# Chords diagrams

Chord diagrams shows relationships among groups of entities. Each block on the exterior circle represents an entity. The links between entities represent shared characteristics of the entities. Additionally, the size of the link area is proportional to the number of common characteristics of the entities. The ticks on the exterior side of the circle aim at providing guides to the number of common characteristics.

## First, second, third and other lines

### Functions in lines

An hypothesis is that functions are attached to lines. If this hypothesis were correct, a given function would be always performed on a given line. Then, lines would be considered as separated by functions.

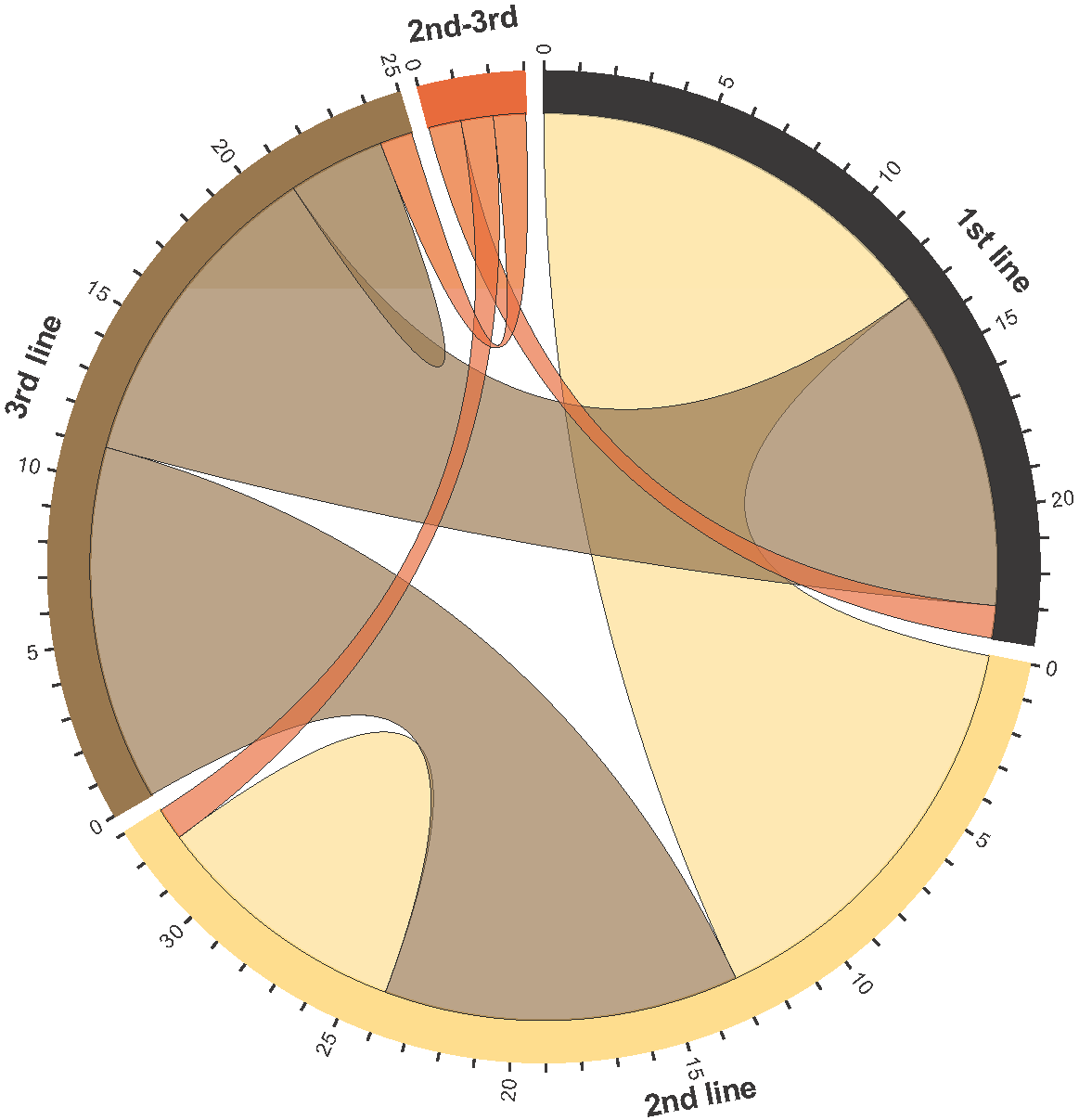
However, the analysis of the incident dataset leads to the disproval of this hypothesis: functions are mainly scattered among lines, as presented on the chord diagram in Figure XXX. On this diagram, four lines “1st line”, “2nd line”, “3rd line”, and “2nd-3rd“, matching the line names provided in the provided data sets, are represented by 4 separated blocks on the outside circle. Two blocks are linked according to the number of functions that they both perform. A particular case is the case of functions performed by a unique line. In this case, an area attached to a single block is drawn.

Figure Function sharing across lines

From the chord diagram, one may identified that no function performed by the 1st line is performed only by the 1st line. 13 functions performed by the 1st line are performed by the 2nd line too, 10 functions are performed by the 3rd line. 11 functions are performed both by the 2nd and 3rd lines. The “2nd-3rd “ line performed 1 function in common with the 1st line, another function with the 2nd line, and a last function with the 3rd line. Only 8 functions are performed only by the 2nd line, and 3 functions are performed only by the 3rd line.

A summary of the number of functions performed by two lines is presented in Table XXX.

Table Functions co-performed by lines

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | 1st line | 2nd line | 3rd line | 2nd-3rd line |
| 1st line | 0 | 13 | 10 | 1 |
| 2nd line | 13 | 8 | 11 | 1 |
| 3rd line | 10 | 11 | 3 | 1 |
| 2nd-3rd line | 1 | 1 | 1 | 0 |

### Organization lines in lines

An hypothesis is that organization lines are spread amongst lines, assuming that the competences are scattered within each organization line among all the lines. If this hypothesis were correct, each organization line would be operating on each line and no organization line would be confined to a single line.

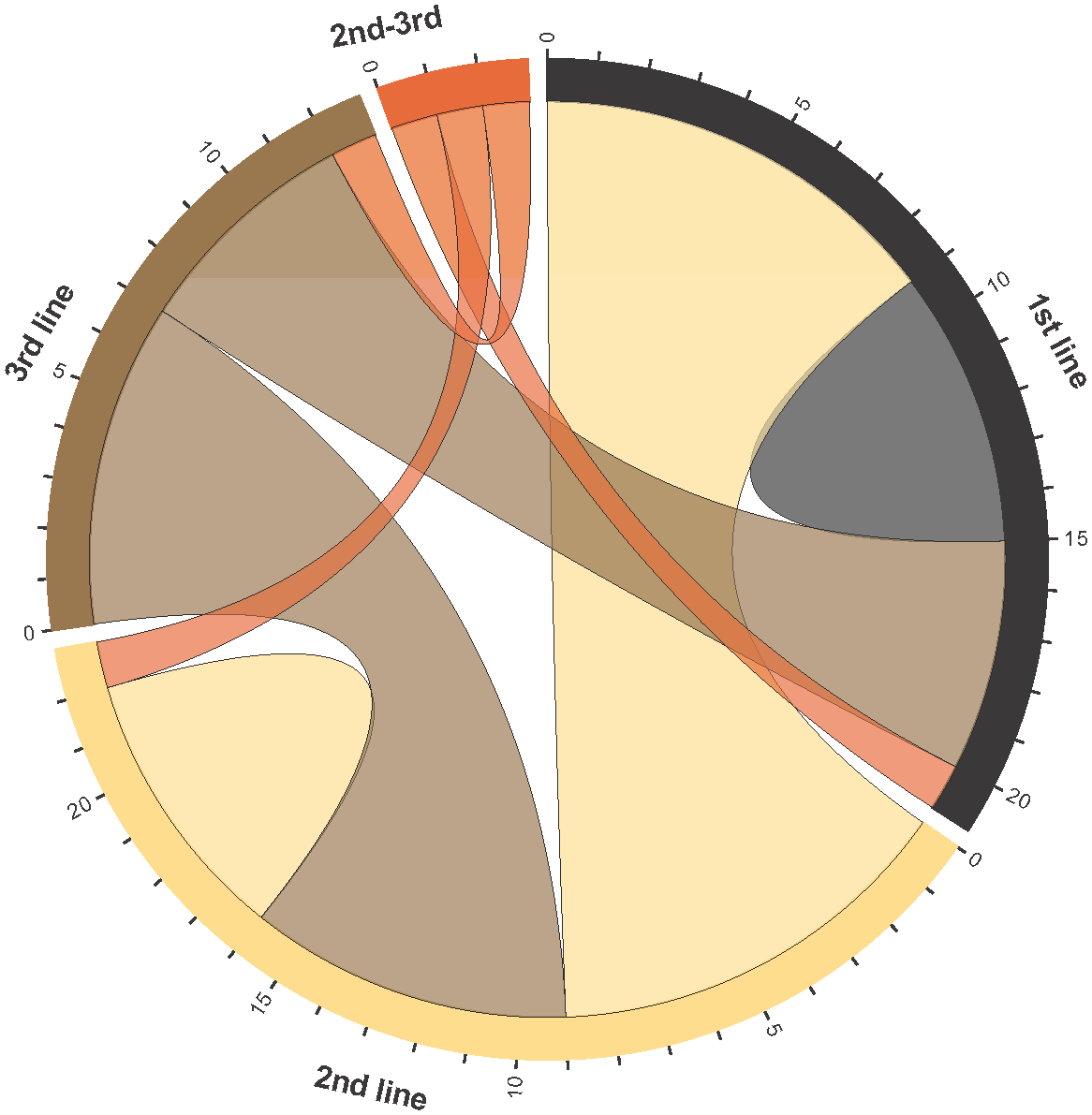
The analysis of the incident dataset leads to a partial confirmation of this hypothesis: organization lines are mainly participating to many lines, as presented on the chord diagram in Figure XXX. On this diagram, four lines “1st line”, “2nd line”, “3rd line”, and “2nd-3rd“, matching the line names provided in the provided data sets, are represented by 4 separated blocks on the outside circle. Two blocks are linked according to the number of organization lines participating in these lines. A particular case is the case of organization lines performing on a single line. In this case, an area attached to a single block is drawn.

Figure Organization lines across lines

From the chord diagram, on may identify that 9 organization lines are active on both the 1st and the 2nd lines, 5 organization lines are active on both the 1st and the 3rd lines, and 7 organization lines are active on both the 2nd and the 3rd lines. One organization line active on the “2nd-3rd” line is also active on the 1st line, another one on the 2nd line, and a last one on the third line. Therefore, organization lines tends to participate in more than one line.

However, 6 organization lines operate only on the 1st line and 6 organization lines operate only on the 2nd lines. Therefore, the hypothesis of a systematic spreading of organization lines across lines is not fully acceptable.

A summary of the number of organization lines operating on various lines is presented in Table XXX.

Table Functions co-performed by lines

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | 1st line | 2nd line | 3rd line | 2nd-3rd line |
| 1st line | 6 | 9 | 5 | 1 |
| 2nd line | 9 | 6 | 7 | 1 |
| 3rd line | 5 | 7 | 0 | 1 |
| 2nd-3rd line | 1 | 1 | 1 | 0 |

### Persons in lines

An hypothesis is that persons are attached to a given line. If this hypothesis were correct, a given person would performed only on the 1st line, 2nd line or 3rd line, and no person would operate on two or more lines.

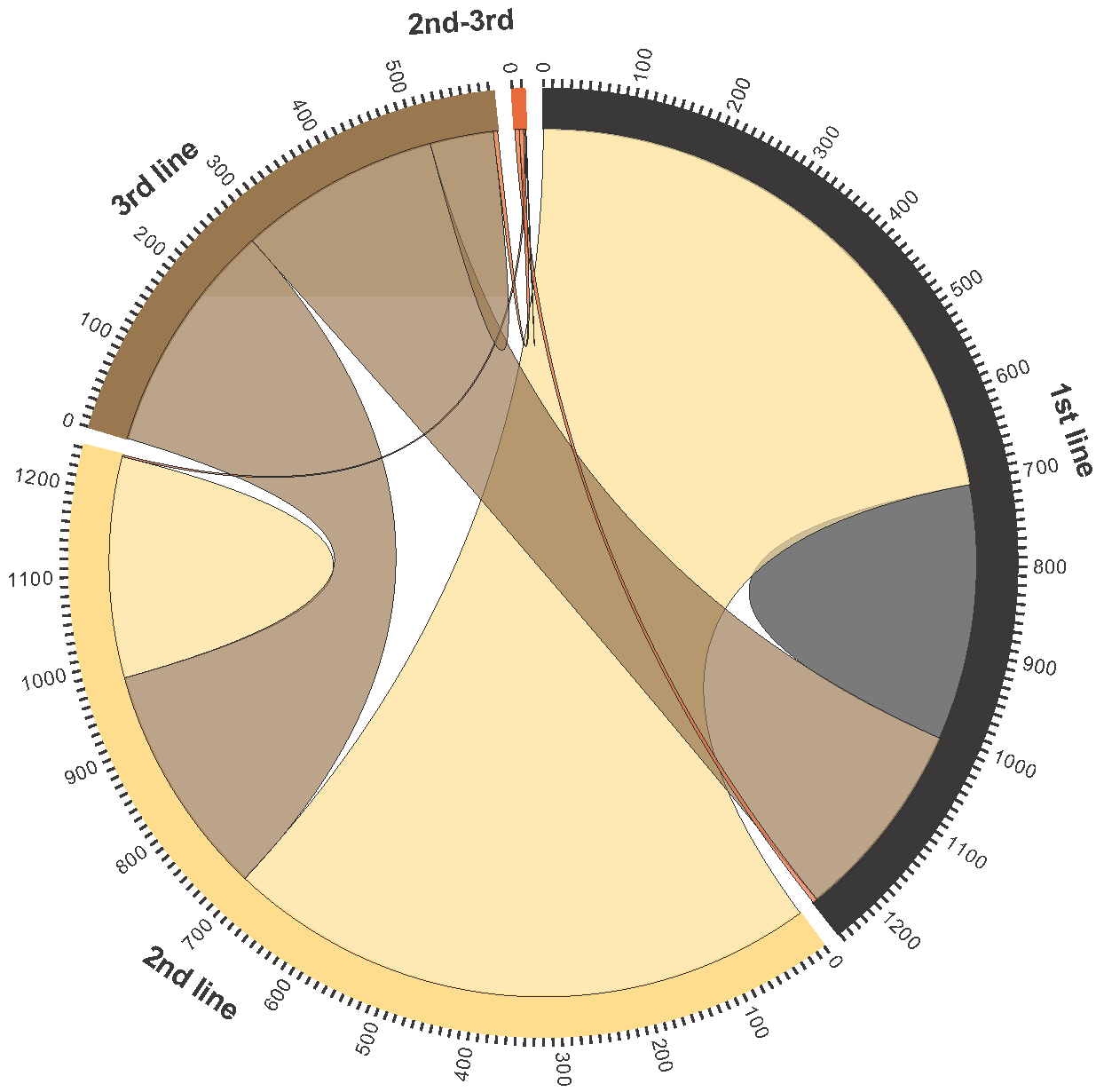
The analysis of the incident dataset leads to the disproval of this hypothesis: a large number of persons is operating on many lines, as presented on the chord diagram in Figure XXX. On this diagram, four lines “1st line”, “2nd line”, “3rd line”, and “2nd-3rd“, matching the line names provided in the provided data sets, are represented by 4 separated blocks on the outside circle. Two blocks are linked according to the number of persons operating on these lines. A particular case is the case of persons operating on a single line. In this case, an area attached to a single block is drawn.

Figure Person involvement in lines

About 700 persons are operating at both the 1st and 2nd lines. About 250 persons are operating at both the 1st and the 3rd lines. Almost 300 persons are operating at both the 2nd and the 3rd lines. The persons operating on the “2nd-3rd” line are operating at the 1st line (5 persons), 2nd line (2 persons), and the 3rd line (5 persons). Therefore, most persons are operating at two lines or more.

However, still a large number of persons are operating on a single line: about 300 persons are working only on the 1st line, about 250 persons are working only on the 2nd line, and less than 100 persons are working only on the 3rd line. Therefore, a core of persons are focusing only on a single line while a large number of persons are operating in a cross-line mode.

A summary of the number of persons operating on various lines is presented in Table XXX.

Table Persons by lines

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | 1st line | 2nd line | 3rd line | 2nd-3rd line |
| 1st line | 303 | 704 | 239 | 5 |
| 2nd line | 704 | 261 | 276 | 2 |
| 3rd line | 239 | 276 | 75 | 5 |
| 2nd-3rd line | 5 | 2 | 5 | 1 |

### STs in lines

An hypothesis is that support teams (STs) are attached to a given line. If this hypothesis were correct, a given ST would operate only on a given line, and no ST would operate on two lines.

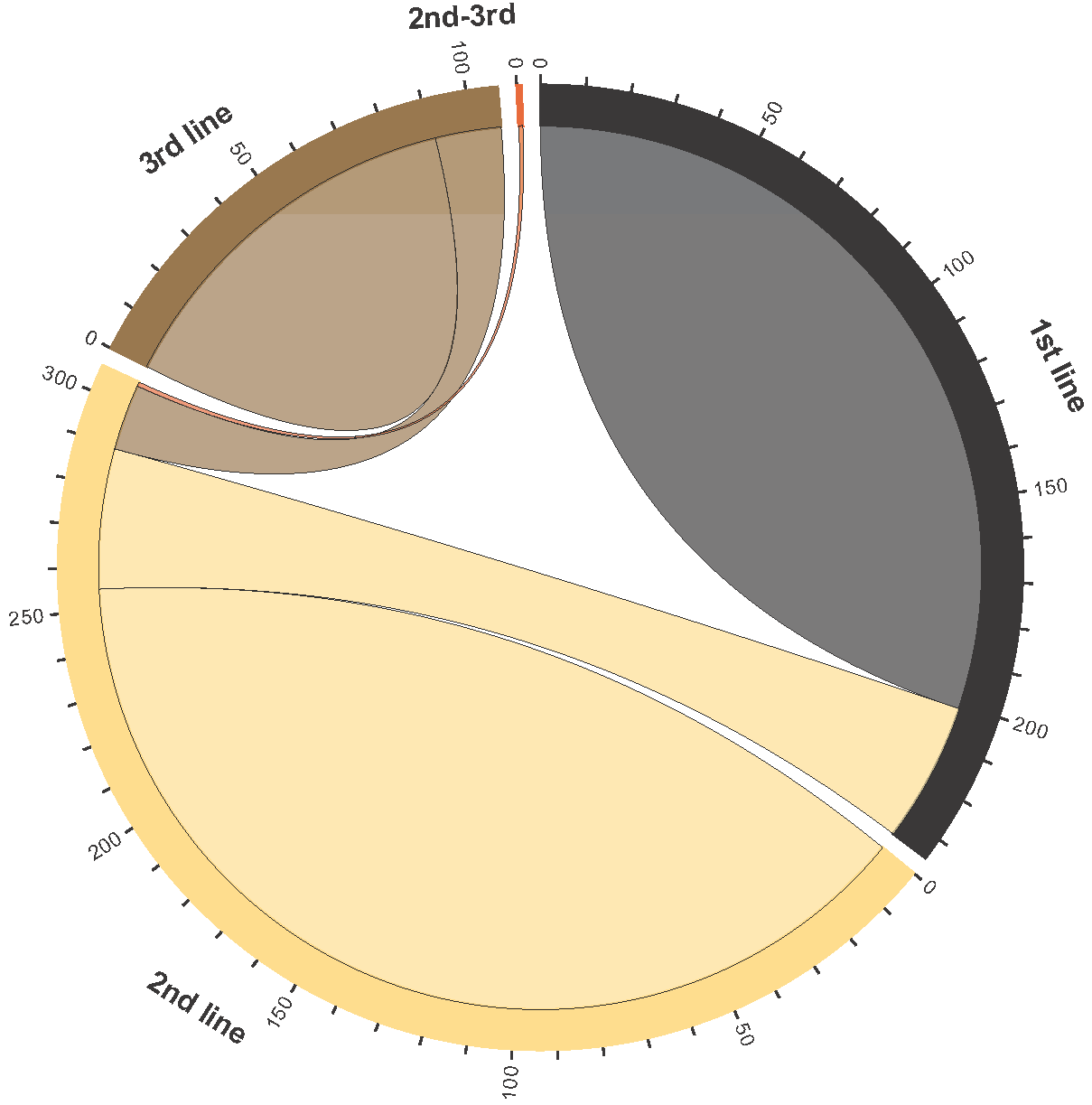
The analysis of the incident dataset leads to a partial confirmation of this hypothesis. Indeed, a large number of STs are operating on a single line. However, a significant number of STs are operating on two lines, as presented on the chord diagram in Figure XXX. . On this diagram, four lines “1st line”, “2nd line”, “3rd line”, and “2nd-3rd“, matching the line names provided in the provided data sets, are represented by 4 separated blocks on the outside circle. Two blocks are linked according to the number of STs operating on these lines. A particular case is the case of STs operating on a single line. In this case, an area attached to a single block is drawn.

Figure STs operating in lines

About 209 STs are performing only on the 1st line. About 250 STs are performing only on the 2nd line. About 90 STs are performing only on the 3rd line. Most STs are confided to a given ST. However, about 30 STs are performing on both the 1st and the 2nd lines, while about 15 STs are performing on both the 2nd and the 3rd lines. The unique ST on the “2nd-3rd” line is performing also on the second line. An interesting aspect is the lack of ST operating on both the 1st and the 3rd line.

Therefore, from this analysis, STs are mostly confined to a given line. However, some STs are connecting the 1st line to the 2nd line as well as the 2nd line to the 3rd line, which may be considered as “bridges” gradually linking the lines.

A summary of the number of STs operating on various lines is presented in Table XXX.

Table STs by lines

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | 1st line | 2nd line | 3rd line | 2nd-3rd line |
| 1st line | 201 | 34 | 0 | 0 |
| 2nd line | 34 | 255 | 16 | 1 |
| 3rd line | 0 | 16 | 91 | 0 |
| 2nd-3rd line | 0 | 1 | 0 | 0 |

## Functions

### Functions and persons

An hypothesis is that persons are performing a given function, which is aligned with a function-oriented management. If this hypothesis were correct, a given person would focus on a given function and specialization would be the rule for employees.

The analysis of the incident dataset leads to a disapproval of this hypothesis. Indeed, a very small number of persons are performing a unique function (cf. the diagonal in the data presented in Table XXX), as presented on the chord diagram in Figure XXX. . On this diagram, each block on the outside circle represents a function. Two blocks are linked according to the number of persons performing the associated functions.

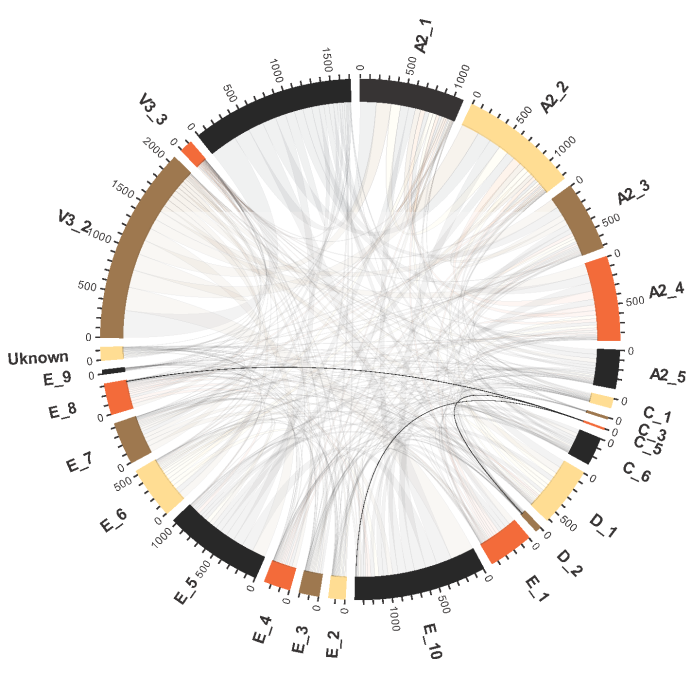
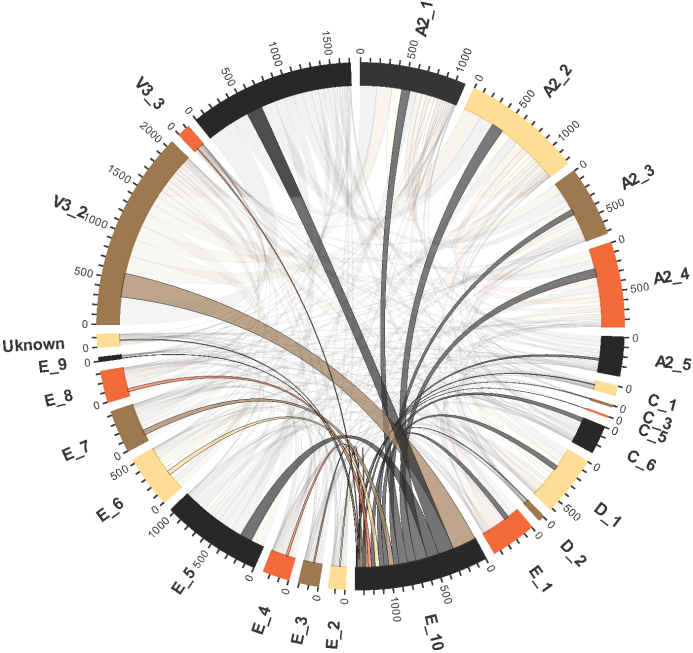
For almost each pair of functions, it exists a person that performs both functions. The rare exception is function C\_5. Persons performing function C\_5 are either performing D\_2, E\_8, or E\_10 (cf. Figure YYY). Therefore function C\_5 is either rare, or requires a very narrow of expertise. Another interesting fact is that persons performing function C\_3 are performing most existing functions but C\_5, D\_2, E\_8. Therefore persons performing function C\_3 seem to “avoid” functions performed by persons performing C\_5.

Figure Functions performed together by employees. On the left side, the set of functions performed by employees performing C\_5 are highlighted. On the right side, the set of functions performed by employees performing E\_10 are highlighted.

A summary of the number of persons performing various functions is presented in Table XXX.

Table Number of persons performing functions

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | A2\_1 | A2\_2 | A2\_3 | A2\_4 | A2\_5 | C\_1 | C\_3 | C\_5 | C\_6 | D\_1 | D\_2 | E\_1 | E\_10 | E\_2 | E\_3 | E\_4 | E\_5 | E\_6 | E\_7 | E\_8 | E\_9 | Uknown | V3\_2 | V3\_3 |  |
| A2\_1 | **30** | 111 | 61 | 53 | 49 | 7 | 0 | 0 | 16 | 42 | 2 | 35 | 94 | 16 | 12 | 20 | 78 | 46 | 28 | 22 | 3 | 9 | 170 | 8 | 30 |
| A2\_2 | 111 | **31** | 65 | 94 | 35 | 8 | 1 | 0 | 17 | 55 | 3 | 37 | 131 | 16 | 14 | 17 | 96 | 51 | 42 | 24 | 3 | 9 | 203 | 11 | 111 |
| A2\_3 | 61 | 65 | **15** | 38 | 25 | 4 | 1 | 0 | 17 | 33 | 1 | 24 | 61 | 9 | 14 | 12 | 50 | 33 | 17 | 12 | 2 | 10 | 107 | 7 | 61 |
| A2\_4 | 53 | 94 | 38 | **64** | 34 | 5 | 1 | 0 | 10 | 41 | 3 | 27 | 96 | 8 | 13 | 11 | 66 | 32 | 27 | 19 | 2 | 8 | 119 | 6 | 53 |
| A2\_5 | 49 | 35 | 25 | 34 | **13** | 2 | 0 | 0 | 7 | 18 | 2 | 10 | 24 | 8 | 5 | 5 | 20 | 13 | 6 | 10 | 1 | 2 | 48 | 5 | 49 |
| C\_1 | 7 | 8 | 4 | 5 | 2 | **0** | 0 | 0 | 3 | 4 | 1 | 10 | 8 | 3 | 2 | 2 | 4 | 3 | 2 | 1 | 1 | 3 | 11 | 2 | 7 |
| C\_3 | 0 | 1 | 1 | 1 | 0 | 0 | **0** | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 1 | 0 | 0 |
| C\_5 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | **0** | 0 | 0 | 2 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 3 | 0 | 0 | 0 | 0 | 0 |
| C\_6 | 16 | 17 | 17 | 10 | 7 | 3 | 0 | 0 | **2** | 7 | 3 | 11 | 44 | 3 | 6 | 5 | 27 | 16 | 6 | 4 | 1 | 6 | 42 | 4 | 16 |
| D\_1 | 42 | 55 | 33 | 41 | 18 | 4 | 1 | 0 | 7 | **51** | 3 | 20 | 54 | 7 | 5 | 10 | 40 | 19 | 15 | 20 | 1 | 3 | 78 | 6 | 42 |
| D\_2 | 2 | 3 | 1 | 3 | 2 | 1 | **0** | **2** | 3 | 3 | **1** | 1 | 6 | 1 | 1 | 1 | 3 | 2 | 1 | 5 | 1 | 1 | 5 | 1 | 2 |
| E\_1 | 35 | 37 | 24 | 27 | 10 | 10 | 1 | 0 | 11 | 20 | 1 | **7** | 52 | 9 | 5 | 8 | 37 | 12 | 14 | 11 | 1 | 4 | 66 | 4 | 35 |
| E\_10 | 94 | 131 | 61 | 96 | 24 | 8 | 1 | **1** | 44 | 54 | 6 | 52 | **24** | 13 | 16 | 24 | 133 | 47 | 56 | 35 | 3 | 9 | 268 | 10 | 94 |
| E\_2 | 16 | 16 | 9 | 8 | 8 | 3 | 0 | 0 | 3 | 7 | 1 | 9 | 13 | **2** | 5 | 2 | 10 | 5 | 7 | 3 | 1 | 4 | 17 | 2 | 16 |
| E\_3 | 12 | 14 | 14 | 13 | 5 | 2 | 1 | 0 | 6 | 5 | 1 | 5 | 16 | 5 | **2** | 4 | 20 | 17 | 5 | 2 | 1 | 4 | 21 | 4 | 12 |
| E\_4 | 20 | 17 | 12 | 11 | 5 | 2 | 1 | 0 | 5 | 10 | 1 | 8 | 24 | 2 | 4 | **16** | 36 | 11 | 9 | 5 | 1 | 1 | 44 | 4 | 20 |
| E\_5 | 78 | 96 | 50 | 66 | 20 | 4 | 1 | 0 | 27 | 40 | 3 | 37 | 133 | 10 | 20 | 36 | **3** | 45 | 47 | 35 | 2 | 7 | 150 | 9 | 78 |
| E\_6 | 46 | 51 | 33 | 32 | 13 | 3 | 1 | 0 | 16 | 19 | 2 | 12 | 47 | 5 | 17 | 11 | 45 | **4** | 17 | 11 | 2 | 7 | 77 | 7 | 46 |
| E\_7 | 28 | 42 | 17 | 27 | 6 | 2 | 1 | 0 | 6 | 15 | 1 | 14 | 56 | 7 | 5 | 9 | 47 | 17 | **4** | 13 | 1 | 4 | 68 | 3 | 28 |
| E\_8 | 22 | 24 | 12 | 19 | 10 | 1 | **0** | **3** | 4 | 20 | 5 | 11 | 35 | 3 | 2 | 5 | 35 | 11 | 13 | **7** | 2 | 2 | 43 | 3 | 22 |
| E\_9 | 3 | 3 | 2 | 2 | 1 | 1 | 0 | 0 | 1 | 1 | 1 | 1 | 3 | 1 | 1 | 1 | 2 | 2 | 1 | 2 | **0** | 2 | 3 | 1 | 3 |
| Uknown | 9 | 9 | 10 | 8 | 2 | 3 | 0 | 0 | 6 | 3 | 1 | 4 | 9 | 4 | 4 | 1 | 7 | 7 | 4 | 2 | 2 | **7** | 13 | 1 | 9 |
| V3\_2 | 170 | 203 | 107 | 119 | 48 | 11 | 1 | 0 | 42 | 78 | 5 | 66 | 268 | 17 | 21 | 44 | 150 | 77 | 68 | 43 | 3 | 13 | **191** | 13 | 170 |
| V3\_3 | 8 | 11 | 7 | 6 | 5 | 2 | 0 | 0 | 4 | 6 | 1 | 4 | 10 | 2 | 4 | 4 | 9 | 7 | 3 | 3 | 1 | 1 | 13 | **0** | 8 |
|  | 184 | 158 | 101 | 101 | 55 | 10 | 1 | 0 | 30 | 83 | 2 | 57 | 176 | 16 | 24 | 21 | 118 | 58 | 41 | 27 | 2 | 13 | 309 | 12 | **184** |

### Functions in organization lines

An hypothesis is that a given organization line is responsible for a given function. If this hypothesis were correct, functions would not be performed by various organization lines.

The analysis of the incident dataset leads to a partial confirmation of this hypothesis. Indeed, some organization lines tend to share functions with a large number of organization lines, while other organizations do either share their functions with a very limited number of organization lines, or perform functions in an exclusive manner.

The first group of organization lines sharing their functions with a large number of STs corresponds to the organization lines B, C, E, F, G1, G4, H, and V\* (except V8). These organizations lines share their functions among themselves, which lead to the supposition that this organization lines are either performing a large range of generic operations or that they are very scattered organization lines having a very similar profile. The presence of the large C organization line is to be noted.

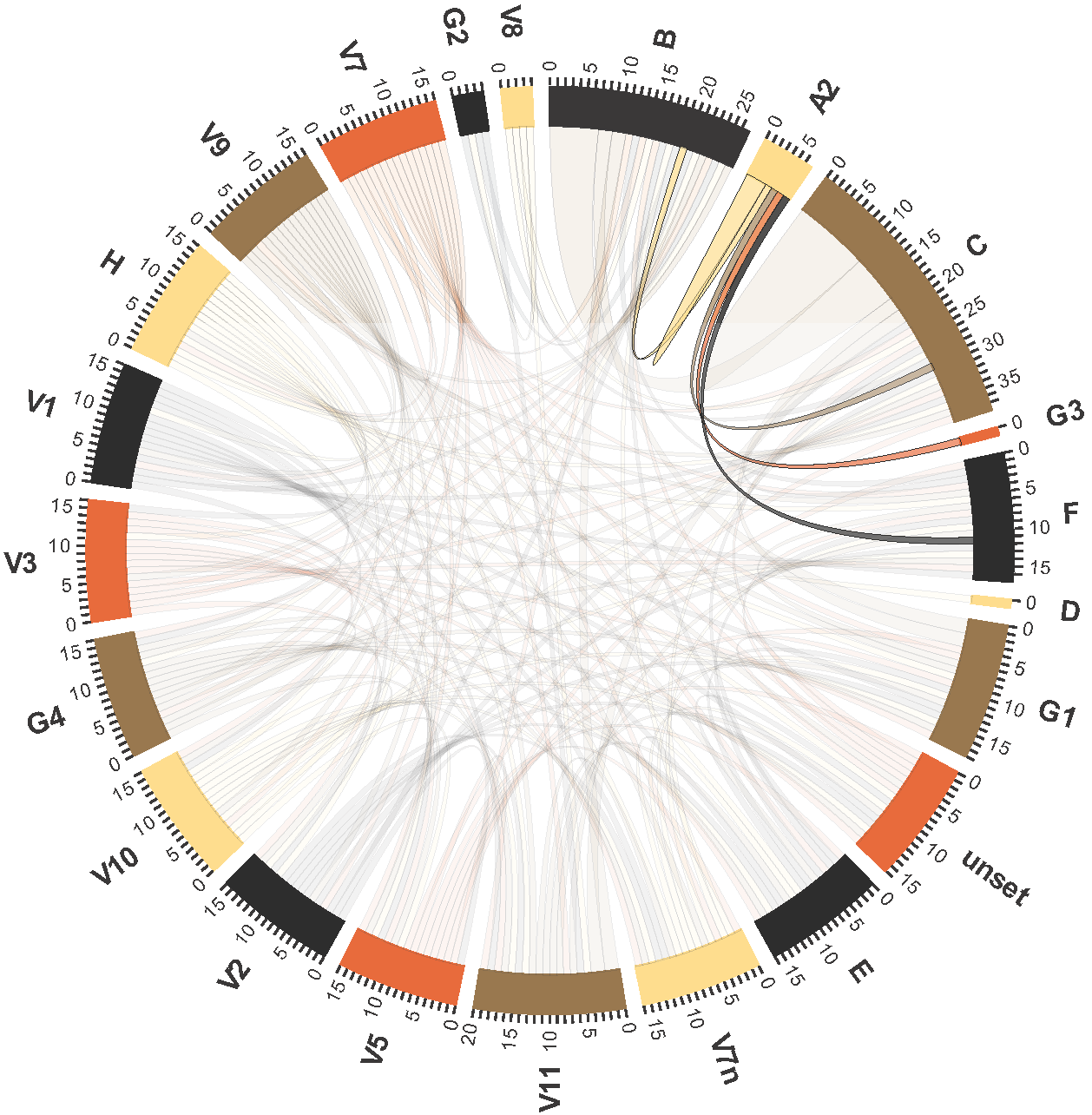
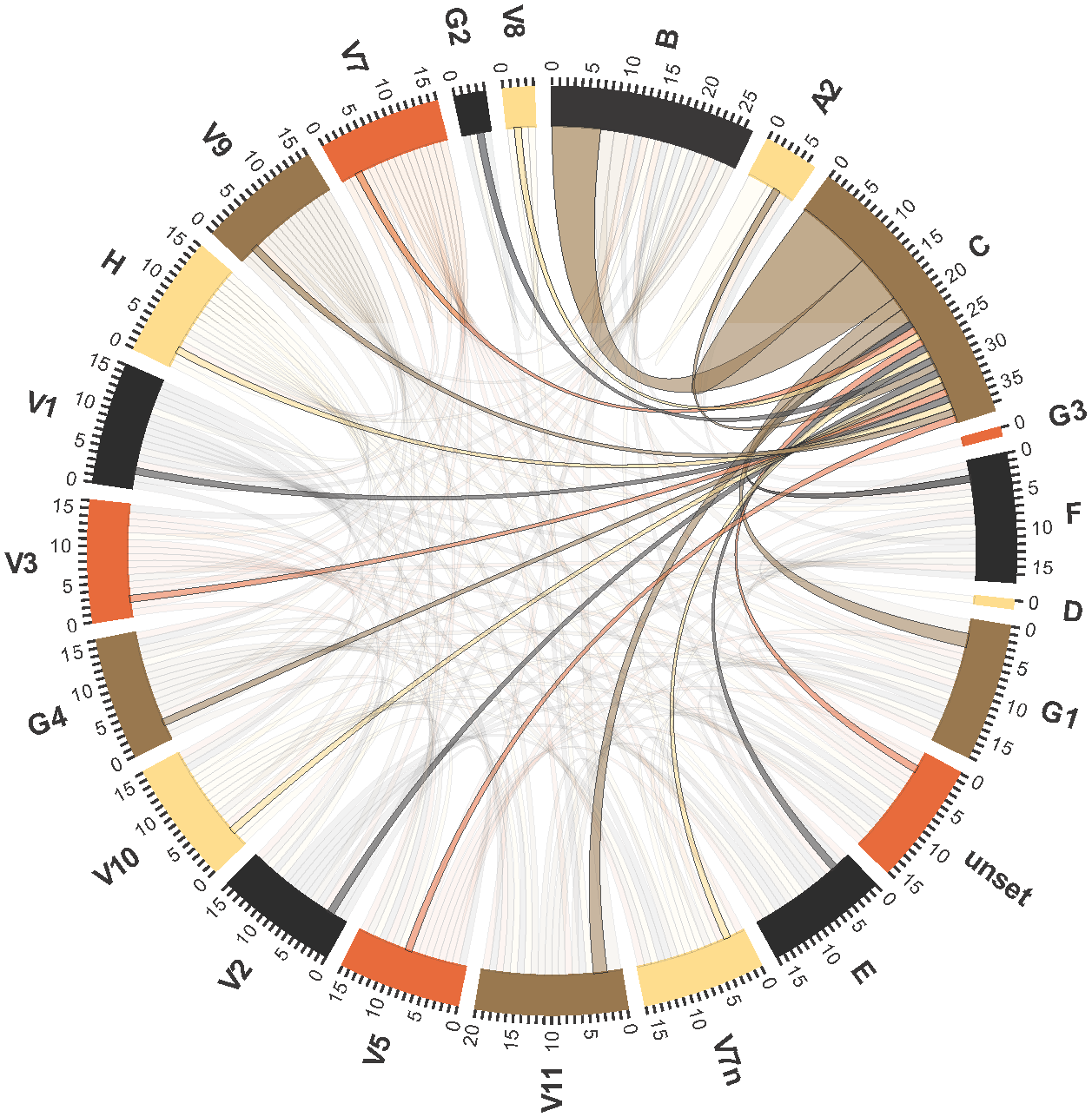
The second group of organization lines contains the organization lines A2, D, G2, G3, and V8. These organization lines don’t share their functions with a large number of organization lines, but rather either share them with organization lines within this second group. The case of A, D, and V8 is special as these organization lines have specific functions that are performed exclusively by these organization lines.

Figure 6 Functions in organization lines. On the left side, the case of the organization line C is highlighted. On the right side, the case of the organization line A2 is highlighted.

The case of the organization line C is interesting as on the one hand organization line C share functions with many organization lines, on the second hand, a large set of functions performed by C is only performed by C.

A summary of the number of persons performing various functions is presented in Table XXX.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | B | A2 | C | G3 | F | D | G1 | unset | E | V7n | V11 | V5 | V2 | V10 | G4 | V3 | V1 | H | V9 | V7 | G2 | V8 |
| B | **1** | 1 | 7 | 0 | 1 | 0 | 2 | 1 | 1 | 1 | 2 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| A2 | 3 | **3** | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| C | 1 | 1 | **11** | 0 | 1 | 0 | 2 | 1 | 1 | 1 | 2 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| G3 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| F | 1 | 1 | 1 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 |
| D | 0 | 0 | 0 | 0 | 0 | **1** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| G1 | 0 | 0 | 2 | 0 | 1 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 |
| unset | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 |
| E | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 |
| V7n | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 |
| V11 | 0 | 0 | 2 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| V5 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 |
| V2 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 |
| V10 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 |
| G4 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 0 | 0 |
| V3 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 0 |
| V1 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 0 | 0 |
| H | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 0 | 0 |
| V9 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 0 |
| V7 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 |
| G2 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| V8 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |

# Cycles

## The cycle graph

A graph, referred to as the *cycle graph*, has been built to capture cycles in process instances.

The case of Siebel is particularly interesting. Seibel has the highest degree in the cycle graph, which means that among all the owners participating into a cycle in a process instance, Siebel is the one connected to the highest number of unique owners.

In the representation of the cycle graph presented in Figure XXX, each node represents a person. The size of each node represents the degree of the person: the bigger the node, the higher the number of connected node. Therefore, a bigger node represents a person involved in cycles with a large number of persons, while a small node represents a person involved in cycles with a small number of persons.

The brightness of a node represent the *betweenness* of a person in the cycle of graph. Betweenness is defined as the ratio of the number of shortest paths through a given node between every two nodes and the overall number of shortest path between every two nodes in a graph. Betweenness represents the importance of a node with regard to information flows. The higher the betweenness of a node, the bigger its important in terms of transmission within the graph. In Figure XXX, the darker a node, the higher its betweenness.

Finally, nodes are laid out as follow in Figure XXX. The central node Siebel. Nodes located on the smaller circle are directly connected to Siebel. Nodes located on the second circle are directly connected to a node on the first circle. Therefore, nodes located on the second circle are two hops away from Siebel, that is they have participate in a cycle with someone who has been participating in a cycle with Siebel, or they have participate in a cycle with someone between them and Siebel.

One may notice that the darkest node is Siebel. Siebel is the most central node in the network of cycles in term of betweenness. Most of the cycles linking two persons are containing Siebel. Additionally, the degree of Siebel is the highest, and therefore, the node representing Siebel is the biggest.

Both the betweenness and the degrees of the nodes are fading down when going away from Siebel. The nodes are becoming brighter and smaller.

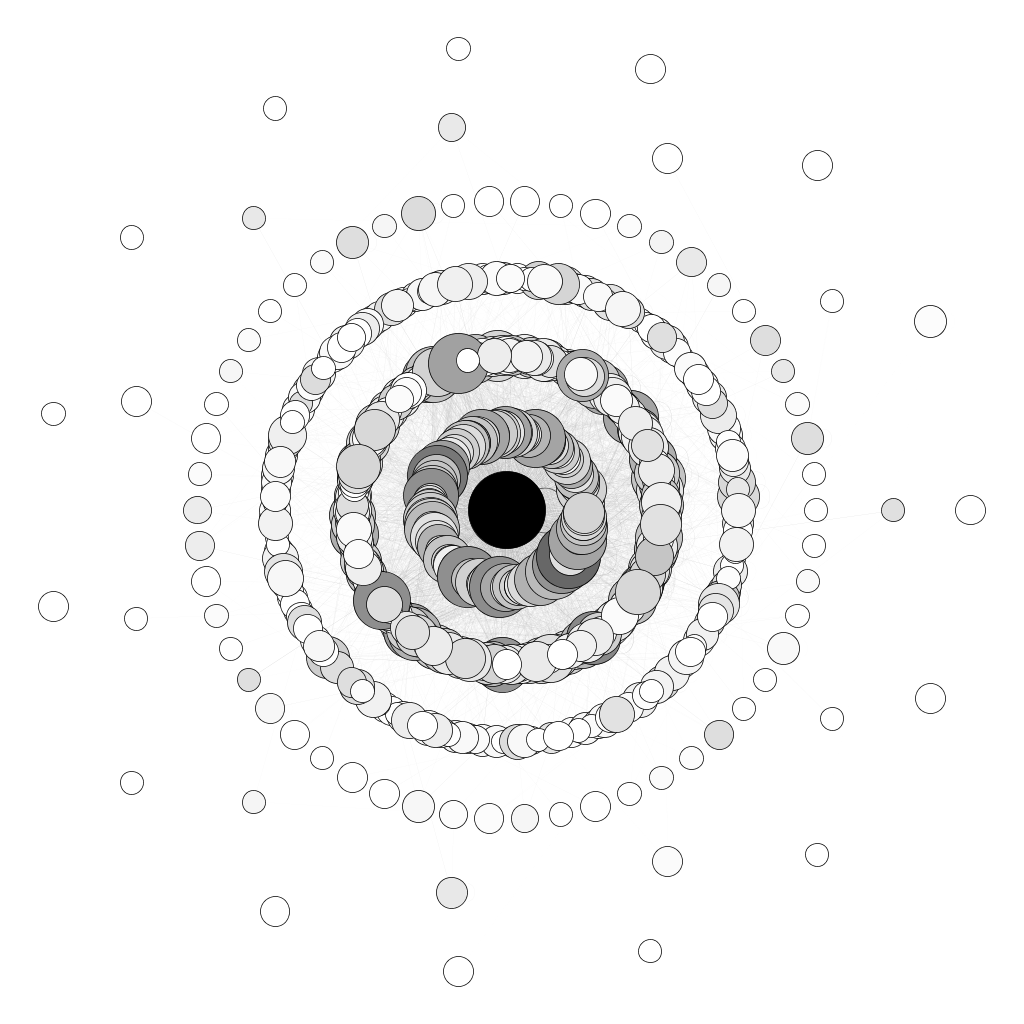


Figure The cycle graph centered on Siebel. The size of the nodes is proportional to their degree, the brightness is proportional to their betweenness centrality.

# Tools

## Data munging

The basic tool for data munging has been **Microsoft Excel 2010**. Its capacities to import CSV file have allow the original datasets to be imported to a tabular format. Next, filtering and sorting capabilities of MS Excel have eased the identification of missing data and have allowed for raw estimations of the count and distribution of the data, such as the number of organization lines.

## Data processing

An important tool for our analyses is the **Python language** and the **Enthought Canopy** environment. Python, in its version 2.7.4, has been used to process the cleaned data to obtain a better understanding of the concept of lines and the relation between lines and persons/organization lines/functions. The data presented in Tables XXX, YYY have been computed in Python.

The following Python modules have been useful for our analyses:

* the **csv** module has been used to import and export data from and to CSV files,
* the **networkx** module has been used to create and analyze graphs, especially the cycle graph used to study ping-pong.

The Enthought Canopy has been chosen as an environment supporting of-the-box the networkx module and encompassing the IPython environment providing support for Python notebooks.

## Data visualization

The visualization of chord diagrams is based on the **D3.js** javascript library, available at http://d3js.org/, and the examples of chord diagrams published by Mike Bostock (http://bl.ocks.org/mbostock/4062006 and http://bl.ocks.org/mbostock/1046712).

The visualization of the cycle graph has been performed in **Gephi** 0.8.2-beta (http://www.gephi.org/). The processing of betweenness and degree centralities has been performed with Gephi, as well as the laying out process and the final rendition of the graph visualization.