

Clustering by Connections in the Input Matrix

Jonathan McFadden
University of Washington Tacoma
mcfaddja@uw.edu

Summer 2016

Contents

I	Introduction, Definitions, Datasets, and Metrics	5
1	Introduction	7
2	Definitions	9
2.1	Input Matrix	9
2.2	SVD Matrices	9
2.2.1	Full SVD	10
2.2.2	Approximate SVD	10
2.3	Useful Functions	11
2.4	Connection Matrices	11
2.4.1	Row Connection Matrix	11
2.4.2	Column Connection Matrix	12
2.5	Relation Sets	12
2.5.1	Row Relation Set	13
2.5.2	Column Relation Set	13
2.6	Cluster Sets	14
2.6.1	Row Cluster Set	15
2.6.2	Column Cluster Set	15
3	Datasets	17
3.1	Contrived Data	17
3.2	Random Data	17
3.3	"Real World" Data	18

4	Metrics	19
4.1	Entropy	19
4.2	Connection Quality	19
II	Algorithms	21
5	SVD Signs	23
5.1	Finding k	23
5.2	Compare Row Sign Patters	23
5.3	Complete Row Cluster Sets	23
5.4	Single Sets for each Row Cluster	23
5.5	Compare Column Sign Patters	23
5.6	Complete Column Cluster Sets	24
5.7	Single Sets for each Column Cluster	24
6	SVD Gaps	25
7	CbC Disjoint Sets	27
7.1	Compute $\bar{\bar{\mathbf{A}}}_U$	28
7.2	Construct \mathbb{U}	28
7.3	Complete the \mathbb{U} Sets	29
7.4	Construct the \mathbb{U}^* Sets	30
7.5	Compute $\bar{\bar{\mathbf{A}}}_V$	30
7.6	Construct \mathbb{V}	31
7.7	Complete the \mathbb{V} Sets	32
7.8	Construct the \mathbb{V}^* Sets	33
8	CbC Non-Disjoint Sets	35

Part I

Introduction, Definitions, Datasets, and Metrics

Chapter 1

Introduction

Introduction goes here

Chapter 2

Definitions

2.1 Input Matrix

Let $\bar{\bar{\mathbf{L}}} \in \mathbb{R}^{m \times n}$ be a matrix which represents how m terms describe a collection of n objects.

We can express this formally as

$$\left(\bar{\bar{\mathbf{L}}}\right)_{ij} \equiv \begin{cases} 1 & \text{if the } i\text{th term describes the } j\text{th object.} \\ 0 & \text{otherwise} \end{cases} \quad (2.1.1)$$

We may also use the alternative definition

$$\left(\bar{\bar{\mathbf{L}}}\right)_{ij} \equiv \begin{cases} x & \text{the } i\text{th term describes the } j\text{th object } x \text{ times} \\ 0 & \text{otherwise} \end{cases} \quad (2.1.2)$$

In both definitions for the $\left(\bar{\bar{\mathbf{L}}}\right)_{ij}$, we require that $i \in [1, m] \subset \mathbb{Z}^+$ and that $j \in [1, n] \subset \mathbb{Z}^+$.

2.2 SVD Matrices

We will be using both the full SVD of $\bar{\bar{\mathbf{L}}}$ and the κ th order approximate SVD of $\bar{\bar{\mathbf{L}}}$ where the order of the approximate SVD, κ , must satisfy the relation

$$\kappa < r$$

with r defined as

$$r = \text{rank} \left[\bar{\bar{\mathbf{L}}} \right] \quad (2.2.1)$$

2.2.1 Full SVD

Let the full singular value decomposition of $\bar{\bar{\mathbf{L}}}$ be given by

$$\bar{\bar{\mathbf{L}}} = \bar{\bar{\mathbf{U}}} \bar{\bar{\mathbf{S}}} \bar{\bar{\mathbf{V}}}^\top$$

where $\bar{\bar{\mathbf{U}}} \in \mathbb{R}^{m \times r}$, $\bar{\bar{\mathbf{S}}} \in \mathbb{R}^{r \times r}$, and $\bar{\bar{\mathbf{V}}} \in \mathbb{R}^{n \times r}$ with r as defined as above in 2.2.1. We call $\bar{\bar{\mathbf{U}}}$ the "*Row Matrix*" of $\bar{\bar{\mathbf{L}}}$, $\bar{\bar{\mathbf{S}}}$ the "*Singular Value Matrix*" of $\bar{\bar{\mathbf{L}}}$, and $\bar{\bar{\mathbf{V}}}$ the "*Column Matrix*" of $\bar{\bar{\mathbf{L}}}$.

2.2.2 Approximate SVD

Let the κ th order approximate singular value decomposition of $\bar{\bar{\mathbf{L}}}$ be expressed as

$$\bar{\bar{\mathbf{L}}} = \bar{\bar{\mathbf{U}}}_\kappa \bar{\bar{\mathbf{S}}}_\kappa \bar{\bar{\mathbf{V}}}_\kappa^\top$$

where $\bar{\bar{\mathbf{U}}}_\kappa \in \mathbb{R}^{m \times \kappa}$, $\bar{\bar{\mathbf{S}}}_\kappa \in \mathbb{R}^{\kappa \times \kappa}$, and $\bar{\bar{\mathbf{V}}}_\kappa \in \mathbb{R}^{n \times \kappa}$ with r as defined as above in 2.2.1. We call $\bar{\bar{\mathbf{U}}}_\kappa$ the "*Approximate Row Matrix*" or "*Reduced Row Matrix*" of $\bar{\bar{\mathbf{L}}}$, $\bar{\bar{\mathbf{S}}}_\kappa$ the "*Approximate Singular Value Matrix*" or "*Reduced Singular Value Matrix*" of $\bar{\bar{\mathbf{L}}}$, and $\bar{\bar{\mathbf{V}}}_\kappa$ the "*Approximate Column Matrix*" or "*Reduced Column Matrix*" of $\bar{\bar{\mathbf{L}}}$.

2.3 Useful Functions

The following function will prove useful for the alternative definitions of the *Connection Matrices* given below. This function indicates if the i th element of $\bar{\bar{\mathbf{L}}}$ can be connected to the j th element of $\bar{\bar{\mathbf{L}}}$ and is denoted by $\delta^*(\bar{\bar{\mathbf{L}}}, i, j)$. We define this function according to the expression

$$\delta^*(\bar{\bar{\mathbf{L}}}, i, j) \equiv \begin{cases} 1 & \text{if } \bar{\bar{\mathbf{L}}}_{ij} \neq 0 \\ 0 & \text{otherwise} \end{cases} \quad (2.3.1)$$

with $i \in [1, m] \subset \mathbb{Z}^+$ and that $j \in [1, n] \subset \mathbb{Z}^+$.

2.4 Connection Matrices

We will create two "*Connection Matrices*" from $\bar{\bar{\mathbf{L}}}$. These matrices represent the connections among all of the row elements of $\bar{\bar{\mathbf{L}}}$ or among all of the column elements of $\bar{\bar{\mathbf{L}}}$. They will provide the information required for clustering either the rows of $\bar{\bar{\mathbf{L}}}$ or for clustering the columns of $\bar{\bar{\mathbf{L}}}$.

2.4.1 Row Connection Matrix

The first of the "*Connection Matrices*" that we will create from $\bar{\bar{\mathbf{L}}}$ is the "*Row Connection Matrix*". This matrix will be used to provide information for clustering the rows of $\bar{\bar{\mathbf{L}}}$. We will denote this matrix by $\bar{\bar{\mathbf{A}}}_U \in \mathbb{R}^{m \times m}$ and define it, in terms of $\bar{\bar{\mathbf{L}}}$, according to the expression

$$\bar{\bar{\mathbf{A}}}_U = \bar{\bar{\mathbf{L}}} \bar{\bar{\mathbf{L}}}^\top \quad (2.4.1)$$

We may also use the alternate definition

$$\left(\bar{\bar{\mathbf{A}}}_U\right)_{ij} = \sum_{k=1}^n \left\{ \delta^* \left(\bar{\bar{\mathbf{L}}}, i, k \right) \bar{\bar{\mathbf{L}}}_{jk} \right\} \quad (1.4.1a)$$

where $i, j \in [1, m] \subset \mathbb{Z}^+$.

2.4.2 Column Connection Matrix

The second "Connection Matrix" to be created from $\bar{\bar{\mathbf{L}}}$ is the "Column Connection Matrix". This matrix will be used to provide information for clustering the columns of $\bar{\bar{\mathbf{L}}}$. We will denote this matrix by $\bar{\bar{\mathbf{A}}}_V \in \mathbb{R}^{n \times n}$ and define it, in terms of $\bar{\bar{\mathbf{L}}}$, according to the expression

$$\bar{\bar{\mathbf{A}}}_V = \bar{\bar{\mathbf{L}}}^\top \bar{\bar{\mathbf{L}}} \quad (2.4.2)$$

We may also use the alternate definition

$$\left(\bar{\bar{\mathbf{A}}}_V\right)_{ij} = \sum_{k=1}^m \left\{ \delta^* \left(\bar{\bar{\mathbf{L}}}, k, i \right) \bar{\bar{\mathbf{L}}}_{kj} \right\} \quad (1.4.2a)$$

where $i, j \in [1, n] \subset \mathbb{Z}^+$.

2.5 Relation Sets

We will now define two sets of 2-tuples. The first will describe the relation or relations each row element has with all the other row elements; while the second will describe the relation or relations each column element has with all the other column elements. One of the elements of the 2-tuple represents the other rows or columns that are related to a given row or column. The other element of the 2-tuple describes the strength of those relations.

2.5.1 Row Relation Set

Let \mathbb{U} be a set of m 2-tuples that can be expressed as

$$\mathbb{U} = \left\{ \{\mathbb{U}_1, \mathbb{M}_1\}, \{\mathbb{U}_2, \mathbb{M}_2\}, \dots, \{\mathbb{U}_m, \mathbb{M}_m\} \right\}$$

where the $\{\mathbb{U}_i, \mathbb{M}_i\}$ are 2-tuples where the first element of the tuple, \mathbb{U}_i , represents which rows in $\bar{\bar{\mathbf{L}}}$ are connected to the i th row of $\bar{\bar{\mathbf{L}}}$; while the second element of the tuple, \mathbb{M}_i , describes the strength of those relations, with i such that $i \in [1, m] \subset \mathbb{Z}^+$. We now define \mathbb{U}_i and \mathbb{M}_i concurrently as

$$\mathbb{U}_i \equiv \left\{ j \mid j \in [1, m] \subset \mathbb{Z}^+ \text{ and } \left(\bar{\bar{\mathbf{A}}}_U \right)_{ij} \neq 0 \right\} \quad (2.5.1)$$

and

$$\mathbb{M}_i \equiv \left\{ \left(\bar{\bar{\mathbf{A}}}_U \right)_{ij} \mid j \in [1, m] \subset \mathbb{Z}^+ \text{ and } \left(\bar{\bar{\mathbf{A}}}_U \right)_{ij} \neq 0 \right\} \quad (2.5.2)$$

respectively. These definitions require that the expression

$$|\mathbb{U}_i| = |\mathbb{M}_i|$$

holds for all $i \in [1, m] \subset \mathbb{Z}^+$. Additionally, we have defined \mathbb{U}_i and \mathbb{M}_i such that the strength of the connection to the l th element in \mathbb{U}_i is represented by the l th element of \mathbb{M}_i .

2.5.2 Column Relation Set

Let \mathbb{V} be a set of n 2-tuples that can be expressed as

$$\mathbb{V} = \left\{ \{\mathbb{V}_1, \mathbb{N}_1\}, \{\mathbb{V}_2, \mathbb{N}_2\}, \dots, \{\mathbb{V}_n, \mathbb{N}_n\} \right\}$$

where the $\{\mathbb{V}_j, \mathbb{N}_j\}$ are 2-tuples where the first element of the tuple, \mathbb{V}_j , represents which columns in $\bar{\bar{\mathbf{L}}}$ are connected to the j th column of $\bar{\bar{\mathbf{L}}}$; while the second element of the tuple, \mathbb{N}_j , describes the strength of those relations, with j such that $j \in [1, n] \subset \mathbb{Z}^+$. We now define \mathbb{V}_j and \mathbb{N}_j concurrently as

$$\mathbb{V}_j \equiv \left\{ i \mid i \in [1, n] \subset \mathbb{Z}^+ \text{ and } \left(\bar{\bar{\mathbf{A}}}_V \right)_{ji} \neq 0 \right\} \quad (2.5.3)$$

and

$$\mathbb{N}_j \equiv \left\{ \left(\bar{\bar{\mathbf{A}}}_V \right)_{ji} \mid i \in [1, n] \subset \mathbb{Z}^+ \text{ and } \left(\bar{\bar{\mathbf{A}}}_V \right)_{ji} \neq 0 \right\} \quad (2.5.4)$$

respectively. These definitions require that the expression

$$|\mathbb{V}_j| = |\mathbb{N}_j|$$

holds for all $j \in [1, n] \subset \mathbb{Z}^+$. Additionally, we have defined \mathbb{V}_j and \mathbb{N}_j such that the strength of the connection to the l th element in \mathbb{V}_j is represented by the l th element of \mathbb{N}_j .

2.6 Cluster Sets

We now define two sets of sets with the first set of sets representing the row clusters and the second set of sets representing the column clusters. Each set of sets is composed of sets which represent the members of each row or column cluster.

2.6.1 Row Cluster Set

The "*Row Cluster Set*" is a set of $\mathbf{u} \in \mathbb{Z}^+$ sets with each component set representing one of the \mathbf{u} row clusters. We will denote this set of sets as \mathbb{U}^* and formally describe it as

$$\mathbb{U}^* = \left\{ \mathbb{U}_1^*, \mathbb{U}_2^*, \dots, \mathbb{U}_{\mathbf{u}}^* \right\}$$

where the \mathbb{U}_μ^* are sets which contain the members of the μ th row cluster, with $\mu \in [1, \mathbf{u}] \subset \mathbb{Z}^+$. Each of the \mathbb{U}_μ^* is constructed from the $\{\mathbb{U}_i, \mathbb{M}_i\}$ of its (the \mathbb{U}_μ^*) constituent elements via a process we will describe later.

2.6.2 Column Cluster Set

The "*Column Cluster Set*" is a set of $\mathbf{v} \in \mathbb{Z}^+$ sets with each component set representing one of the \mathbf{v} column clusters. We will denote this set of sets as \mathbb{V}^* and formally describe it as

$$\mathbb{V}^* = \left\{ \mathbb{V}_1^*, \mathbb{V}_2^*, \dots, \mathbb{V}_{\mathbf{v}}^* \right\}$$

where the \mathbb{V}_ν^* are sets which contain the members of the ν th column cluster, with $\nu \in [1, \mathbf{v}] \subset \mathbb{Z}^+$. Each of the \mathbb{V}_ν^* is constructed from the $\{\mathbb{V}_j, \mathbb{N}_j\}$ of its (the \mathbb{V}_ν^*) constituent elements via a process we will describe later.

Chapter 3

Datasets

Initial Dataset info goes here

There 3 types of datasets

1. Contrived Datasets
2. Random Datasets
3. "*Real World*" Datasets

The "*Real World*" will initially be truncated due to limits on processing and memory; however, later, a few of the "*Real World*" datasets will be run in full.

3.1 Contrived Data

Contrived Dataset info goes here

3.2 Random Data

Random Dataset info goes here

3.3 ”*Real World*” Data

”*Real World*” Dataset info goes here

Chapter 4

Metrics

Initial Metrics info goes here

4.1 Entropy

Initial Entropy Metric info goes here

4.2 Connection Quality

Initial Connection Quality Metric info goes here

Part II

Algorithms

Chapter 5

SVD Signs

Intro SVD Signs Algorithm info goes here

5.1 Finding k

Finding k section SVD Signs Algorithm info goes here

5.2 Compare Row Sign Patterns

Compare Row Sign Patterns section SVD Signs Algorithm info goes here

5.3 Complete Row Cluster Sets

Complete Row Cluster Sets section SVD Signs Algorithm info goes here

5.4 Single Sets for each Row Cluster

Single Sets for each Row Cluster section SVD Signs Algorithm info goes here

5.5 Compare Column Sign Patterns

Compare Column Sign Patterns section SVD Signs Algorithm info goes here

5.6 Complete Column Cluster Sets

Complete Column Cluster Sets section SVD Signs Algorithm info goes here

5.7 Single Sets for each Column Cluster

Single Sets for each Column Cluster section SVD Signs Algorithm info goes here

Chapter 6

SVD Gaps

Intro SVD Gaps Algorithm info goes here

Chapter 7

CbC Disjoint Sets

When $\bar{\bar{\mathbf{L}}}$ consists of a series of disjoint sets, our clustering algorithm proceeds according to the following simple steps

1. Compute $\bar{\bar{\mathbf{A}}}_U$
2. Construct $\mathbb{U} = \left\{ \{U_1, \mathbb{M}_1\}, \{U_2, \mathbb{M}_2\}, \dots, \{U_m, \mathbb{M}_m\} \right\}$
3. Complete the $\{U_i, \mathbb{M}_i\} \in \mathbb{U}$
4. Construct the \mathbb{U}^* set of sets
5. Compute $\bar{\bar{\mathbf{A}}}_V$
6. Construct $\mathbb{V} = \left\{ \{V_1, \mathbb{N}_1\}, \{V_2, \mathbb{N}_2\}, \dots, \{V_n, \mathbb{N}_n\} \right\}$
7. Complete the $\{V_j, \mathbb{N}_j\} \in \mathbb{V}$
8. Construct the \mathbb{V}^* set of sets
9. Reorder $\bar{\bar{\mathbf{L}}}$ in the a new, temporary matrix, $\bar{\bar{\mathbf{L}}}_0 \in \mathbb{R}^{m \times n}$.
10. Reorder the temporary matrix, $\bar{\bar{\mathbf{L}}}_0$ into the final matrix $\bar{\bar{\mathbf{L}}}^* \in \mathbb{R}^{m \times n}$.

The description of our algorithm for the case when $\bar{\bar{\mathbf{L}}}$ is composed of non-disjoint sets is given later in chapter 3.

7.1 Compute $\bar{\bar{\mathbf{A}}}_U$

We use the expression

$$\bar{\bar{\mathbf{A}}}_U = \bar{\bar{\mathbf{L}}} \bar{\bar{\mathbf{L}}}^\top \quad (1.4.1)$$

to compute $\bar{\bar{\mathbf{A}}}_U \in \mathbb{R}^{m \times m}$.

If a different weighting of the relations between the elements of $\bar{\bar{\mathbf{L}}}$ is desired, we can use the alternative expression

$$\left(\bar{\bar{\mathbf{A}}}_U\right)_{ij} = \sum_{k=1}^n \left\{ \delta^* \left(\bar{\bar{\mathbf{L}}}, i, k \right) \bar{\bar{\mathbf{L}}}_{jk} \right\} \quad (1.4.1a)$$

where $i, j \in [1, m] \subset \mathbb{Z}^+$.

7.2 Construct \mathbb{U}

We initially construct \mathbb{U} by following the sub-routine (sub algorithm) given below

Data: Connection Matrix for rows, $\bar{\bar{\mathbf{A}}}_U$, and the number of rows in $\bar{\bar{\mathbf{L}}}$, m .

Result: Initial value for each of the $\{\mathbb{U}_i, \mathbb{M}_i\} \in \mathbb{U}$.

begin

for $i = 1 : m$ **do**

$\mathbb{U}_i = \emptyset; \mathbb{M}_i = \emptyset;$

for $j = 1 : m$ **do**

if $\left(\bar{\bar{\mathbf{A}}}_U\right)_{ij} \neq 0$ **then**

$\mathbb{U}_i = \mathbb{U}_i \cup \{j\};$

$\mathbb{M}_i = \mathbb{M}_i \cup \left\{ \left(\bar{\bar{\mathbf{A}}}_U\right)_{ij} \right\};$

end

end

$\mathbb{U}[i, 1] = \mathbb{U}_i; \mathbb{U}[i, 2] = \mathbb{M}_i;$

end

end

Algorithm 1: Computing the initial value for each of the $\{\mathbb{U}_i, \mathbb{M}_i\} \in \mathbb{U}$.

After the completion of this sub-routine, we will make a copy of this initial state of \mathbb{U} . We denote this copy of this initial state as \mathbb{U}_0 .

7.3 Complete the \mathbb{U} Sets

The elements of \mathbb{U} must now be "completed" so that they include any indirectly related elements. Taking advantage of the disjointedness of $\bar{\bar{\mathbf{L}}}$, the "completion" of the elements in \mathbb{U} can be accomplished using the simple subroutine below

Data: Initial values for the elements of \mathbb{U} , $\{\mathbb{U}_i, \mathbb{M}_i\}$, and the number of rows in $\bar{\bar{\mathbf{L}}}$, m .

Result: The final values for each of the $\{\mathbb{U}_i, \mathbb{M}_i\} \in \mathbb{U}$.

begin

```

    boolean isChanged = true ;
    while isChanged do
        isChanged = false ;
        for i = 1 : m do
             $\mathbb{U}_i = \mathbb{U}[i, 1]; \mathbb{M}_i = \mathbb{U}[i, 2];$ 
            for j = 1 : m do
                 $\mathbb{U}_j = \mathbb{U}[j, 1]; \mathbb{M}_j = \mathbb{U}[j, 2];$ 
                if  $\mathbb{U}_i \neq \mathbb{U}_j \ \&\& \ \mathbb{U}_i \cap \mathbb{U}_j \neq \emptyset$  then
                     $\mathbb{U}_i = \mathbb{U}_i \cup \{\mathbb{U}_j, \mathbb{U}_i\};$ 
                     $\mathbb{M}_i = \mathbb{M}_i \cup \{\mathbb{M}_j, \mathbb{M}_i\};$ 
                     $\mathbb{U}_j = \mathbb{U}_j \cup \{\mathbb{U}_i, \mathbb{U}_j\};$ 
                     $\mathbb{M}_j = \mathbb{M}_j \cup \{\mathbb{M}_i, \mathbb{M}_j\};$ 
                    isChanged = true ;
                end
                if isChanged then
                     $\mathbb{U}[j, 1] = \mathbb{U}_j; \mathbb{U}[j, 2] = \mathbb{M}_j;$ 
                end
            end
            if isChanged then
                 $\mathbb{U}[i, 1] = \mathbb{U}_i; \mathbb{U}[i, 2] = \mathbb{M}_i;$ 
            end
        end
    end
end
```

Algorithm 2: Computing the final value for each $\{\mathbb{U}_i, \mathbb{M}_i\} \in \mathbb{U}$.

7.4 Construct the \mathbb{U}^* Sets

To construct the sets of \mathbb{U}^* , we look for unique $\mathbb{U}_i \in \mathbb{U}$ and then store each unique \mathbb{U}_i as its own set in \mathbb{U}^* . We accomplish this by using the following subroutine

Data: The final values for the elements of \mathbb{U} , $\{\mathbb{U}_i, \mathbb{M}_i\}$, and the number of rows in $\bar{\bar{\mathbf{L}}}$, m .
Result: The sets of \mathbb{U}^* , with each set in \mathbb{U}^* representing a row cluster and containing its elements.

```

begin
  int nClust = 1;  $\mathbb{U}^*[nClust] = \mathbb{U}[1, 1]$  ;
  for  $i = 2 : m$  do
    boolean isNew = true;  $\mathbb{U}_i = \mathbb{U}[i, 1]$  ;
    for  $j = 1 : nClust$  do
       $\mathbb{U}_j^* = \mathbb{U}^*[j]$  ;
      if  $\mathbb{U}_j^* = \mathbb{U}_i$  then
        isNew = false ;
        break ;
      end
    end
    if isNew then
      nClust ++ ;
       $\mathbb{U}^*[nClust] = \mathbb{U}_i$  ;
    end
  end
end

```

Algorithm 3: Compute the set of all row clusters sets, \mathbb{U}^* .

7.5 Compute $\bar{\bar{\mathbf{A}}}_V$

We use the expression

$$\bar{\bar{\mathbf{A}}}_V = \bar{\bar{\mathbf{L}}}^\top \bar{\bar{\mathbf{L}}} \quad (1.4.2)$$

to compute $\bar{\bar{\mathbf{A}}}_V \in \mathbb{R}^{n \times n}$.

If a different weighting of the relations between the elements of $\bar{\bar{\mathbf{L}}}$ is desired, we can use the alternative expression

$$\left(\bar{\bar{\mathbf{A}}}_V\right)_{ij} = \sum_{k=1}^m \left\{ \delta^* \left(\bar{\mathbf{L}}, k, i \right) \bar{\mathbf{L}}_{kj} \right\} \quad (1.4.2a)$$

where $i, j \in [1, n] \subset \mathbb{Z}^+$.

7.6 Construct \mathbb{V}

We initially construct \mathbb{V} by following the sub-routine (sub algorithm) given below

Data: Connection Matrix for columns, $\bar{\bar{\mathbf{A}}}_V$, and the number of columns in $\bar{\mathbf{L}}$, n .

Result: Initial value for each of the $\{\mathbb{V}_j, \mathbb{N}_j\} \in \mathbb{V}$.

begin

for $j = 1 : n$ **do**

$\mathbb{V}_j = \emptyset; \mathbb{N}_j = \emptyset;$

for $i = 1 : n$ **do**

if $\left(\bar{\bar{\mathbf{A}}}_V\right)_{ji} \neq 0$ **then**

$\mathbb{V}_j = \mathbb{V}_j \cup \{i\};$

$\mathbb{N}_j = \mathbb{N}_j \cup \left\{ \left(\bar{\bar{\mathbf{A}}}_V\right)_{ji} \right\};$

end

end

$\mathbb{V}[j, 1] = \mathbb{V}_j; \mathbb{V}[j, 2] = \mathbb{N}_j;$

end

end

Algorithm 4: Computing the initial value for each of the $\{\mathbb{V}_j, \mathbb{N}_j\} \in \mathbb{V}$.

After the completion of this sub-routine, we will make a copy of this initial state of \mathbb{V} . We denote this copy of this initial state as \mathbb{V}_0 .

7.7 Complete the \mathbb{V} Sets

The elements of \mathbb{V} must now be "completed" so that they include any indirectly related elements. Taking advantage of the disjointedness of $\bar{\bar{\mathbf{L}}}$, the "completion" of the elements in \mathbb{V} can be accomplished using the simple subroutine below

Data: Initial values for the elements of \mathbb{V} , $\{\mathbb{V}_j, \mathbb{N}_j\}$, and the number of columns in $\bar{\bar{\mathbf{L}}}$, n .

Result: The final values for each of the $\{\mathbb{V}_j, \mathbb{N}_j\} \in \mathbb{V}$.

begin

```

    boolean isChanged = true ;
    while isChanged do
        isChanged = false ;
        for  $j = 1 : n$  do
             $\mathbb{V}_j = \mathbb{V}[j, 1]; \mathbb{N}_j = \mathbb{V}[j, 2];$ 
            for  $i = 1 : n$  do
                 $\mathbb{V}_i = \mathbb{V}[i, 1]; \mathbb{N}_i = \mathbb{V}[i, 2];$ 
                if  $\mathbb{V}_j \neq \mathbb{V}_i \ \&\& \ \mathbb{V}_j \cap \mathbb{V}_i \neq \emptyset$  then
                     $\mathbb{V}_j = \mathbb{V}_j \cup \{\mathbb{V}_i\};$ 
                     $\mathbb{N}_j = \mathbb{N}_j \cup \{\mathbb{N}_i\};$ 
                     $\mathbb{V}_i = \mathbb{V}_i \cup \{\mathbb{V}_j\};$ 
                     $\mathbb{N}_i = \mathbb{N}_i \cup \{\mathbb{N}_j\};$ 
                    isChanged = true ;
                end
                if isChanged then
                     $\mathbb{V}[i, 1] = \mathbb{V}_i; \mathbb{V}[i, 2] = \mathbb{N}_i;$ 
                end
            end
            if isChanged then
                 $\mathbb{V}[j, 1] = \mathbb{V}_j; \mathbb{V}[j, 2] = \mathbb{N}_j;$ 
            end
        end
    end
end

```

Algorithm 5: Computing the final value for each $\{\mathbb{V}_j, \mathbb{N}_j\} \in \mathbb{V}$.

7.8 Construct the \mathbb{V}^* Sets

To construct the sets of \mathbb{V}^* , we look for unique $\mathbb{V}_i \in \mathbb{V}$ and then store each unique \mathbb{V}_i as its own set in \mathbb{V}^* . We accomplish this by using the following subroutine

Data: The final values for the elements of \mathbb{V} , $\{\mathbb{V}_i, \mathbb{M}_i\}$, and the number of columns in $\bar{\mathbf{L}}$, n .
Result: The sets of \mathbb{V}^* , with each set in \mathbb{V}^* representing a column cluster and containing its elements.

```

begin
  int nClust = 1;  $\mathbb{V}^*[nClust] = \mathbb{V}[1, 1]$  ;
  for  $i = 2 : n$  do
    boolean isNew = true;  $\mathbb{V}_i = \mathbb{V}[i, 1]$  ;
    for  $j = 1 : nClust$  do
       $\mathbb{V}_j^* = \mathbb{V}^*[j]$  ;
      if  $\mathbb{V}_j^* = \mathbb{V}_i$  then
        isNew = false ;
        break ;
      end
    end
    if isNew then
      nClust ++ ;
       $\mathbb{V}^*[nClust] = \mathbb{V}_i$  ;
    end
  end
end

```

Algorithm 6: Compute the set of all column clusters sets, \mathbb{V}^* .

Chapter 8

CbC Non-Disjoint Sets

Connection by Clustering on Non-Disjoint Algorithm info goes here