Suppose that there are 25 elements a-y.

In the gold standard clustering, there are three clusters:
\[ A = \{ a,b,c,d,e,f,g,h,i,j \} \]
\[ B = \{ k,l,m,n,o \} \]
\[ C = \{ p,q,r,s,t,u,v,w,x,y \} \]

The clustering algorithm outputs three clusters:
\[ X = \{ a,b,c,d,e,f,g,k,l,m,p \} \]
\[ Y = \{ h,i,n,o,q \} \]
\[ Z = \{ j,r,s,t,u,v,w,x,y \} \]

So we can write the following table.

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td>7</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>Y</td>
<td>2</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Z</td>
<td>1</td>
<td>0</td>
<td>8</td>
</tr>
</tbody>
</table>

The elements of the table are the size of the intersection of the row and column label. For instance \( A \cap X = \{ a,b,c,d,e,f,g \} \) so the size is 7.

Now consider the space of all pairs of elements. Since there are 25 elements in total, the number of pairs is
\[
\binom{25}{2} = \frac{25 \cdot (25 - 1)}{2} = 300
\]

The two clustering schemes divide these pairs into four categories.

- **True positives**: Pairs \( \alpha - \beta \) which are in the same cluster in both the gold standard and the algorithmic clustering; for example, a-b, a-c, n-o, w-x. These must be in the same cell in the above table. If a cell \( C \) has \( k \) elements, then it contributes \( k(k-1)/2 \) such pairs. So the total number of true positives is \( 7 \cdot 6/2 + 3 \cdot 2/2 + 2 \cdot 1/2 + 2 \cdot 1/2 + 8 \cdot 7/2 = 21 + 3 + 1 + 1 + 28 = 54 \).
  (I have omitted the cells with 1 or 0 elements since they have 0 cells.)

- **False positives**: Pairs \( \alpha - \beta \) which are in the same cluster in the algorithmic cluster but different clusters in the gold standard. Examples: a-p, h-q, j-y. The two elements \( \alpha \) and \( \beta \) must be in two different cells that are in the same row but not the same column. If \( \alpha \) is in cell \( F \) and \( \beta \) is in cell \( G \), then the number of edges connecting \( \alpha \) to \( \beta \) is \( |F| \cdot |G| \). The following combinations of cells are in this category:
  - A,X-B,X with 7 \cdot 3 pairs.
  - A,X-C,X with 7 \cdot 1 pairs.
  - B,X-C,X with 3 \cdot 1 pairs.
  - A,Y-B,Y with 2 \cdot 2 pairs.
  - A,Y-C,Y with 2 \cdot 1 pairs.
  - B,Y-C,Y with 2 \cdot 1 pairs.
  - A,Z-C,Z with 1 \cdot 8 pairs.

So the total number of pairs is \( 21 + 7 + 3 + 4 + 2 + 2 + 8 = 47 \) pairs.
• False negatives: Pairs $\alpha - \beta$ which are in the same cluster in the gold standard, but different algorithmic clusters. Examples: a-j, k-o, r-y. The two elements $\alpha$ and $\beta$ must be in two different cells that are in the same column but not the same row. The following combinations of cells are in this category:
  A,X-A,Y with 7 \cdot 2 pairs.
  A,X-A,Z with 7 \cdot 1 pairs.
  A,Y-A,Z with 2 \cdot 1 pairs.
  B,X-B,Y with 3 \cdot 2 pairs.
  C,X-C,Y with 1 \cdot 1 pairs.
  C,X-C,Z with 1 \cdot 8 pairs.
  C,Y-C,Z with 1 \cdot 8 pairs.
So the total number of pairs is 14+7+2+6+1+8+8 = 46.

• True negatives: Pairs $\alpha - \beta$ that are in different clusters in both the gold standard and the algorithmic clustering. These must be in cells in different rows and columns. The following combinations of cells are in this category:
  A,X-B,Y with 7 \cdot 2 pairs.
  A,X-C,Y with 7 \cdot 1 pairs.
  A,X-C,Z with 7 \cdot 8 pairs.
  A,Y-B,X with 2 \cdot 3 pairs.
  A,Y-C,X with 2 \cdot 1 pairs.
  A,Y-C,Z with 2 \cdot 8 pairs.
  A,Z-B,X with 1 \cdot 3 pairs.
  A,Z-B,Y with 1 \cdot 2 pairs.
  A,Z-C,X with 1 \cdot 1 pairs.
  A,Z-C,Y with 1 \cdot 1 pairs.
  B,X-C,Y with 3 \cdot 1 pairs.
  B,X-C,Z with 3 \cdot 8 pairs.
  B,Y-C,X with 2 \cdot 1 pairs.
  B,Y-C,Z with 2 \cdot 8 pairs.
So the total number of pairs is 14+7+56+6+16+3+2+1+1+3+24+2+8 = 153.

Note that 54+47+46+153 = 300, so all the pairs are accounted for.

Now,

Accuracy = (TP+TN) / Total = 207/300 = 0.69.
Precision = TP / (TP+FP) = 54/101 = 0.5346.
Recall = TP / (TP + FN) = 54/100 = 0.54.
F score = 2PR/(P+R) = 0.5373