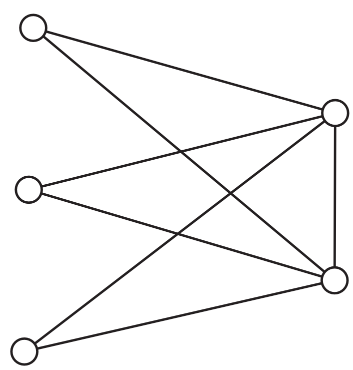

Figure 3.1 The graph $K_5 - K_3$ also denoted $F_{3,2}$

3.2 Probability recurrence for $\mathbf{K_n - K_t}$

Notice that in the construction of the graph, what is left is an independent set on t vertices attached by every possible edge to the complete graph K_{n-t}. We will use this fact to prove the following theorem.

Theorem 3.2.1. *Let $G = K_n - K_t$ with $n \geq t$, and let $K_{t,n-t}(a,b,\ell)$ be the probability that the forest building process will terminate with k trees given that $t-a$ vertices have already been chosen from the t vertices of the independent set, $n-t-b$ vertices have already been chosen from the $n-t$ vertices of K_{n-t}, and there are currently $k-\ell$ trees. Then*

$$\begin{aligned} K_{t,n-t}(a,b,\ell) = & \frac{2ab}{-2ab-2at+2an-b^2-b+2bn} K_{t,n-t}(a-1,b-1,\ell-1) \\ & + \frac{b(b-1)}{-2ab-2at+2an-b^2-b+2bn} K_{t,n-t}(a,b-2,\ell-1) \\ & + \frac{2b(n-a-b)}{-2ab-2at+2an-b^2-b+2bn} K_{t,n-t}(a,b-1,\ell) \\ & + \frac{2a(n-t-b)}{-2ab-2at+2an-b^2-b+2bn} K_{t,n-t}(a-1,b,\ell) \end{aligned}$$

with

$$K_{t,n-t}(a,0,\ell) = K_{t,n-t}(0,1,\ell) = \begin{cases} 0 & \ell \neq 0 \\ 1 & \ell = 0. \end{cases}$$

Proof. We use a similar argument as before. In this graph, there are exactly four situations that can happen given the current state of the forest building algorithm.

- We add an edge between a vertex in the set of a vertices and a vertex in the set with b vertices. There are ab edges where this happens, and if it does, the set of a vertices loses one vertex, the set of b vertices loses one vertex, and one new tree is formed.
- We add an edge between two vertices in the set of b vertices. There are $\binom{b}{2}$ edges where this happens and if it does the set of b vertices loses two vertices, and one new tree is formed.
- We add an edge between a vertex in the set of a vertices and the set of $n-t-b$ vertices. There are $a(n-t-b)$ edges where this happens and if it does the set of a vertices loses one vertex, and no new tree is formed.
- We add an edge between a vertex in the set of b vertices and the set of $n-t-b$ vertices or the set of $t-a$ vertices. There are $b(n-a-b)$ edges where this happens and if it does the set of b vertices loses one vertex, and no new tree is formed.

Therefore there are a total of $ab + \binom{b}{2} + a(n-t-b) + b(n-a-b) = \frac{-2ab-2at+2an-b^2-b+2bn}{2}$ edges. Thus we get the recurrence

$$\begin{aligned} K_{t,n-t}(a,b,l) = &\frac{2ab}{-2ab-2at+2an-b^2-b+2bn} K_{t,n-t}(a-1,b-1,\ell-1) \\ &+ \frac{b(b-1)}{-2ab-2at+2an-b^2-b+2bn} K_{t,n-t}(a,b-2,\ell-1) \\ &+ \frac{2b(n-a-b)}{-2ab-2at+2an-b^2-b+2bn} K_{t,n-t}(a,b-1,\ell) \\ &+ \frac{2a(n-t-b)}{-2ab-2at+2an-b^2-b+2bn} K_{t,n-t}(a-1,b,\ell) \end{aligned}$$

□

Solving this recurrence is also an open question; however, we can use it to find $p_G(x)$ for the Fan graph $F_{n-2,2} = K_n - K_{n-2}$. This polynomial was first found in (3) using a different method. We find it using the above recurrence.

3.3 The Fan graph $\mathbf{F_{n-2,2}}$

Observe that $F_{n-2,2}$ can have at most two trees. Because of this, we only need to find $K_{n-2,2}(n-2,2,1)$, since $K_{n-2,2}(n-2,2,1) + K_{n-2,2}(n-2,2,2) = 1$. In the recurrence for $K_{n-2,2}(n-2,2,1)$, $K_{n-2,2}(n-3,1,0)$ appears, so if we can find a solution to this we are done.

Lemma 3.3.1. *Let $G = K_n - K_{n-m}$, and let $K_{n-m,m}(a,b,\ell)$ be the probability that the forest building process will terminate with k trees given that $n-m-a$ vertices have already been chosen from the $n-m$ vertices of the independent set, $m-b$ vertices have already been chosen from the m vertices of K_m, and there are currently $k-\ell$ trees. Then*

$$K_{n-m,m}(r,1,0) = \frac{n-(-m+r+2)}{n+m-2}.$$

Proof. We prove this for the case when $m=2$ and show $K_{n-2,2}(r,1,0) = \frac{n-r}{n}$. The general case is identical with a bit more bookkeeping. We proceed by induction on r. Let $r=0$,

$K_{n-2,2}(0,1,0) = 1 = \frac{n}{n}$ by definition. Now assume up to $r = k$, $K_{n-2,2}(k,1,0) = \frac{n-k}{n}$. Consider $K_{n-2,2}(k+1,1,0)$. Using the recurrence we have that

$$\begin{aligned} K_{n-2,2}(k+1,1,0) &= \frac{n-k-2}{k+n} K_{n-2,2}(k+1,0,0) + \frac{k+1}{k+n} K_{n-2,2}(k,1,0) \\ &= \frac{n-k-2}{k+n} + \frac{(k+1)(n-k)}{(k+n)n} \\ &= \frac{n-(k+1)}{n}. \end{aligned}$$

Thus by induction, $K_{n-2,2}(r,1,0) = \frac{n-r}{n}$. □

Using this lemma we can prove the following.

Theorem 3.3.2. *(Butler-Hamanaka-Hardt) The probability that the forest building process ends with one tree for* $F_{n-2,2}$ *is*

$$\frac{7n-12}{n(2n-3)}.$$

Proof. To prove this we need to show $K_{n-2,2}(n-2,2,1) = \frac{7n-12}{n(2n-3)}$. In the recurrence we have

$$\begin{aligned} K_{n-2,2}(n-2,2,1) &= \frac{2(n-2)}{2n-3} K_{n-2,2}(n-3,1,0) + \frac{1}{(2n-3)} K_{n-2,2}(n-2,0,0) \\ &= \frac{2(n-2)}{2n-3}\frac{3}{n} + \frac{1}{(2n-3)} \qquad \text{(by Lemma 3.2)} \\ &= \frac{7n-12}{n(2n-3)}. \end{aligned}$$

□

We get the following corollary because $F_{n-2,2}$ can have at most two trees in the forest building process.

Corollary 3.3.3. *For* $G = F_{n-2,2}$,

$$p_G(x) = \frac{7n-12}{n(2n-3)}x + \frac{2(n-2)(n-3)}{n(2n-3)}x^2.$$

Proof. Since in the forest building process $F_{n-2,2}$ can have at most two trees, and the probability the process ends with one tree is $\frac{7n-12}{n(2n-3)}$, the probability the process ends with two trees is $1 - \frac{7n-12}{n(2n-3)} = \frac{2(n-2)(n-3)}{n(2n-3)}$. Thus, $p_G(x) = \frac{7n-12}{n(2n-3)}x + \frac{2(n-2)(n-3)}{n(2n-3)}x^2$. □

3.4 The graphs $\mathbf{K_n - K_{n-3}}$

We again observe that $K_n - K_{n-3}$ can have at most three trees. So we only need to find $K_{n-3,3}(n-3,3,1)$ and $K_{n-3,3}(n-3,3,2)$. We have not yet proven the results for these two cases of the recurrence; however, we do have the following conjecture.

Conjecture 3.4.1. *The probability that the forest building process ends with one tree for $K_n - K_{n-3}$ is*

$$\frac{2(37n^2 - 149n + 144)}{(2n-1)(2n-3)(n+1)(n-2)},$$

and the probability that the forest building process ends with two trees is

$$\frac{(n-3)(42n^2 - 205n + 218)}{(2n-1)(2n-3)(n+1)(n-2)}.$$

This would then give us that for $G = K_n - K_{n-3}$

$$\begin{aligned} p_G(x) = &\frac{2(37n^2 - 149n + 144)}{(2n-1)(2n-3)(n+1)(n-2)}x \\ &+ \frac{(n-3)(42n^2 - 205n + 218)}{(2n-1)(2n-3)(n+1)(n-2)}x^2 \\ &+ \frac{2(n-3)(n-4)(n-5)}{(2n-1)(n+1)(n-2)}x^3. \end{aligned}$$

So far we have proven this conjecture up to $n = 100$. We have also solved a piece of $K_{n-3,3}(n-3,3,2)$.

Lemma 3.4.2. *Let $G = K_n - K_{n-m}$, and let $K_{n-m,m}(a,b,\ell)$ be the probability that the forest building process will terminate with k trees given that $n-m-a$ vertices have already been chosen from the $n-m$ vertices of the independent set, $m-b$ vertices have already been chosen from the m vertices of K_m, and there are currently $k-\ell$ trees. Then*

$$K_{n-m,m}(r,1,1) = \frac{r}{n+m-2}.$$

The proof of Lemma 3.6 is omitted as it is almost identical to the proof of Lemma 3.2. All that remains to be shown for this conjecture to be true is that

$$K_{n-3,3}(n-4,2,0) = \frac{2(29n-46)}{(n+1)(2n-3)(2n-1)}$$

and

$$K_{n-3,3}(n-4,2,1) = \frac{28n^2-197n+215}{(n+1)(2n-3)(2n-1)}.$$

It is also not a coincidence that $K_{n-m,m}(r,1,1)+K_{n-m,m}(r,1,0)=1$. This tells us that if every vertex except one is chosen from the the complete graph portion of the graph, then the process can only end with one or two trees.

3.5 Guessing game

The reader at this point may wonder how these probabilities are found. The answer is looking at a lot of data by coding the recurrences into Sage. By looking at the data for various n values, we can strategically guess what the probability should be. For $K_n - K_{n-3}$, the denominators consistently factored into nice terms that consisted of small primes which let us guess their general form; once we had the denominators, we guessed that the numerators were polynomials in n and used several values to find the coefficients. This method can work well for graphs that cannot have many trees in the forest building process; however, for graphs with many possible trees, a pattern must be found for the a, b and ℓ values.

CHAPTER 4. THE $\mathbf{G_{2n+1}}$ CONJECTURE

We now construct the graph G_{2n+1}, find its probability recurrence, and present a new way of thinking about a previously published conjecture.

4.1 Construction of $\mathbf{G_{2n+1}}$ and its probability recurrence

Let G be the graph on $m+n+r$ vertices where r of the vertices form a complete graph, n of the vertices are an independent set with every possible edge attached to the complete graph on r vertices, and m of the vertices are an independent set with with ever possible edge between the independent set with n vertices. For this graph we have the following.

Theorem 4.1.1. *Let G be the graph described above, and let $Y_{m,n,r}(a,b,c,\ell)$ be the probability that the forest building process will terminate with k trees given that $m-a$ vertices have already been chosen in the the independent set of size m, $n-b$ vertices have already been chosen in the independent set of size n, $r-c$ vertices have already been chosen in K_r, and there are currently $k-\ell$ trees. Let $X = 2cr - c^2 - c + 2cn + 2br - 2bc + 2an + 2bm - 2ba$, then*

$$
\begin{aligned}
Y_{m,n,r}(a,b,c,\ell) = {} & \frac{c(c-1)}{X} Y_{m,n,r}(a,b,c-2,\ell-1) \\
& + \frac{2c(r-c)+2c(n-b)}{X} Y_{m,n,r}(a,b,c-1,\ell) \\
& + \frac{2b(r-c)+2b(m-a)}{X} Y_{m,n,r}(a,b-1,c,\ell) \\
& + \frac{2a(n-b)}{X} Y_{m,n,r}(a-1,b,c,\ell) \\
& + \frac{2cb}{X} Y_{m,n,r}(a,b-1,c-1,\ell-1) \\
& + \frac{2ab}{X} Y_{m,n,r}(a-1,b-1,c,\ell-1)
\end{aligned}
$$

with

$$
Y_{m,n,r}(0,b,0,\ell) = Y_{m,n,r}(a,0,0,\ell) = Y_{m,n,r}(a,0,1,\ell) = \begin{cases} 0 & \ell \neq 0 \\ 1 & \ell = 0. \end{cases}
$$

The proof of this will be omitted since it uses the same basic counting arguments as the previous recurrences.

4.2 Conjectures

To see why we care about the graph G_{2n+1} and its recurrence, we first introduce a conjecture that was published in (1).

Conjecture 4.2.1. *For $n \geq 1$ the graph $G_{2n+1} = K_{n,n+1} + e$, where e is an edge connecting two vertices in the larger part of $K_{n,n+1}$, has the same polynomial as $K_{n,n+1}$.*

Notice that in our above graph, if we let $m = n-1$, and $r = 2$, we get G_{2n+1} described in the above conjecture. This means we can rewrite the conjecture as follows.

Conjecture 4.2.2. *Let G be the graph described above, and let $Y_{n-1,n,2}(a,b,c,\ell)$ be the probability that the forest building process will terminate with k trees given that $n-a-1$ vertices have already been chosen in the the independent set of size $n-1$, $n-b$ vertices* *have already been chosen in the independent set of size n, $2-c$ vertices have already been chosen in K_2, and there are currently $k-\ell$ trees. Then*

$$Y_{n-1,n,2}(n-1,n,2,\ell) = \frac{\binom{n}{\ell}\binom{n}{n-\ell+1}}{\binom{2n}{n+1}}$$

where $\frac{\binom{n}{\ell}\binom{n}{n-\ell+1}}{\binom{2n}{n+1}}$ *is the solution to the $K_{n,n+1}$ recurrence in the forest building process.*

To prove this conjecture we need to show

$$\begin{aligned}
Y_{n-1,n,2}(n-1,n,2,\ell) &= \frac{1}{n^2+n+1}Y_{n-1,n,2}(n-1,n,0,\ell-1) \\
&+ \frac{2n}{n^2+n+1}Y_{n-1,n,2}(n-1,n-1,1,\ell-1) \\
&+ \frac{n(n-1)}{n^2+n+1}Y_{n-1,n,2}(n-2,n-1,2,\ell-1) \\
&= \frac{\binom{n}{\ell}\binom{n}{n-\ell+1}}{\binom{2n}{n+1}}.
\end{aligned}$$

Which means we need to solve the general recurrence for $c = 0, 1, 2$. Although we have not yet proved the conjecture, we have used the recurrence to verify the conjecture up to 23 vertices. We have also solved the $c = 0$ case for the recurrence leaving only two more cases. Interestingly enough, for $c = 0$,

$$\begin{aligned}
Y_{n-1,n,2}(a,b,0,\ell) &= \frac{2b+b(n-1-a)}{-ab+an+bn+b}Y_{n-1,n,2}(a,b-1,0,\ell) \\
&+ \frac{a(n-b)}{-ab+an+bn+b}Y_{n-1,n,2}(a-1,b,0,\ell) \\
&+ \frac{ab}{-ab+an+bn+b}Y_{n-1,n,2}(a-1,b-1,0,\ell-1) \\
&= \frac{\binom{b}{\ell}\binom{2n-b}{a-\ell}}{\binom{2n}{a}}
\end{aligned}$$

which is also the solution to the complete bipartite recurrence in (1). Once we have the case for $c = 1$, we should be able to solve $Y_{n-1,n,2}(n-1,n,2,\ell)$ in a similar way to how we solved $K_{n-3,3}(n-3,3,\ell)$, since we know what the answer should be. We hope to solve this case, and also find a solution for $K_n - K_{n-k}$ in future work.

BIBLIOGRAPHY

[1] Z. Berikkyzy, S. Butler, J. Cummings, K. Heysse, P. Horn, R. Luo, and B. Moran. A forest building process on simple graphs. *Discrete Mathematics*, 341:497–507, 2018.

[2] S. Butler, F. Chung, J. Cummings, and R. Graham. Edge flipping in the complete graph. *Advances in Applied Mathematics*, 69:46–64, 2015.

[3] S. Butler, M. Hamanaka, and M. Hardt. Graphs with at most two trees in a forest building process. *preprint*, 2019.