

Bounds for elements of the degree sequence of an unknown vertex set in a balanced bipartite graph

Sam Pine
Elmira College, spine14@elmira.edu

Follow this and additional works at: <https://scholar.rose-hulman.edu/rhumj>

Recommended Citation

Pine, Sam (2013) "Bounds for elements of the degree sequence of an unknown vertex set in a balanced bipartite graph," *Rose-Hulman Undergraduate Mathematics Journal*: Vol. 14 : Iss. 1 , Article 11.
Available at: <https://scholar.rose-hulman.edu/rhumj/vol14/iss1/11>

ROSE-
HULMAN
UNDERGRADUATE
MATHEMATICS
JOURNAL

BOUNDS FOR ELEMENTS OF THE
DEGREE SEQUENCE OF AN UNKNOWN
VERTEX SET IN A BALANCED
BIPARTITE GRAPH

Sam Pine^a

VOLUME 14, No. 1, SPRING 2013

Sponsored by

Rose-Hulman Institute of Technology

Department of Mathematics

Terre Haute, IN 47803

Email: mathjournal@rose-hulman.edu

<http://www.rose-hulman.edu/mathjournal>

^aElmira College

BOUNDS FOR ELEMENTS OF THE DEGREE
SEQUENCE OF AN UNKNOWN VERTEX SET IN A
BALANCED BIPARTITE GRAPH

Sam Pine

Abstract. Consider the set of all balanced bipartite graphs. Given the degree sequence of one vertex set in one of these graphs, we find bounds for any given position in the degree sequence of the unknown vertex set. Additionally, we establish bounds for the median of the unknown degree sequence, as well as bounds for any given percentile. We discuss the connection between this paper and the High School Prom Theorem.

Acknowledgements: Supported in part by the Elmira College Summer Research Program. Thanks to Dr. Charlie Jacobson for his guidance, and to an anonymous referee for his helpful suggestions.

1 Introduction

In 2001, a British study [JOH] reported that men have an average of 12.7 female sex partners in a lifetime, whereas women have 6.5 male sex partners per lifetime, on average. Then in 2007, the Department of Health and Human Services released a report [FRY] stating that the median number of female partners men have in a lifetime is 7, while women have a median of 4 male partners. David Gale, a prominent mathematician and economist, criticized these publications and questioned whether the data sets were logically possible [DAV]. Later in 2007, a *New York Times* article [KOL] discussed the two studies mentioned here; the article even included a proof of Gale's High School Prom theorem (HSPT), in which Gale demonstrates that the mean number of partners for men and women must necessarily be the same. The theorem is a direct result of the structure inherent to bipartite graphs, leading to the possibility that the structure of bipartite graphs may yield other statistical information. Given the degree sequence of one vertex set, we place bounds on positions in the degree sequence of the other vertex set. Furthermore, this paper gives bounds for a given percentile in the degree sequence of an unknown vertex set, as well as sharp bounds for the median of the unknown vertex set.

In the next section, we provide an overview of the background material one should study prior to reading this paper. In section 3, we develop a technique for placing bounds on a given position in the degree sequence of one vertex set, given the degree sequence of the other set. In section 4, we modify these bounds to account for the possibility that we know the maximum degree of any vertex in the unknown set. Section 4 also contains a method for bounding a given percentile in the degree sequence of the unknown vertex set. In section 5, we adapt the previous techniques to calculate bounds for the median of the degree sequence of the unknown vertex set. Finally, in section 6 we discuss some unsolved problems related to the researched contained here.

2 Background

Before continuing, the reader should familiarize himself with the basic terminology of graph theory; a brief overview of this information is included for convenience. A *graph* is a pictorial representation of a set of objects. The objects are symbolized by *vertices*, and lines drawn

between two vertices, called *edges*, represent some relationship between those objects. The *degree* of a vertex refers to the number of edges connected to that vertex. We sometimes classify a graph by its *order*, the number of vertices in the graph, and its *size*, the number of edges in the graph. *Bipartite graphs* are a special class of graph in which the vertices are divided into two groups, called *vertex sets*, and no edges exist between two vertices in the same set. Bipartite graphs in which both vertex sets have the same number of vertices are *balanced* bipartite graphs.

Every situation to which the HSPT applies involves a balanced bipartite graph. One vertex set corresponds to the male participants, and the other set corresponds to the female participants. An edge drawn between two vertices indicates a sexual interaction—or a dance at the prom. We call problems based on situations of this type *high school prom problems*. The degree sum of either set is necessarily equal to the total number of edges in the graph, m . Thus, we may think of the degree sequences of each set as a special type of partition of m . Since vertices of degree zero are possible, we will use *partition of m* to describe any means of writing a number m as the unordered sum of non-negative integers. This is non-standard, but allows us to maintain the common graph theory practice of referring to m as the number of edges in a graph.

The partitions we are concerned with will have exactly n parts, where n is the number of vertices in one vertex set. Generating all the partitions of m into n parts with each part less than or equal to n is equivalent to finding all of the possible *degree sequences* of one vertex set in a balanced bipartite graph of order $2n$ and size m . If it is possible to assign two degree sequences to the vertex sets of a balanced bipartite graph, they are *compatible*. Suppose that we generate the possible degree sequences of a graph by using this partition method. When we attempt to pair these sequences and assign them to the vertex sets of a bipartite graph, we find that some pairings are impossible. For example, consider the possible degree sequences $[5,4,3,3,2,1]$ and $[6,6,3,2,1,0]$. The second degree sequence has two vertices of degree six, which means that we would expect each of the vertices in the other set to be of at least degree 2. Thus these degree sequences are incompatible.

Every bipartite graph corresponds to a *biadjacency matrix*, a binary matrix in which a 1 indicates an edge between two vertices, and a 0 indicates that no edge exists. The row sum vector and column sum vector of a biadjacency matrix correspond to the degree sequences of the two

vertex sets in a bipartite graph. The row sum vector and column sum vector of a matrix M will now be denoted $r(M)$ and $c(M)$ respectively. Given one degree sequence of a bipartite graph, $A = [a_1, a_2, a_3, \dots, a_k]$, $a_1 \geq a_2 \geq a_3 \geq \dots \geq a_k$, we may construct a matrix with a_k ones followed by zeros in each row, k . This matrix corresponds to the Ferrers diagram for the partition $a_1 + a_2 + a_3 + \dots + a_k$; we borrow terminology from a work by Krause [KRA] and refer to this matrix as the *Ferrers matrix of A* . Consider the partition $[5,4,3,3,2,1]$ from above. The partition corresponds to the Ferrers matrix below; notice that the row sums form the original partition, and that the column sums form another partition.

$$\begin{array}{rcccccc}
 & 6 & 5 & 4 & 2 & 1 & 0 \\
 5 & \left[\begin{array}{cccccc} 1 & 1 & 1 & 1 & 1 & 0 \end{array} \right. \\
 4 & \left[\begin{array}{cccccc} 1 & 1 & 1 & 1 & 0 & 0 \end{array} \right. \\
 3 & \left[\begin{array}{cccccc} 1 & 1 & 1 & 0 & 0 & 0 \end{array} \right. \\
 3 & \left[\begin{array}{cccccc} 1 & 1 & 1 & 0 & 0 & 0 \end{array} \right. \\
 2 & \left[\begin{array}{cccccc} 1 & 1 & 0 & 0 & 0 & 0 \end{array} \right. \\
 1 & \left[\begin{array}{cccccc} 1 & 0 & 0 & 0 & 0 & 0 \end{array} \right.
 \end{array}$$

The partition formed by the column sums of A is the *conjugate partition*, or just the *conjugate* of the original partition. From the above diagram, we see that the conjugate of $[5,4,3,3,2,1]$ is $[6,5,4,2,1,0]$. We may construct the Ferrers matrix for any degree sequence, and so a degree sequence and its conjugate always correspond to a biadjacency matrix, and are therefore always compatible. The Gale-Ryser theorem, proved independently by both David Gale and Herbert Ryser [RYS], provides a clear condition for compatibility based on the conjugate of a given degree sequence. Krause presents the Gale-Ryser theorem in a straightforward manner [KRA] by first defining *domination* for two partitions p and q ; p dominates q if $\sum_{i=1}^m p_i \geq \sum_{i=1}^m q_i$ for all positive integers m . For example, the partition $[5,5,3,2,1]$ dominates $[4,4,4,3,1]$. Then, the theorem states that there exists a binary matrix A such that $c(A) = p$ and $r(A) = q$ iff q is dominated by the conjugate of p .

3 Bounds on a Given Position

With the domination condition and the Gale-Ryser theorem, we are now prepared to place bounds on values in the degree sequence of one vertex set, given the other vertex set.

Theorem**1**

Let U and V be the vertex sets of a balanced bipartite graph of order $2n$, with degree sequences A and C , respectively. If the conjugate of A is $B = [b_1, b_2, b_3, \dots, b_n]$, then c_p , the entry in position p of C , is bounded by:

$$\left\lceil \frac{\sum_{i=p}^n b_i}{n-(p-1)} \right\rceil \leq c_p \leq \left\lfloor \frac{\sum_{i=1}^p b_i}{p} \right\rfloor.$$

Proof

Suppose we are given the degree sequence of U with n vertices:

$$A = [a_1, a_2, a_3, \dots, a_n], \quad a_1 \geq a_2 \geq a_3 \geq \dots \geq a_n.$$

Let B be the conjugate of A : $B = [b_1, b_2, b_3, \dots, b_n], \quad b_1 \geq b_2 \geq b_3 \geq \dots \geq b_n.$

By extension of the Gale-Ryser Theorem, all compatible degree sequences of V must be dominated by B .

Thus for degree sequence: $C = [c_1, c_2, c_3, \dots, c_n], \quad c_1 \geq c_2 \geq c_3 \geq \dots \geq c_n$

we have:

$$\sum_{i=1}^m b_i \geq \sum_{i=1}^m c_i \text{ for } 1 \leq i \leq m \leq n.$$

In particular, for any p , $\sum_{i=1}^p b_i \geq \sum_{i=1}^p c_i \geq pc_p$. Through an application of the division algorithm, it follows that $\left\lfloor \frac{\sum_{i=1}^p b_i}{p} \right\rfloor \geq c_p$, where $\lfloor x \rfloor$ denotes the floor function.

Since $\sum_{i=1}^{p-1} b_i \geq \sum_{i=1}^{p-1} c_i$, we have $\sum_{i=p}^n b_i \leq \sum_{i=p}^n c_i \leq (n - (p - 1))c_p$. It follows that

$$\left\lceil \frac{\sum_{i=p}^n b_i}{n-(p-1)} \right\rceil \leq c_p, \text{ where } \lceil x \rceil \text{ denotes the ceiling function. } \blacksquare$$

Example

Suppose we have vertex sets U and V in a balanced bipartite graph. Let the degree sequence of U be $A = [5, 3, 3, 2, 1, 0]$, and let the degree sequence of V be $C = [c_1, c_2, c_3, c_4, c_5, c_6]$. What is the greatest value c_4 can achieve? The least value?

The conjugate of A is $B = [5, 4, 3, 1, 1, 0]$. Then $\sum_{i=1}^4 b_i = 13$. To make c_4 as large as possible, we distribute the sum over the first four positions in C as evenly as possible, without spilling over into the fifth position: $[4, 3, 3, 3, 1, 0]$.

Thus $c_4 \leq \left\lfloor \frac{13}{4} \right\rfloor = 3$.

For the lower bound, we find $\sum_{i=4}^6 b_i = 2$. To make c_4 as small as possible, we distribute the sum over the last three positions in C as evenly as possible: $[5,4,3,1,1,0]$.

Thus $c_4 \geq \left\lfloor \frac{2}{3} \right\rfloor = 1$.

For low-order bipartite graphs, this method yields reasonable bounds. For larger graphs, however, utilizing this method will generate less useful bounds with respect to problems of the high school prom type. For example, suppose that we polled 10,000 women and included a question about their total number of male sex partners. The results are summarized in the table below.

NUMBER OF MALE PARTNERS	FREQUENCY
4	1000
2	2000
1	5000
0	2000

The degree sequence of the vertex set that models this situation would be $[4, \dots, 4, 2, \dots, 2, 1, \dots, 1, 0, \dots, 0]$, and its conjugate would be $[8000, 3000, 1000, 1000, 0, \dots, 0]$, a poor reflection of the sexual history of men. While it is technically possible for a man to have 8000 female sex partners, it is still highly unlikely. Instead of basing our bounds on this conjugate, it makes more sense to limit the maximum degree of any vertex in the unknown set. Suppose that we limited the maximum degree in our compatible degree sequences to 16, something less fantastic. Since the sum of the elements in the conjugate is 13,000, an application of the method of the previous theorem implies that the minimum values of the first $\left\lfloor \frac{13000}{16} \right\rfloor = 812$ values in the unknown degree sequence will no longer be 0.

4 Revised Bounds Given a Maximum Degree

The following theorem is an adaptation of the first theorem that takes the maximum degree of the unknown degree sequence into account.

Theorem

2

Let U and V be the vertex sets of a balanced bipartite graph of order $2n$, with degree sequences A

and C , respectively. Let the maximum degree of V be q . If the conjugate of A is $B = [b_1, b_2, b_3, \dots, b_n]$, then c_p , the entry in position p of C is bounded by:

$$\left\lfloor \frac{\sum_{i=1}^n b_i - \min(\sum_{i=1}^{p-1} b_i, q(p-1))}{n-(p-1)} \right\rfloor \leq c_p \leq \min \left(\left\lfloor \frac{\sum_{i=1}^p b_i}{p} \right\rfloor, q \right).$$

Proof

Suppose we are given the degree sequence of U with n vertices:

$$A = [a_1, a_2, a_3, \dots, a_n], \quad a_1 \geq a_2 \geq a_3 \geq \dots \geq a_n.$$

Let B be the conjugate of A and C be a compatible degree sequence, as before.

Since B dominates C , and $c_i \geq c_{i+1}$ for all i , $\sum_{i=1}^p b_i \geq \sum_{i=1}^p c_i \geq pc_p$. As above, we have

$$\left\lfloor \frac{\sum_{i=1}^p b_i}{p} \right\rfloor \geq c_p. \text{ Also, } c_i \leq q \text{ for all } 1 \leq i \leq n. \text{ Thus } c_p \leq \min \left(\left\lfloor \frac{\sum_{i=1}^p b_i}{p} \right\rfloor, q \right).$$

Suppose the first $p-1$ elements of the compatible sequence are as large as possible, that is, suppose $\sum_{i=1}^{p-1} c_i = \min(\sum_{i=1}^{p-1} b_i, q(p-1))$. This minimizes the sum of the remaining elements in the compatible sequence, given by $\sum_{i=p}^n c_i = \sum_{i=1}^n b_i - \min(\sum_{i=1}^{p-1} b_i, q(p-1))$.

Note that $(n - (p - 1))c_p \geq \sum_{i=p}^n c_i$, so it follows that $c_p \geq \left\lfloor \frac{\sum_{i=1}^n b_i - \min(\sum_{i=1}^{p-1} b_i, q(p-1))}{n-(p-1)} \right\rfloor$. ■

Example

Consider the hypothetical sample of women from above. We have the degree sequence

$$A = [4, \dots, 4, 2, \dots, 2, 1, \dots, 1, 0, \dots, 0] \quad \text{with conjugate} \quad B = [8000, 3000, 1000, 1000, 0, \dots, 0].$$

Suppose the maximum degree of the unknown degree sequence $C = [c_1, c_2, c_3, \dots, c_{10000}]$ is 16.

What is the greatest possible value of c_{10} ? The least possible value?

$$\sum_{i=1}^{10} b_i = 13,000, \text{ and so } c_{10} \leq \left\lfloor \frac{13000}{10} \right\rfloor = 1300. \text{ The maximum degree of } C \text{ is } 16, \text{ so } c_{10} \leq 16.$$

Since the maximum degree of C is 16, the compatible degree sequence containing the lowest possible value of c_{10} will be of the form: $[16, \dots, 16, c_{10}, \dots, c_{10000}]$. Thus there remains

$$\sum_{i=1}^{10000} b_i - 16(9) = 12,856 \text{ to distribute over the remaining vertices, } c_{10}, \dots, c_{10000}. \text{ } c_{10} \text{ is the greatest of these values, and so } c_{10} \geq \left\lfloor \frac{12856}{9991} \right\rfloor = 2.$$

Previously, the minimum value of c_{10} was 0, and so we see a slight constriction of the bounds by restricting the maximum value of the compatible degree sequences, q . A lower value of q will produce bounds equivalent to or tighter than a greater value of q .

For convenience, the following corollary demonstrates a technique for generating bounds for a given percentile within a compatible degree sequence, given the degree sequence of one vertex set in a balanced bipartite graph. Note that the bounds are not sharp, as in the theorems above, but that percentile values calculated using the method approved by the National Institute of Standards and Technology [NIST] will fall within these bounds.

Corollary

Let U and V be the vertex sets of a balanced bipartite graph of order $2n$, with degree sequences A and C , respectively. Let the maximum degree of V be q . Let $d = \max\left(\left\lfloor \frac{n(100-k)}{100} \right\rfloor, 1\right)$. If the conjugate of A is $B = [b_1, b_2, b_3, \dots, b_n]$, then the k^{th} percentile of C , P , is bounded by:

$$\left\lfloor \frac{\sum_{i=1}^n b_i - \min(\sum_{i=1}^d b_i, q(d))}{n-d} \right\rfloor \leq P \leq \min\left(\left\lfloor \frac{\sum_{i=1}^d b_i}{d} \right\rfloor, q\right).$$

Notice that these are the same bounds as in theorem 2, except that the maximum is for position d , while the minimum is calculated for position $d + 1$. For any given percentile, the value $\frac{n(100-k)}{100}$ gives the position of this value in the compatible degree sequences. Only integer values correspond to positions in the compatible degree sequences, so we calculate the max of the position d , and the minimum of the position $d + 1$. We cannot select a position before the first, so $d \geq 1$ for all k .

5 Bounds for the Median

One can use the percentile procedure above to bound the medians of compatible degree sequences, but the resulting bound is not necessarily sharp. The median is important enough to warrant its own procedure for producing sharp bounds, with a process slightly more complicated than shown in the previous arguments. Theorem 3 demonstrates a procedure for placing sharp bounds on the median of a compatible degree sequence.

Theorem 3

Let U and V be the vertex sets of a balanced bipartite graph of order $2n$, with degree sequences A and C , respectively. Let the maximum degree of V be q . Let the median degree of V be denoted M . If the conjugate of A is $B = [b_1, b_2, b_3, \dots, b_n]$, then M is bounded:

for odd n :

$$\left\lfloor \frac{2\left(\sum_{i=1}^n b_i - \min\left(\sum_{i=1}^{(n-1)/2} b_i, \frac{q(n-1)}{2}\right)\right)}{n+1} \right\rfloor \leq M \leq \min\left(\left\lfloor \frac{2\sum_{i=1}^{(n+1)/2} b_i}{n+1} \right\rfloor, q\right)$$

for even n :

$$M \leq \min\left(\left\lfloor \frac{2\sum_{i=1}^{n/2+1} b_i}{n+2} \right\rfloor - \frac{1}{2}, q\right) \quad \text{or} \quad M \leq \min\left(\left\lfloor \frac{2\sum_{i=1}^{n/2+1} b_i}{n+2} \right\rfloor, q\right)$$

$$M \geq \left\lfloor \frac{2\left(\sum_{i=1}^n b_i - \min\left(\sum_{i=1}^{n/2-1} b_i, q\left(\frac{n}{2}-1\right)\right)\right)}{n+2} \right\rfloor - \frac{1}{2} \quad \text{or}$$

$$M \geq \left\lfloor \frac{2\left(\sum_{i=1}^n b_i - \min\left(\sum_{i=1}^{n/2-1} b_i, q\left(\frac{n}{2}-1\right)\right)\right)}{n+2} \right\rfloor.$$

Proof

For odd n , the median is an element of the compatible degree sequence. In every case, the median falls on position $\frac{n+1}{2}$. Apply theorem 2 with $p = \frac{n+1}{2}$ and we are done.

For even n , the median is not an element of the compatible degree sequence. For each compatible degree sequence, the median is the average of the $\frac{n}{2}$ element and the $\frac{n}{2} + 1$ element.

We first compute an upper bound. To compute this bound, we consider a specific form of compatible degree sequences that have the greatest possible median for any compatible degree sequence. Because of the domination condition, the maximum size of the $\frac{n}{2}$ element and the $\frac{n}{2} + 1$ element depend on the sum of the first $\frac{n}{2} + 1$ elements in B ; we see that the two elements in question are largest when this sum is distributed over the first $\frac{n}{2} + 1$ elements of C as evenly as possible. Let C be of this form. Since B dominates C , $\sum_{i=1}^{n/2+1} b_i \geq \sum_{i=1}^{n/2+1} c_i$. Apply the division algorithm to dividend $\sum_{i=1}^{n/2+1} b_i$, divisor $\frac{n}{2} + 1$, and non-negative quotient, Q , and remainder, R .

For the first R elements in C ,

$$c_i = \left\lfloor \frac{\sum_{i=1}^{\frac{n}{2}+1} b_i}{\frac{n}{2}+1} \right\rfloor = \left\lfloor \frac{2 \sum_{i=1}^{\frac{n}{2}+1} b_i}{n+2} \right\rfloor.$$

For the remaining elements in C ,

$$c_i \leq \left\lfloor \frac{\sum_{i=1}^{\frac{n}{2}+1} b_i}{\frac{n}{2}+1} \right\rfloor = \left\lfloor \frac{2 \sum_{i=1}^{\frac{n}{2}+1} b_i}{n+2} \right\rfloor.$$

Also, $c_i \leq q$ for all $1 \leq i \leq n$.

For $R = \frac{n}{2}$, we have

$$M \leq \min \left(\left\lfloor \frac{2 \sum_{i=1}^{\frac{n}{2}+1} b_i}{n+2} \right\rfloor - \frac{1}{2}, q \right).$$

For $R < \frac{n}{2}$, we have

$$M \leq \min \left(\left\lfloor \frac{2 \sum_{i=1}^{\frac{n}{2}+1} b_i}{n+2} \right\rfloor, q \right).$$

We now compute a lower bound. To compute this bound, we consider a specific form of compatible degree sequences that have the least possible median of any compatible degree sequence. To minimize the $\frac{n}{2}$ and $\frac{n}{2} + 1$ elements, the first $\frac{n}{2} - 1$ elements in this degree sequence should be as large as possible. Since the $\frac{n}{2}$ and $\frac{n}{2} + 1$ elements are largest of the remaining elements, and $\sum_{i=1}^n b_i = \sum_{i=1}^n c_i$, the two elements are minimized by distributing $\sum_{i=n/2}^n b_i$ as evenly as possible over the last $\frac{n}{2} + 1$ terms in the compatible degree sequence. Let C be of this form.

Suppose $\sum_{i=1}^{n/2-1} c_i = \min \left(\sum_{i=1}^{n/2-1} b_i, q \left(\frac{n}{2} - 1 \right) \right)$.

Then $\sum_{i=n/2}^n c_i = \sum_{i=1}^n b_i - \min \left(\sum_{i=1}^{n/2-1} b_i, q \left(\frac{n}{2} - 1 \right) \right)$.

Now apply the division algorithm to dividend $\sum_{i=1}^n b_i - \min \left(\sum_{i=1}^{n/2-1} b_i, q \left(\frac{n}{2} - 1 \right) \right)$, divisor $\frac{n}{2} + 1$, and non-negative Q and R .

For the elements $c_{n/2}$ to $c_{n/2+(R-1)}$: $c_i \geq \left\lfloor \frac{2 \left(\sum_{i=1}^n b_i - \min \left(\sum_{i=1}^{n/2-1} b_i, q \left(\frac{n}{2} - 1 \right) \right) \right)}{n+2} \right\rfloor$.

For the elements $c_{\frac{n}{2}+R}$ to c_n : $c_i \leq \left\lfloor \frac{2 \left(\sum_{i=1}^n b_i - \min \left(\sum_{i=1}^{n/2-1} b_i, q \left(\frac{n}{2} - 1 \right) \right) \right)}{n+2} \right\rfloor$.

Then for $R = 1$, we have

$$M \geq \left\lfloor \frac{2\left(\sum_{i=1}^n b_i - \min\left(\sum_{i=1}^{n/2-1} b_i, q\left(\frac{n}{2}-1\right)\right)\right)}{n+2} \right\rfloor - \frac{1}{2}.$$

And for $R > 1$, we have

$$M \geq \left\lfloor \frac{2\left(\sum_{i=1}^n b_i - \min\left(\sum_{i=1}^{n/2-1} b_i, q\left(\frac{n}{2}-1\right)\right)\right)}{n+2} \right\rfloor. \blacksquare$$

6 Utilization and Further Research

The tools developed so far could be used to analyze data sets like those that led to the HSPT. For example, given the full data set reported by women in such a study, we could compute bounds for the median of the men's data set. If the reported median for men falls within this range, then the study's results are plausible. If not, we know that the reported data is unreliable.

Dealing with the large, sparsely populated graphs that correspond to studies done by the likes of the Department of Health and Human Services makes analysis more complicated. Through more analysis of these large bipartite graphs and the biadjacency matrices associated with them, it may be possible to predict the median of a bipartite data set, given the degree sequence of the other set. For small graphs, it is possible to create a frequency distribution of the possible medians of one vertex set, given the degree sequence of the other vertex set. For example, suppose we are given the degree sequence, $[3,2,1,0]$, of a balanced bipartite graph. Then the other degree sequence must be one of four possibilities: $[2,2,1,0]$, $[2,2,2,0]$, $[3,1,1,1]$ or $[3,2,1,0]$. In order create a frequency distribution, we require a method of enumerating bipartite graphs with given degree sequences. Rather than considering the graphs, we study the corresponding biadjacency matrices. If we attempt to construct a biadjacency matrix with row sums $[3,2,1,0]$ and column sums that correspond to one of the four compatible degree sequences, we find that there are many matrices which satisfy each set of sums. Among binary matrices with row sums $[3,2,1,0]$, there are 48 with column sums $[2,2,1,1]$, 12 with column sums $[2,2,2,0]$, 12 with column sums $[3,1,1,1]$, and 24 with column sums $[3,2,1,0]$. From this, we find that 96 balanced bipartite graphs have a vertex set with degree sequence $[3,2,1,0]$. Categorizing these

graphs by the median degree of the other vertex set, we find that 12 graphs have a median of 2, 72 have a median of 1.5, and 12 have a median of 1. Thus, given the degree sequence $[3,2,1,0]$ of one vertex set in a balanced bipartite graph, we have the following frequency distribution for the median degree of the other vertex set:

MEDIAN	FREQUENCY	PERCENT
1	12	12.5
1.5	72	75
2	12	12.5

The difficulty in this approach lies in computing the number of matrices with given row and column sums. A brute force method can be applied to smaller matrices; permuting the ones and zeroes in each of the rows will create all of the desired matrices, but some matrices created by this process do not satisfy the row and column sums. Identifying these matrices leads to a branching problem. While a solution to this problem has not been presented yet, both Brualdi and Ryser published work that is relevant to the issue. Their combined work details the terminology and ideas used in the field of combinatorial matrix theory [BR1], but contains little in the way of enumerating matrices with specific properties. Brualdi introduces the idea of a structure matrix in his work [BR2]. This matrix identifies all of the invariable positions among binary matrices with given row and column sums; finding the invariable positions in a given type of matrix seems a promising start to enumeration of those matrices. If the reader wishes to generate additional examples for study, tools for generating the bounds discussed in this paper are available on Maplesoft [PIN].

References

- [AS] Armen S. Asratian, Tristan M.J. Denley, Roland Häggkvist, *Bipartite Graphs and Their Applications*, Cambridge University Press, New York, 1998.
- [AND] G.E. Andrews and K. Eriksson, *Integer Partitions*, Cambridge University Press, 2004.
- [BR1] Richard A. Brualdi, Matrices of Zeros and Ones with Fixed Row and Column Sum Vectors, *Linear Algebra and its Applications*, 33(Oct. 1980) 159-231.

- [BR2] Richard A. Brualdi and Herbert J. Ryser, *Combinatorial Matrix Theory*, Cambridge University Press, New York, 1991. P 1-11, 198-209.
- [CHA] Gary Chartrand and Ping Zhang, *A First Course in Graph Theory*, Dover Publications, New York, 2012.
- [DAV] Martin Davis, Sex and the Mathematician: The High School Prom Theorem, *Games and Economic Behavior*, 66.2(July 2009). 600.
- [FRY] Cheryl Fryar *et al.*, Drug Use and Sexual Behavior Reported by Adults: United States, 1999-2002, *Advance Data from Vital and Health Statistics*; no. 384. Hyattsville, MD: National Center for Health Statistics. 2007.
- [KOL] Gina Kolatta, The Myth, the Math, the Sex, *New York Times*, 12 Aug. 2007.
- [KRA] Manfred Krause, A Simple Proof of the Gale- Ryser Theorem, *The American Mathematical Monthly*, 103.4(Apr. 1996). 335-337.
- [JOH] Anne M. Johnson *et al.*, Sexual Behaviour in Britain: Partnerships, Practices, and HIV Risk Behaviours, *The Lancet*, 358(2001) 1835-1842.
- [NIST] *NIST/SEMATECH e-Handbook of Statistical Methods*, www.itl.nist.gov/div898/handbook, Accessed 3 Jun 2012.
- [PIN] Sam Pine, Bipartite Degree Sequences, submitted to Maplesoft on July 20, 2012
- [RYS] Herbert J. Ryser, Combinatorial Properties of Matrices of Zeros and Ones, *Canadian Journal of Mathematics*, 9(1957) 371-377.