# On Strongly Multiplicative Graphs

Matthew Rauen

under the direction of
Mr. Jesse Geneson
Massachusetts Institute of Technology

Research Science Institute
July 26, 2011

### Abstract

We analyze the problem of finding the maximal number of edges on a strongly multiplicative graph on $n$ vertices, a problem which has applications to network modeling. Such a graph has $n$ vertices labeled by the integers $1, 2, \ldots, n$ such that if each edge is labeled with the product of adjacent vertices, no two edges have the same label. This value is denoted by $\lambda(n)$, and we construct an analogous function where the two factors, $a$ and $b$, are chosen from sets of differing cardinalities, denoted by $f(x, y)$. We establish the difference function $\delta_f(x, y)$ which is equal to the number of products constructible for the first time as the cardinality of the set of cardinality $y - 1$ is increased by 1. We prove the periodicity and symmetry of $\delta_f(x, y)$ and use it to create a linear approximation for $f(x, y)$ for fixed $x$ in terms of $y$, and prove this approximation is the least squares regression line.



### Summary

We analyze the problem of finding the maximal number of edges on a specific type of graph, a problem which has applications to network modeling. We equate this to finding the number of distinct products that can be made by multiplying two factors together. These factors are taken from separate sets that contain consecutive natural numbers starting at one. By holding the size of one of the sets constant, and adding numbers to the other set, we obtain a sequence of values for the number of products. We then establish a linear approximation for this sequence. Finally, we prove that this approximation is statistically the best possible line to fit the values of the sequence.

# 1 Introduction

Graphs with labeled edges are commonly used to model networks, with restrictions on the network represented as restrictions on the labels of edges. For instance, when modeling transportation networks, such labels can be used to represent a variety of factors, from cost to level of traffic flow. More generally, Ahuja, Magnati and Orlin [1] point out various applications in statistical physics, particle physics, computer science, biology, economics, operations research, and sociology. One of the more classical examples was first introduced by Beineke and Hegde [2], who developed the notion of the *strongly multiplicative graph*. Such a graph has $n$ vertices labeled by the integers $1, 2, \ldots, n$ such that, if each edge is labeled with the product of adjacent vertices, no two edges have the same label. Using this, they then address the question of constructing a strongly multiplicative graph on $n$ vertices with the maximal number of edges, which we denote $\lambda(n)$. See Figures 1 and 2 in the Appendix for examples of strongly multiplicative graphs on 5 and 6 vertices with $\lambda(5)$ and $\lambda(6)$ edges, respectively. Additionally, they give the upper bounds

$$\lambda(4r) \leq 6r^2,$$

$$\lambda(4r+1) \leq 6r^2 + 4r,$$

$$\lambda(4r+2) \leq 6r^2 + 6r + 1,$$

$$\lambda(4r+3) \leq 6r^2 + 10r + 3.$$

However, no nontrivial lower bound for $\lambda(n)$ has been found [3].

# 2 Strongly Multiplicative Graph Notations

We establish notation to more efficiently represent the problem.

Define $[n] = \{1, 2, \ldots, n\}$, and let $\text{lcm}[n]$ be the least common multiple of all integers from 1 to $n$.

The set

$$A_n = \{z | z = ab, \quad a, b \in [n], \quad a \neq b\}$$

represents all possible edge labels on a strongly multiplicative graph on $n$ vertices. Since $\lambda(n)$ is equal to the maximal number of distinct edge labels on a strongly multiplicative graph, $\lambda(n) = |A_n|$.

We next construct the analogous function where the two factors are taken from separate sets $[x]$ and $[y]$. We denote the set of all such products by

$$A_{x,y} = \{z | z = ab, \quad a \in [x], \quad b \in [y], \quad a \neq b\},$$

and, as above, $\lambda(x, y) = |A_{x,y}|$. This number is the maximal number of distinct edge labels for a matching constructed between $x$ points labeled $1, \ldots, x$ and $y$ points labeled $1, \ldots, y$ such that no two points of the same label have a connecting edge.

We also define $\delta(n) = \lambda(n) - \lambda(n-1)$, which is the number of new edge labels that can be constructed by adding an $n$th vertex. Similarly, let $\delta(x, y) = \lambda(x, y) - \lambda(x, y-1)$ be the number of new edge labels that can be constructed when adding a $y$th vertex to a matching between $x$ vertices and $y - 1$ vertices.

We define a similar set, in which we allow the factors $a, b$ to be equal. Let

$$C_{x,y} = \{z | z = ab, \quad a \in [x], \quad b \in [y]\}.$$

Then $f(x, y) = |C_{x,y}|$ is the maximal number of edge labels allowing indistinct factors. Again, let $\delta_f(x, y) = f(x, y) - f(x, y-1)$, analogous to the difference function for $\lambda(x, y)$.

Next, we would like to count those edge labels in $C_{x,y}$ which are solely constructible by using the same vertex label twice. If

$$D_{x,y} = \{z|z = a^2, \quad a \in [x] \cap [y], \quad \nexists b, c, \quad b \in [x], \quad c \in [y], \quad b \neq c, \quad z = bc\},$$

then $g(x, y) = |D_{x,y}|$ is the desired count.

By studying the function $\delta_f(x, y)$ and how it determines $f(x, y)$, and also analyzing $g(x, y)$, we can establish general bounds for $\lambda(x, y)$, since by definition, $\lambda(x, y) = f(x, y) - g(x, y)$. These bounds can be modified for the specific case of $\lambda(n)$.

# 3 Properties of $\delta(n)$

Recall that $\delta(n)$ is the number of new edge labels in which $n$ is one of the factors. We begin by examining the special case of $\delta(p)$ where $p$ is prime.

**Proposition 1.** *If $p$ is a prime, then $\delta(p) = p - 1$.*

*Proof.* We can construct $p - 1$ edges between the $p$th vertex and one vertex from 1 through $p - 1$. If $k \leq p - 1$ such that $kp$ already exists as a label for some edge in a graph with $p - 1$ vertices. Then $ab = kp$ for some $a, b < p$, which is impossible. Therefore, there does not exist such a vertex $k$. Hence, all such edges yield new edge labels, and we can form $p - 1$ new edges. Thus, $\delta(p) = p - 1$. □

Now, we construct a lower bound for $\delta(n)$ when $n$ is composite.

**Proposition 2.** *If $p$ is a prime and $q \geq 2$ is an integer, then $\delta(pq) \geq p$.*

*Proof.* We can construct $p$ edges between the $pq$th vertex and a vertex labeled between $pq - p$ and $pq - 1$. Suppose that there exists some vertex $k$ such that $pq - p \leq k \leq pq - 1$ and $kpq$ already exists as a label for some edge with $pq - 1$ vertices. Then $ab = kpq$. Since $a < pq$,

$b > k$, and by symmetry, $a > k$. This implies that $p \nmid a$ and $p \nmid b$, which is a contradiction. Hence, all $p$ of these edges yield new edge labels, so $\delta(pq) \geq p$. □

The strongest bound that can be given by Proposition 2 is presented in Corollary 1.

**Corollary 1.** *Let $n$ be composite with greatest prime divisor $r$. Then $\delta(n) \geq r$.*

# 4 Properties of $\delta_f(x, y)$

When considering the function $\lambda(n)$ as $n$ increases, both of the cardinalities of the sets of possible values for each factor are increasing. Therefore, we simplify the problem by fixing the cardinality of one of the sets, which gives $\lambda(x, y)$. Furthermore, we have the weak bound $g(x, y) \leq \min(x, y)$ so we focus on finding a lower bound for $f(x, y)$.

By definition, $f(x, y) = \sum_{i=1}^{y} \delta_f(x, i)$, so we analyze $\delta_f(x, y)$ to decompose $f(x, y)$.

## 4.1 Characterizing the Structure of $\delta_f(x, y)$

First, we analyze when the minimal values of $\delta_f(x, y)$, or equivalently, the number of products $y, 2y, 3y, \ldots, xy$ that are not expressible as $ab$ where $a \in [x]$ and $b \in [y-1]$, for a given $x$.

**Proposition 3.** *$x|y$ if and only if $\delta_f(x, y) = 1$.*

*Proof.* Assume $x|y$. The maximal product in $C_{x,y-1}$ is $x(y-1)$. Let $ky$ be a product in $C_{x,y-1}$ for $k \leq x-1$. Then $\frac{ky}{x} \in \mathbb{Z}$ and $\frac{ky}{x} < y$. Therefore, $ky \in C_{x,y-1}$ for $k \leq x-1$, but $xy \notin C_{x,y-1}$ so $\delta_f(x, y) = 1$.

Assume $\delta_f(x, y) = 1$. Since $xy \notin C_{x,y-1}$, $(x-1)y \in C_{x,y-1}$. Hence, there exists some $a \in [x]$ and some $b \in [y-1]$ such that $ab = (x-1)y$. $a = \frac{(x-1)y}{b}$ and $b < y \implies a > x-1$, so $a = x$. Then $x|(x-1)y$, and since $\gcd(x, x-1) = 1$, $x|y$. □

Next, we analyze where the maximal values of $\delta_f(x, y)$ for given $x$ occur.

**Proposition 4.** $\gcd(\text{lcm}[x], y) = 1$ *if and only if* $\delta_f(x, y) = x$.

*Proof.* Assume $\gcd(\text{lcm}[x], y) = 1$. Suppose there exists some product $ky$ such that $ky \in C_{x,y-1}$. Then $ky = ab$ for some $a \in [x]$ and $b \in [y-1]$. Since $\gcd(\text{lcm}[x], y) = 1$, $\gcd(a, y) = 1$ so $y|b$. This is a contradiction. Hence, $\delta_f(x, y) = x$.

Assume $\delta_f(x, y) = x$. Then $y$ is not an element of $C_{x,y-1}$, which implies that $\nexists a, b$ such that $a \in [x]$ and $b \in [y-1]$ where $y = ab$. Suppose there exists some prime $p$ such that $p|\gcd(\text{lcm}[x], y)$. Then $\frac{y}{p} \in \mathbb{Z}$ and $\frac{y}{p} < y$, which implies that $y \in C_{x,y-1}$, so we have a contradiction. Thus, $\gcd(\text{lcm}[x], y) = 1$. □

Now that we have characterized where the minimal and maximal values of $\delta_f(x, y)$ occur for some given $x$, we continue our analysis by characterizing the structure of $\delta_f(x, y)$ more generally. We prove that it is periodic for fixed $x$ and symmetric within its period.

**Proposition 5.** *For all positive integers* $x, y$, $\delta_f(x, y) = \delta_f(x, \text{lcm}[x] \pm y)$ *holds.*

*Proof.* Consider the products $ky$, where $k \in [x]$ and $ky \in C_{x,y-1}$. Since $\delta_f(x, y)$ is the number of products $ly$ where $l \in [x]$ that are not elements of $C_{x,y-1}$, there are $x - \delta_f(x, y)$ such products $ky$. For each $ky$, there exists some $a$ such that $a|ky$, $a \in [x]$, and $\frac{ky}{a} < y$. We then consider the corresponding product $k(\text{lcm}[x] \pm y)$. Since $a \in [x]$, $a|\text{lcm}[x]$, so $a|k(\text{lcm}[x] \pm y)$. Additionally, $\frac{ky}{a} < y \implies \frac{k(\text{lcm}[x] \pm y)}{a} < \text{lcm}[x] \pm y$. Therefore, $k(\text{lcm}[x] \pm y) \in C_{x,\text{lcm}[x] \pm y-1}$. Hence, at least $x - \delta_f(x, y)$ of the products of the form $k(\text{lcm}[x] \pm y)$ where $k \in [x]$ are elements of $C_{x,\text{lcm}[x] \pm y-1}$, so $\delta_f(x, \text{lcm}[x] \pm y)$ is at most $x - (x - \delta_f(x, y)) = \delta_f(x, y)$. Hence, $\delta_f(x, y) \geq \delta_f(x, \text{lcm}[x] \pm y)$.

By a similar argument, $\delta_f(x, \text{lcm}[x] \pm y) \geq \delta_f(x, y)$.

Therefore, $\delta_f(x, y) = \delta_f(x, \text{lcm}[x] \pm y)$. □

### 4.2 Properties of $\delta_f(x, y)$ for Fixed $y$

Now we analyze the behavior of $\delta_f(x, y)$ for fixed $y$, rather than for fixed $x$, because this will potentially allow for a recursive method of defining $f(x, y)$. We begin by bounding the values of $\delta_f(x+1, y)$ with $\delta_f(x, y)$.

**Proposition 6.** *For all positive integers $x, y$, $\delta_f(x, y) + 1 \geq \delta_f(x+1, y)$ holds.*

*Proof.* Consider the products $ky$ where $k \in [x]$ and $ky \in C_{x,y-1}$. As shown in the proof of Proposition 5, there are $x - \delta_f(x, y)$ such products. For any such product $ky$, $\exists a \in [x]$ and $\exists b \in [y-1]$ such that $ab = ky$. Now consider the products $y, 2y, 3y, \ldots, xy, xy + y$. We analyze the number of these products that are elements of $C_{x+1,y-1}$, that is, for how many $ly$ where $l \leq x + 1$, $\exists c \in [x+1]$ and $\exists d \in [y-1]$ such that $cd = ly$. The number of these products that are elements of $C_{x+1,y-1}$ is at least $x - \delta_f(x, y)$, so the number of elements of $C_{x+1,y}$ that are not elements of $C_{x+1,y-1}$, or, equivalently, $\delta_f(x+1, y)$, is at most $(x+1) - (x - \delta_f(x, y)) = \delta_f(x, y) + 1$. Hence, $\delta_f(x, y) + 1 \geq \delta_f(x+1, y)$. □

Now we establish some instances in which equality holds for the inequality established in Proposition 6.

**Proposition 7.** $\gcd(x+1, y) = 1 \implies \delta_f(x, y) + 1 = \delta_f(x+1, y)$.

*Proof.* Assume $\gcd(x+1, y) = 1$. Consider the products $y, 2y, 3y, \ldots, xy$. Exactly $x - \delta_f(x, y)$ of these are elements of $C_{x,y-1}$. Now consider the products $y, 2y, 3y, \ldots, xy, xy + y$. At least $x - \delta_f(x, y)$ of these are elements of $C_{x+1,y-1}$ because the same factors can be used. Now we consider the case where $a = x + 1$. If $(x+1)b = ky$ for some $k \leq x + 1$, then $x + 1 | ky$, and since $\gcd(x+1, y) = 1$, $x + 1 | k$. Hence, $k = x + 1$. Then, $b = y$, which is a contradiction, so all of the remaining $\delta_f(x, y) + 1$ products are not elements of $C_{x+1,y-1}$. Thus, $\delta_f(x, y) + 1 = \delta_f(x+1, y)$. □

We now consider the sum of the values of $\delta_f(k, y)$ as $k$ goes from 1 to $m-1$ where $m$ is a multiple of lcm$[x]$. This is a special case of the analogue to the function $f(x, y)$ that is formed by summing the values of $\delta_f(x, k)$ but for fixed $x$ and with $k$ going from 1 to $y$ instead.

**Lemma 1.** *If* lcm$[n]$ *divides* $m$ *then* $\sum_{i=1}^{m-1} \delta_f(i, n) = \frac{m}{2}\delta_f(m-1, n)$.

*Proof.* By definition, $\delta_f(i, n) = f(i, n) - f(i, n-1)$. Therefore,

$$\sum_{i=1}^{m-1} \delta_f(i, n) = \sum_{i=1}^{m-1} f(i, n) - f(i, n-1).$$

By the symmetry of $f(x, y)$,

$$\sum_{i=1}^{m-1} \delta_f(i, n) = \sum_{i=1}^{m-1} f(n, i) - f(n-1, i). \tag{1}$$

We then separate $\sum_{i=1}^{m-1} f(n, i) - f(n-1, i)$ into two summations, $\sum_{i=1}^{m-1} f(n, i)$ and $\sum_{i=1}^{m-1} f(n-1, i)$. By definition,

$$\sum_{i=1}^{m-1} f(n, i) = \sum_{i=1}^{m-1} \sum_{j=1}^{i} \delta_f(n, j). \tag{2}$$

In this expression, $\delta_f(n, z)$ for $1 \leq z \leq m$ will appear $m - z$ times, so

$$\sum_{i=1}^{m-1} \sum_{j=1}^{i} \delta_f(n, j) = \sum_{i=1}^{m-1} (m-i)\delta_f(n, i). \tag{3}$$

By Propositions 3 and 4, lcm$[n]|m \implies \delta_f(n, i) = \delta_f(n, m-i)$, so

$$\sum_{i=1}^{m-1} (m-i)\delta_f(n, i) = \sum_{i=1}^{m-1} \frac{m}{2}\delta_f(n, i) = \frac{m}{2} f(n, m-1). \tag{4}$$

By equations (2), (3), and (4),

$$\sum_{i=1}^{m-1} f(n,i) = \frac{m}{2} f(n, m-1). \tag{5}$$

Similarly,

$$\sum_{i=1}^{m-1} f(n-1,i) = \frac{m}{2} f(n-1, m-1). \tag{6}$$

By combining equations (1), (5), and (6),

$$\sum_{i=1}^{m-1} \delta_f(i,n) = \frac{m}{2}(f(n,m-1) - f(n-1,m-1)).$$

By the symmetry of $f(x,y)$,

$$\sum_{i=1}^{m-1} \delta_f(i,n) = \frac{m}{2}(f(m-1,n) - f(m-1,n-1)).$$

So by definition, we have that

$$\sum_{i=1}^{m-1} \delta_f(i,n) = \frac{m}{2} \delta_f(m-1,n).$$

□

### 4.3 Establishing the Average Value of $\delta_f(x,y)$

We define $\overline{x}$ by $\overline{x} = \lim_{y\to\infty} \dfrac{f(x,y)}{y}$ if such a limit exists. Analyzing this value allows us to obtain a linear approximation for $f(x,y)$ by using the equation $\hat{y} = \overline{x}y + x_c$ for some constant $x_c$ that is dependent on $x$. We begin by establishing that this limit does exist and obtain a formula for $\overline{x}$.

**Theorem 1.** $\overline{x} = \frac{f(x,lcm[x])}{\mathrm{lcm}[x]}$.

*Proof.* By definition, we can rewrite $\lim_{y\to\infty} \frac{f(x,y)}{y}$ as

$$\lim_{y\to\infty} \frac{\sum_{i=1}^{y} \delta_f(x,i)}{y}.$$

Since the values of $\delta_f(x,i)$ are periodic by Proposition 5, and are bounded by the inequality $1 \leq \delta_f(x,y) \leq x$, this limit is the average value of the period, so

$$\overline{x} = \frac{\sum_{i=1}^{\text{lcm}[x]} \delta_f(x,i)}{\text{lcm}[x]} = \frac{f(x,\text{lcm}[x])}{\text{lcm}[x]}.$$

□

Now that we have established that the average value can be determined by $\overline{x} = \frac{f(x,\text{lcm}[x])}{\text{lcm}[x]}$, we use a result given in Lemma 1 to create an expression for $\overline{n}$.

**Lemma 2.** *If* lcm$[n]$ *divides* $m$*, then* $\overline{n} = \frac{1}{m} + \frac{2\sum_{i=1}^{m-1} f(n,i)}{m^2}$.

*Proof.* Recall that the proof of Lemma 1 showed that

$$\sum_{i=1}^{m-1} f(n,i) = \frac{m}{2} f(n,m-1).$$

as this was equation (5). Multiplying through by $\frac{2}{m}$ gives

$$\frac{2}{m}\sum_{i=1}^{m-1} f(n,i) = f(n,m-1). \tag{7}$$

By Proposition 3, $\delta_f(n,m) = 1$, so $f(n,m) = f(n,m-1) + 1$. By combining this result with equation (7),

$$f(n,m) = 1 + \frac{2}{m}\sum_{i=1}^{m-1} f(n,i).$$

Since $\overline{n} = \frac{f(n,m)}{m}$, we divide through by $m$ to obtain the desired result of

$$\overline{n} = \frac{1}{m} + \frac{2\sum_{i=1}^{m-1} f(n,i)}{m^2}.$$

□

We consider the difference between the function $f(x,y)$ and our approximation, $\overline{x}y$, and prove that these differences are periodic for fixed $x$.

**Proposition 8.** *For all positive integers $x, y$, $f(x,y) - \overline{x}y = f(x, \text{lcm}[x] + y) - \overline{x}(\text{lcm}[x] + y)$ holds.*

*Proof.* Note that

$$f(x, \text{lcm}[x] + y) = \sum_{i=1}^{\text{lcm}[x]+y} \delta_f(x,i),$$

and by Proposition 5,

$$\sum_{i=1}^{\text{lcm}[x]+y} \delta_f(x,i) = \sum_{i=1}^{\text{lcm}[x]} \delta_f(x,i) + \sum_{i=1}^{y} \delta(x,i) = f(x, \text{lcm}[x]) + f(x,y).$$

Hence, by Theorem 1,

$$f(x, \text{lcm}[x] + y) - \overline{x}(\text{lcm}[x] + y) = f(x, lcm[x]) + f(x,y) - \overline{x}(\text{lcm}[x] + y) = f(x,y) - \overline{x}y.$$

□

We use the periodicity of the error values to establish the average error, which we will then shift our approximation by to get a more accurate linear regression.

**Theorem 2.** *The average value of $f(n,i) - \overline{n}i$ for fixed $n$ is $\frac{\overline{n}-1}{2}$.*

*Proof.* We wish to find

$$\lim_{x\to\infty} \frac{\sum_{i=1}^{x} f(n,i) - \overline{n}i}{x}.$$

By Proposition 8, these values are periodic and bounded, so this limit is equal to

$$\frac{\sum_{i=1}^{\mathrm{lcm}[n]} f(n,i) - \overline{n}i}{\mathrm{lcm}[n]}. \tag{8}$$

Algebraic manipulation on the result given in Lemma 2 gives that

$$\sum_{i=1}^{\mathrm{lcm}[n]-1} f(n,i) = \frac{\mathrm{lcm}[n]^2\overline{n} - \mathrm{lcm}[n]}{2}.$$

By Theorem 1,

$$\sum_{i=1}^{\mathrm{lcm}[n]} f(n,i) = \frac{\mathrm{lcm}[n]^2\overline{n} - \mathrm{lcm}[n]}{2} + \mathrm{lcm}[n]\overline{n}. \tag{9}$$

Additionally, we rewrite the second portion of the summation as

$$\sum_{i=1}^{\mathrm{lcm}[n]} \overline{n}i = \frac{\mathrm{lcm}[n](\mathrm{lcm}[n]+1)}{2}\overline{n} = \frac{\mathrm{lcm}[n](\mathrm{lcm}[n]\overline{n}+\overline{n})}{2}. \tag{10}$$

We substitute equations (9) and (10) into equation 8 to obtain

$$\frac{\sum_{i=1}^{\mathrm{lcm}[n]} f(n,i) - \overline{n}i}{\mathrm{lcm}[n]} = \frac{\frac{\mathrm{lcm}[n]^2\overline{n}-\mathrm{lcm}[n]}{2} + \mathrm{lcm}[n]\overline{n} - \frac{\mathrm{lcm}[n](\mathrm{lcm}[n]\overline{n}+\overline{n})}{2}}{\mathrm{lcm}[n]} = \frac{\overline{n}-1}{2}$$

□

# 5 A Linear Approximation for $f(x,y)$

We now take our approximation of $\hat{y} = \overline{x}y + \frac{\overline{x}-1}{2}$ and prove that this is the line of best fit with respect to $y$. See Figures 3 and 4 in the Appendix for examples of this approximation.

**Theorem 3.** *$\hat{y} = \overline{x}y + \frac{\overline{x}-1}{2}$ is the least squares regression line for $f(x,y)$ with respect to $y$.*

*Proof.* First, we show that the slope of the least squares regression line is equal to $\overline{x}$. Suppose that the slope is actually $\overline{x}+\epsilon$ for some $\epsilon \neq 0$, and we let the approximation be $\hat{y} = (\overline{x}+\epsilon)y+c$ for some constant $c$. By Proposition 8, we note that

$$\sum_{i=1}^{k\,\mathrm{lcm}[x]} \left(f(x,i) - \left(\overline{x}i + \frac{\overline{x}-1}{2}\right)\right)^2 = k \sum_{i=1}^{\mathrm{lcm}[x]} \left(f(x,i) - \left(\overline{x}i + \frac{\overline{x}-1}{2}\right)\right)^2.$$

Since the sum on the right hand side is finite, let

$$d = \sum_{i=1}^{\mathrm{lcm}[x]} \left(f(x,i) - \left(\overline{x}i + \frac{\overline{x}-1}{2}\right)\right)^2.$$

Then we have that

$$\sum_{i=1}^{k\,\mathrm{lcm}[x]} \left(f(x,i) - \left(\overline{x}i + \frac{\overline{x}-1}{2}\right)\right)^2 = kd.$$

If we then consider the sum of the squares of the errors for the other approximation, it is clear that

$$\sum_{i=1}^{k\,\mathrm{lcm}[x]} \left(f(x,i) - \left(\left(\overline{x}+\epsilon\right)i + c\right)\right)^2 \geq \left(f(x,k\,\mathrm{lcm}[x]) - \left(\left(\overline{x}+\epsilon\right)k\,\mathrm{lcm}[x] + c\right)\right)^2.$$

With a bit of algebra and the result from Theorem 1, we obtain that this is equivalent to

$$\sum_{i=1}^{k\,\mathrm{lcm}[x]} \left(f(x,i) - \left(\left(\overline{x}+\epsilon\right)i + c\right)\right)^2 \geq (k\epsilon\,\mathrm{lcm}[x] + c)^2.$$

It is clear that $(k\epsilon\,\mathrm{lcm}[x] + c)^2 > k(d+1)$ for sufficiently large $k$, which means that the regression line with slope $\overline{x} + \epsilon$ is not the line of best fit.

Next, we show that the constant $\frac{\overline{x}-1}{2}$ is the optimal constant for the line of best fit. We begin by demonstrating that for every error of $\epsilon$, there exists a corresponding error of $-\epsilon$. We only show this for the first $\text{lcm}[x]$ errors, since by Proposition 8, these values are periodic.

First, note that

$$f(x, \text{lcm}[x]) - \left(\overline{x}\,\text{lcm}[x] + \frac{\overline{x}-1}{2}\right) = -\left(f(x, \text{lcm}[x]-1) - \left(\overline{x}(\text{lcm}[x]-1) + \frac{\overline{x}-1}{2}\right)\right),$$

since by Theorem 1, the left hand side is equivalent to $\frac{1-\overline{x}}{2}$. Also using Proposition 3, we know that $\delta_f(x, \text{lcm}[x]) = 1$, or equivalently, $f(x, \text{lcm}[x]-1) = f(x, \text{lcm}[x]) - 1$, the right hand side is also equivalent to $\frac{1-\overline{x}}{2}$.

Next, note that

$$f(x,k) - \left(\overline{x}k + \frac{\overline{x}-1}{2}\right) = -\left(f(x, \text{lcm}[x]-1-k) - \left(\overline{x}(\text{lcm}[x]-1-k) + \frac{\overline{x}-1}{2}\right)\right). \quad (11)$$

This is because $f(x,k) + f(x, \text{lcm}[x]-1-k) = \sum_{i=1}^{k} \delta_f(x,i) + \sum_{i=1}^{\text{lcm}[x]-1-k} \delta_f(x,i)$. By Proposition 5, we can rewrite this further as $f(x,k) + f(x, \text{lcm}[x]-1-k) = \sum_{i=\text{lcm}[x]-k}^{\text{lcm}[x]-1} \delta_f(x,i) + \sum_{i=1}^{\text{lcm}[x]-1-k} \delta_f(x,i) = \sum_{i=1}^{\text{lcm}[x]-1} \delta_f(x,i) = f(x, \text{lcm}[x]-1)$.

Additionally, $\left(\overline{x}(\text{lcm}[x]-1-k) + \frac{\overline{x}-1}{2}\right) + \left(\overline{x}k + \frac{\overline{x}-1}{2}\right) = \overline{x}\,\text{lcm}[x] - 1$.

By Proposition 3 and Theorem 1, we have that $f(x, \text{lcm}[x]-1) = \overline{x}\,\text{lcm}[x]-1$, so equation (11) holds.

Therefore every error $\epsilon$ has a corresponding error $-\epsilon$. If we add a constant $a$ to the current constant of $\frac{\overline{x}-1}{2}$, then the new errors become $\epsilon - a$ and $-\epsilon - a$. The sum of the squares of these new errors is $2\epsilon^2 + 2a^2$, which is strictly greater than the sum of the squares of the old errors, $2\epsilon^2$, assuming $a \neq 0$. Hence, the constant we have chosen is the optimal constant for the least squares regression line. □

## 6 A Recursive Formula for the Average Value for Primes

Now that we have established the least squares regression line for $f(x, y)$ in terms of $\overline{x}$, we create a recursive definition for $\overline{p}$ where $p$ is a prime.

**Proposition 9.** *For $p$ a prime, $\overline{p} = \frac{(p-1)\overline{p-1}}{p} + 1$.*

*Proof.* Consider $f(p, \mathrm{lcm}[p])$. By definition,

$$f(p, \mathrm{lcm}[p]) = \sum_{i=1}^{\mathrm{lcm}[p]} \delta_f(p, i) = \sum_{i=1}^{p\,\mathrm{lcm}[p-1]} \delta_f(p, i).$$

Now we rewrite $\sum_{i=1}^{p\,\mathrm{lcm}[p-1]} \delta_f(p, i)$ as two summations by separating those indices which are divisible by $p$ into the second summation. This yields

$$f(p, \mathrm{lcm}[p]) = \sum_{i=1}^{\mathrm{lcm}[p-1]} \sum_{j=(i-1)p+1}^{ip-1} \delta_f(p, j) + \sum_{i=1}^{\mathrm{lcm}[p-1]} \delta(p, pi).$$

Suppose that for some $i$, $p \nmid i$. Then $\gcd(p, i) = 1$, so by Lemma 1, $\delta_f(p, i) = \delta_f(p-1, i) + 1$. If $p|i$, then by Proposition 3, $\delta_f(p, i) = 1$. Hence,

$$f(p, \mathrm{lcm}[p]) = \sum_{i=1}^{\mathrm{lcm}[p-1]} \sum_{j=(i-1)p+1}^{ip-1} (\delta_f(p-1, j) + 1) + \sum_{i=1}^{\mathrm{lcm}[p-1]} 1.$$

Since there are in total, $\mathrm{lcm}[p]$ terms, we can remove all of the 1's and rewrite this as

$$f(p, \mathrm{lcm}[p]) = \mathrm{lcm}[p] + \sum_{i=1}^{\mathrm{lcm}[p-1]} \sum_{j=(i-1)p+1}^{ip-1} \delta_f(p-1, j). \tag{12}$$

Now consider the integers $j, j+\mathrm{lcm}[p-1], \ldots, j+(p-1)\,\mathrm{lcm}[p-1]$ such that $1 \leq j \leq \mathrm{lcm}[p-1]$. Since $p \nmid \mathrm{lcm}[p-1]$, all $p$ of these integers have different residues modulo $p$. Therefore, exactly $(p-1)$ of these have residues that are nonzero modulo $p$. Let $k$ be one of these $p-1$ integers.

Then $\delta_f(p-1,k)$ will be included in $\sum_{i=1}^{\text{lcm}[p-1]} \sum_{j=(i-1)p+1}^{ip-1} \delta_f(p-1,j)$. Additionally, by Proposition 5, for all such $k$, $\delta_f(p-1,k) = \delta_f(p-1,j)$. Therefore,

$$\sum_{i=1}^{\text{lcm}[p-1]} \sum_{j=(i-1)p+1}^{ip-1} \delta_f(p-1,j) = (p-1) \sum_{i=1}^{\text{lcm}[p-1]} \delta(p-1,i).$$

Then, by definition, we have

$$\sum_{i=1}^{\text{lcm}[p-1]} \sum_{j=(i-1)p+1}^{ip-1} \delta_f(p-1,j) = (p-1) f(p-1, \text{lcm}[p-1]). \tag{13}$$

By combining equations 7 and 8, we obtain

$$f(p, \text{lcm}[p]) = (p-1) f(p-1, \text{lcm}[p-1]) + \text{lcm}[p]. \tag{14}$$

By Theorem 1,

$$\overline{p} = \frac{f(p, \text{lcm}[p])}{\text{lcm}[p]}. \tag{15}$$

By combining equations 9 and 10,

$$\overline{p} = \frac{(p-1) f(p-1, \text{lcm}[p-1]) + \text{lcm}[p]}{\text{lcm}[p]} = \frac{(p-1) f(p-1, \text{lcm}[p-1])}{p \, \text{lcm}[p-1]} + 1.$$

So, by Theorem 1,

$$\overline{p} = \frac{(p-1)\overline{p-1}}{p} + 1.$$

□

# 7 Conclusion

Since $\lambda(n)$ is a special case of $\lambda(x, y)$, which is equal to $f(x, y) - g(x, y)$, and we have given an upper bound for $g(x, y)$, determining a lower bound for $f(x, y)$ would allow for the construction of a lower bound on $\lambda(x, y)$, and by extension, $\lambda(n)$. We have analyzed $\delta_f(x, y)$, the difference function of $f(x, y)$ as $x$ is held constant, and concluded that this function is both periodic and symmetric within this period. More specifically, we have characterized the maximal and minimal values of $\delta_f(x, y)$ and given their exact locations in the period. Finally, we have established the least squares regression line for $f(x, y)$ that involves the average value of $\delta_f(x, y)$ and proved a recursive relation for obtaining the average value of the period for primes $p$. This is in terms of the average value of the period for $p - 1$ and uses proven relationships between $\delta_f(x - 1, y)$ and $\delta_f(x, y)$ in general, and for the specific case in which $\gcd(x, y) = 1$. Further analysis either in how $f(x, y)$ can be bounded based on $f(x - 1, y)$ or on the values of $\overline{x}$ would potentially be able to obtain a strong lower bound on the function.

# 8 Acknowledgments

I would like to thank my mentor, Jesse Geneson (MIT) for his consistent guidance. I would also like to thank Dr. John Rickert (Rose-Hulman Institute of Technology), Dr. Kartik Venkatram (MIT), Dr. Tanya Khovanova (MIT), and Chris Olund for their advice in improving the quality of my paper. I must additionally acknowledge Dr. David Jerison (MIT), the Center for Excellence in Education, the Research Science Institute, and the MIT math department for making this research possible. Finally, I would like to thank my sponsors, SAP America, the Petroleum Service Company, and FinAid.org, without whom I would not have been able to conduct this research.

## References

[1] R. K. Ahuja, T. L. Magnanti, and J. B. Orlin. Network Flows: Theory, Algorithms, and Applications. Prentice Hall, 1993.

[2] L. W. Beineke and S. M. Hegde. Strongly multiplicative graphs. Discuss. Math. Graph Theory 21 (2001), 63–75.

[3] J. A. Gallian. A Dynamic Survey of Graph Labeling. The Electronic Journal of Combinatorics 5 (2005), DS6. http://www.combinatorics.org/Surveys/ds6.pdf.

[4] J. B. Rosser and L. Schoenfeld. Approximate formulas for some functions of prime numbers. Illinois Journal of Mathematics 6 (1962), no. 1, 64–94.

[5] S. K. Vaidya and K. K. Kanani. Some Strongly Multiplicative Graphs in the Context of Arbitrary Supersubdivision. International Journal of Applied Mathematics and Computation 3 (2011), no. 1, 60–64.

# A Appendix

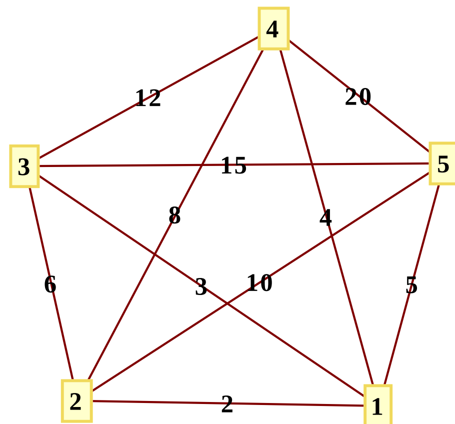


Figure 1: A strongly multiplicative graph on 5 vertices with $\lambda(5)$ edges.

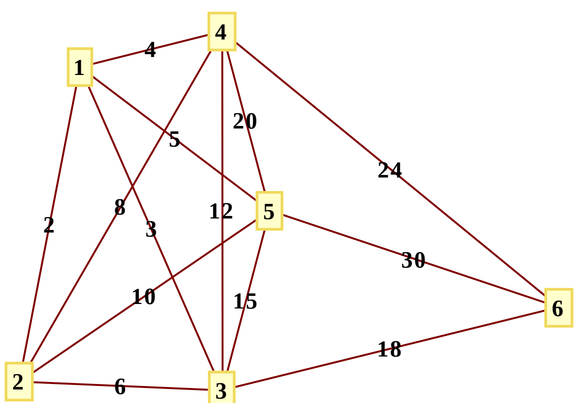


Figure 2: A strongly multiplicative graph on 6 vertices with $\lambda(6)$ edges.

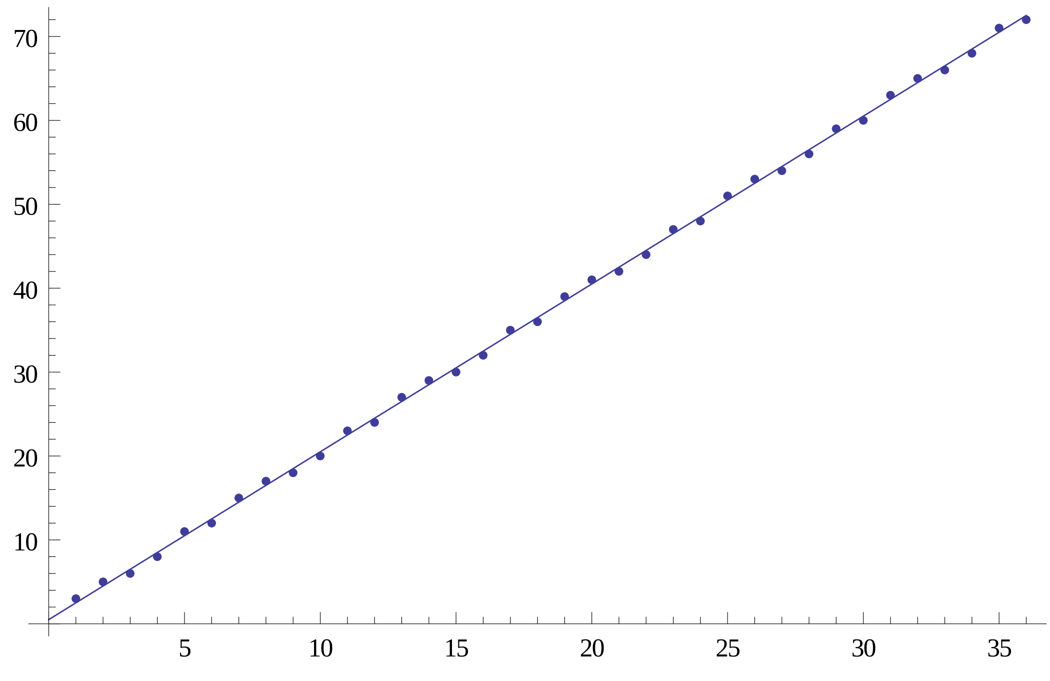


Figure 3: The least squares regression line for $f(3, y)$ vs. $y$.

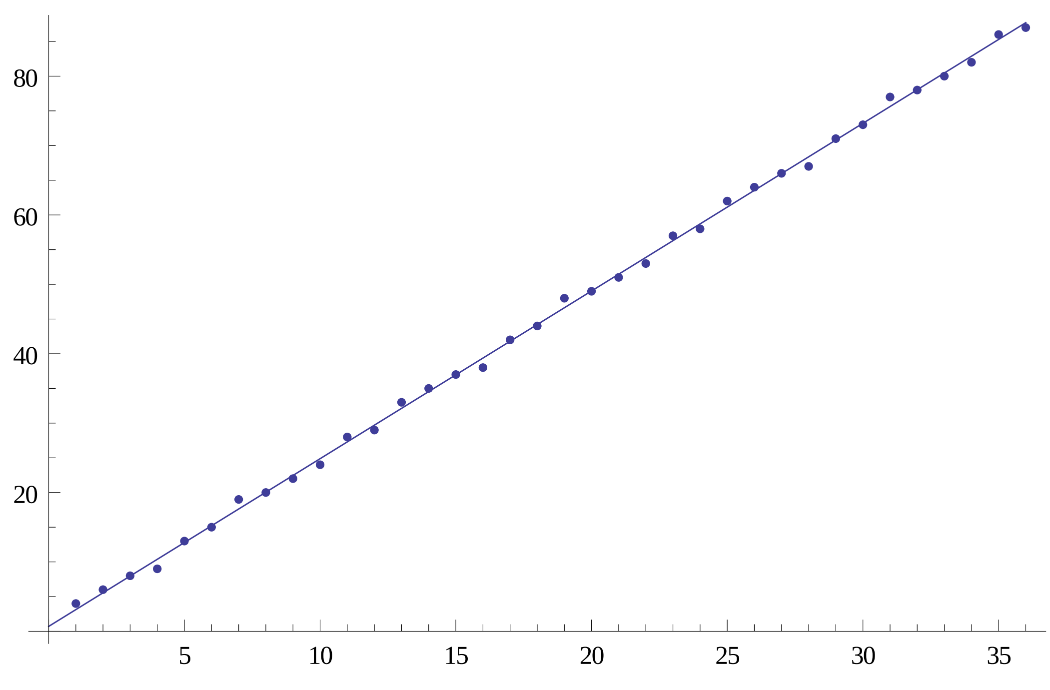


Figure 4: The least squares regression line for $f(4, y)$ vs. $y$.