Visualizing symmetric square matrices with rainbow boxes:

Methods and application to character co-occurrence matrices in literary texts

Jean-Baptiste Lamy

jean-baptiste.lamy @ univ-paris13.fr

LIMICS
Université Paris 13, Sorbonne Paris Cité, 93017 Bobigny
Sorbonne Universités, Paris
INSERM UMRS 1142

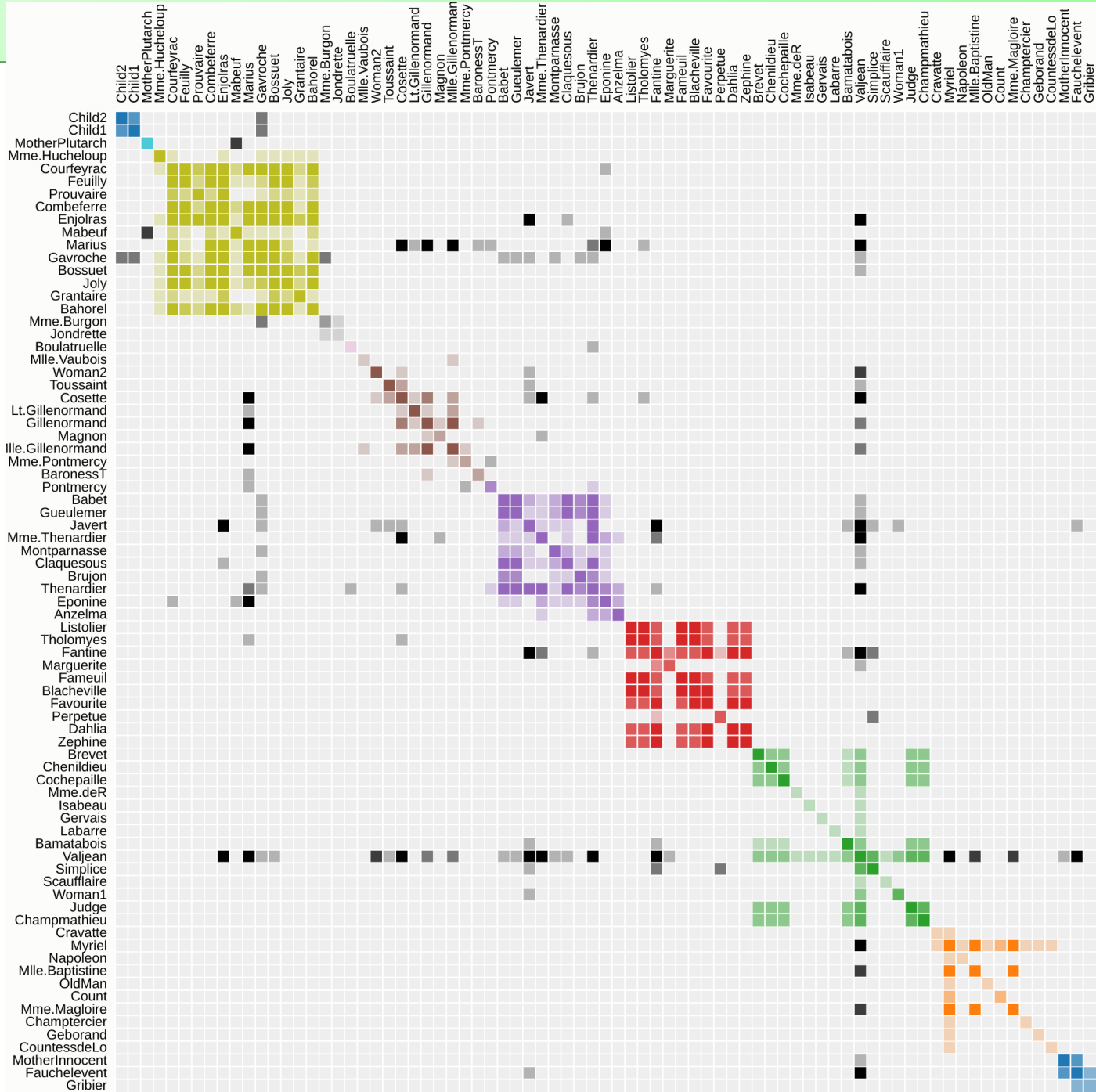
Introduction

- **Symmetric square matrices / co-occurrence matrices [Leydesdorff]**
  - Frequent type of dataset
    - Example: distant reading in digital humanities [Jänicke]
    - Character matrix in a novel

	Tienn	Rakenn	Alyse	Le Capitann	Gauve	Méric	Rashkang	Chéram	Lescantelles	La Rasinne	Werfam	Auguss	Le Lamineur	Nicol	Crépusculine	Auroraline	King Saphir	Thérald	Malfred	Mabine	Érard
Tienn		1	1	1	1	0	0	1	0	2	2	2	2	5	3	0	0	0	4	4	4
Rakenn			1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Alyse				1	1	2	2	2	2	0	0	0	0	0	0	0	0	0	0	0	0
Le Capitann					0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Gauve						2	2	2	2	0	0	0	0	0	0	0	0	0	0	0	0
Méric							2	2	2	0	0	0	0	0	0	0	0	0	0	0	0
Rashkang								2	2	0	0	0	0	0	0	0	0	0	0	0	0
Chéram									1	0	0	0	0	0	0	0	0	0	0	0	0
Lescantelles										1	1	0	0	0	0	0	0	0	0	0	0
La Rasinne											1	1	1	5	0	0	0	0	0	0	0
Werfam												1	1	5	0	0	1	0	0	0	0
Auguss													1	5	0	0	0	0	0	0	0
Le Lamineur														5	0	0	0	0	0	0	0
Nicol															0	0	0	0	0	0	0
Crépusculine																1	1	1	0	0	0
Auroraline																	1	1	0	0	0
King Saphir																		1	0	0	0
Thérald																			0	0	0
Malfred																				4	4
Mabine																					4
Érard																					

Matrix-based visualization

- **Example on ‘*Les Misérables*’ (Victor Hugo)**
- **Problem**
  - A character can belong to at most one group / cluster

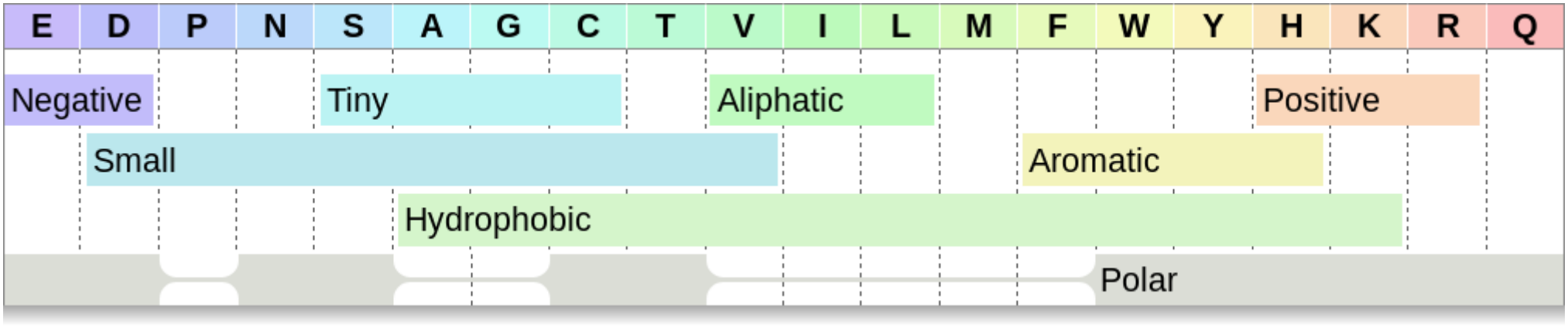


Objectives

- **Propose a visualization techniques for symmetric square matrices**
  - By transforming the matrix into overlapping sets
  - And visualizing these overlapping sets with rainbow boxes
- **Focus on the representation of subsets of interrelated elements**
  - For example in a novel: groups of characters that know each other
- **Rainbow boxes have been used only in the biomedical domain**
  - => a new application domain

Rainbow boxes

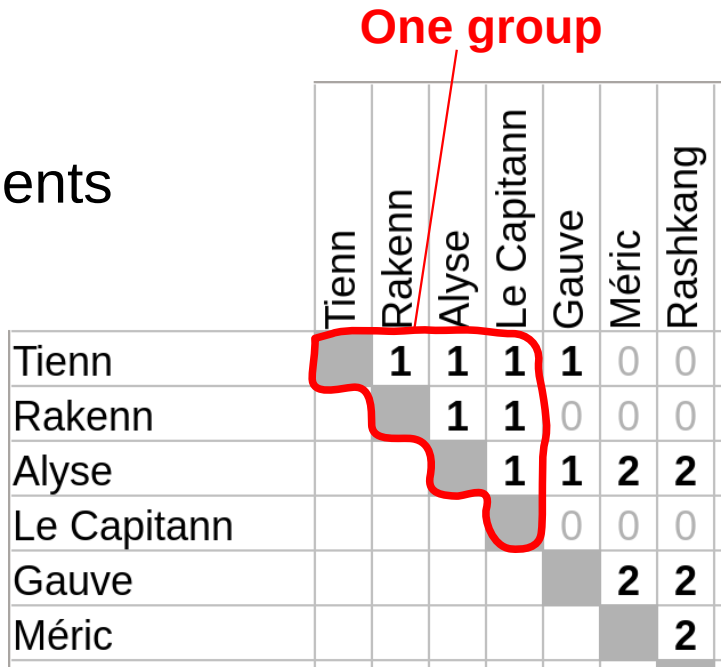
- **Example on amino acids**



General methods

- **A symmetric square matrix** $M = (M_{i,j})_{1 \leq i \leq n, 1 \leq j \leq n} \in \mathbb{R}, \text{ with } M_{i,j} = M_{j,i}$
- **Extracting overlapping sets**
  - One element for each row / column
  - One set for each group of interrelated elements
- **The selection function**
  - Returns True if two elements are related, depending on the value in the matrix

$$select : \mathbb{R} \rightarrow \{True, False\}$$



  - Compute S, the set of sets:

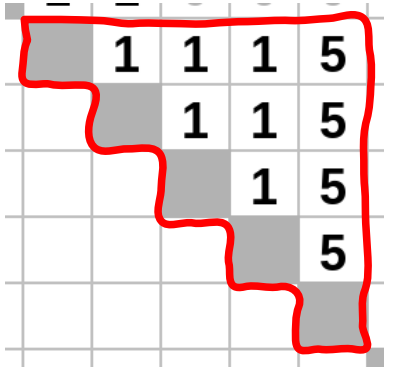
$$S_0 = \left\{ s \subseteq E \mid |s| > 1 \wedge \forall (i,j) \in (s,s) \, with \, i \neq j, select(M_{i,j}) = True \right\}$$

$$S = \{ s \in S_0 \mid \nexists s' \in S_0, s \subset s' \}$$

General methods

- **Each set corresponds to several values in the matrix**
- **The aggregation function returns a single value from these values:**

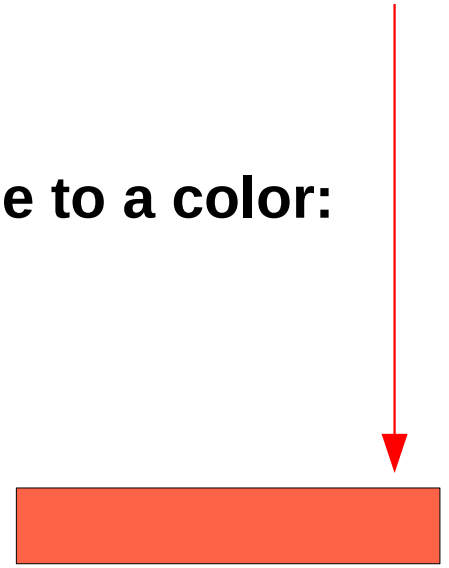
$$aggregate : \mathbb{R}^p \rightarrow \mathbb{R}, \text{ with } p \geq 1$$



- **The colorization function converts the resulting value to a color:**

$$colorize : \mathbb{R} \rightarrow color$$

- **Each set => One rectangular box in rainbow boxes**



- **Visualization parameters: the 3 functions**
  - select(), aggregate(), colorize()

Application to a small dataset
(21 characters)

- **Matrix produced manually by the authors**
- **Number in the matrix: part of the novel in which the two characters begin their relation**
  - 0: never
  - 1: before the novel begins
  - 2-5: parts 1-4
- **Definitions of the 3 functions:**

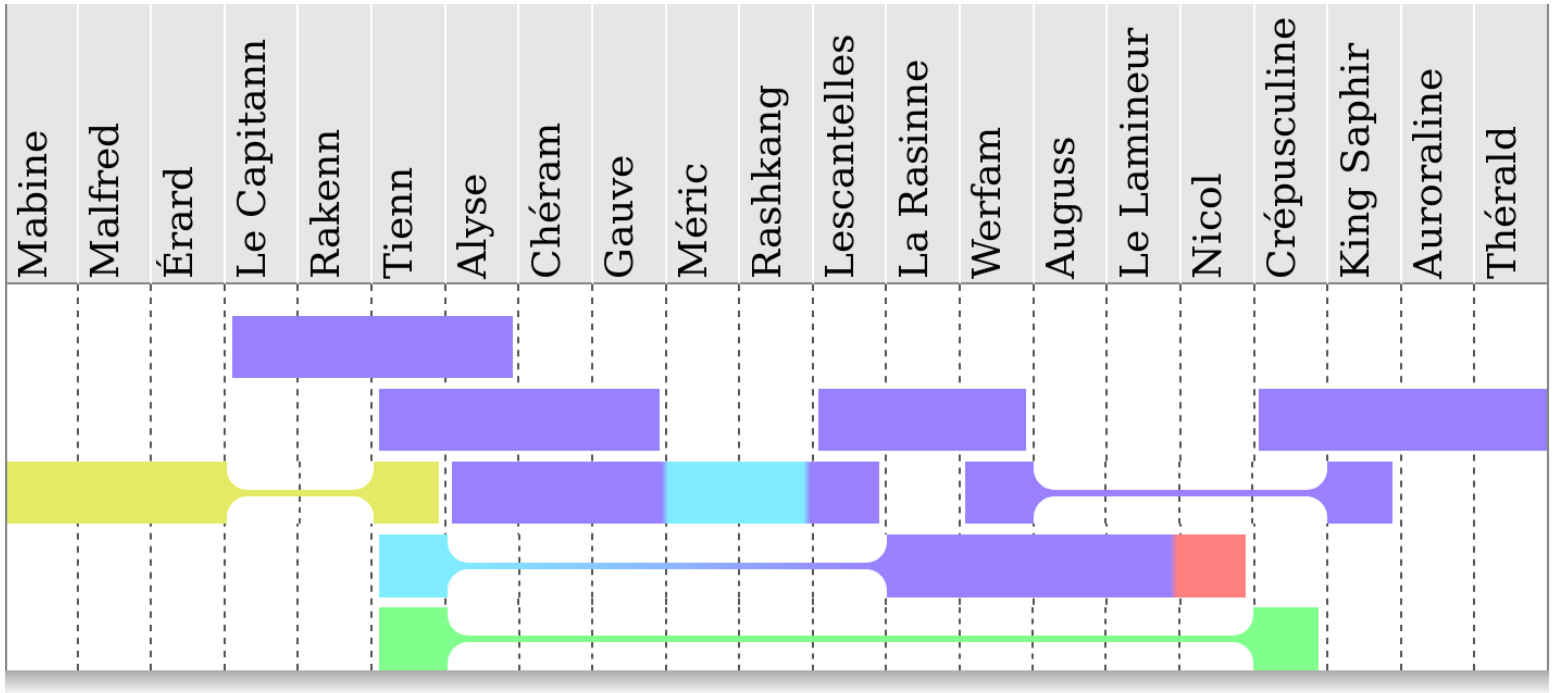
$$select : p \mapsto True\, if\, p \neq 0, False\, otherwise$$

$$aggregate : p_1, p_2, ... \mapsto \min(p_1, p_2, ...)$$

$$colorize : a \mapsto \begin{cases} violet & if\ a = 1 \\ cyan & if\ a = 2 \\ green & if\ a = 3 \\ yellow & if\ a = 4 \\ red & if\ a = 5 \end{cases}$$

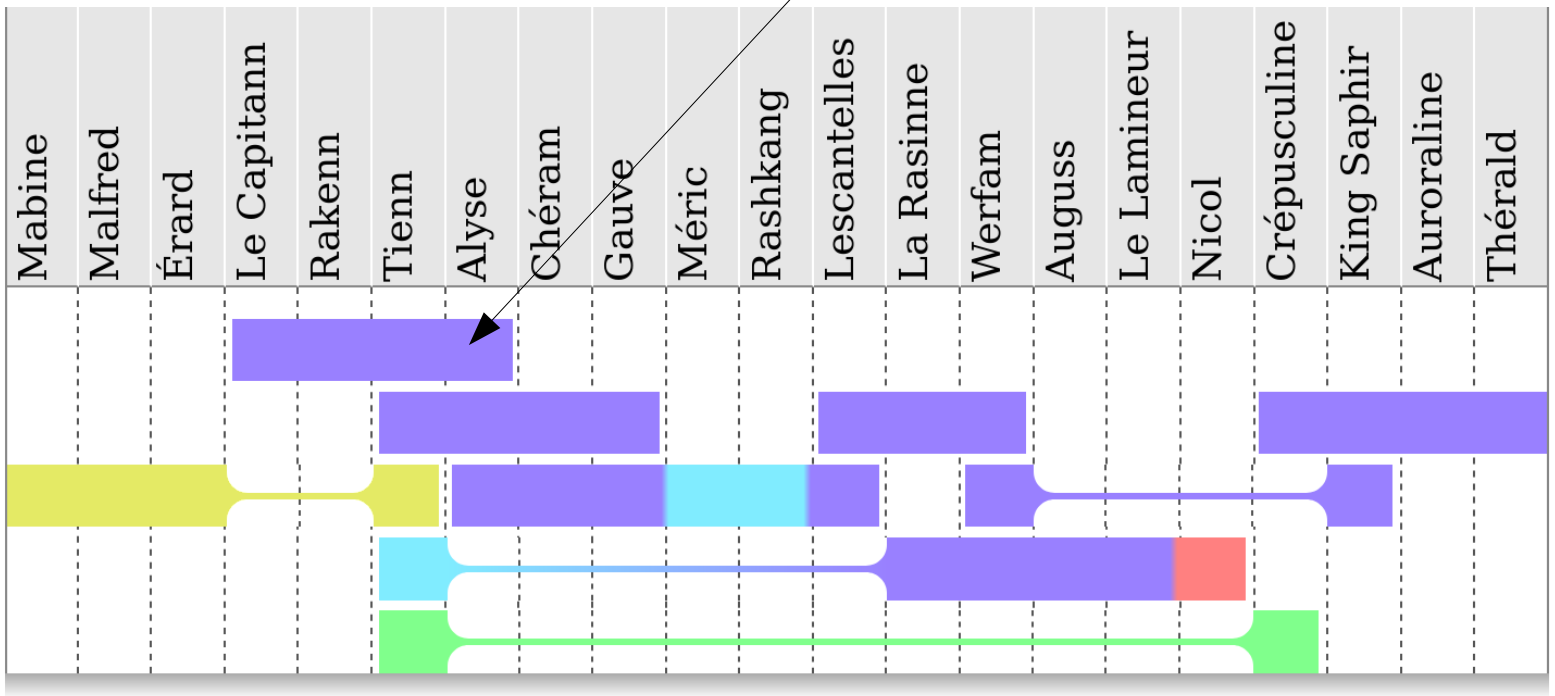
	Tienn	Rakenn	Alyse	Le Capitann	Gauve	Méric	Rashkang	Chéram	Lescantelles	La Rasinne	Werfam	Auguss	Le Lamineur	Nicol	Crépusculine	Auroraline	King Saphir	Thérald	Malfred	Mabine	Érard
Tienn		1	1	1	1	0	0	1	0	2	2	2	2	5	3	0	0	0	4	4	4
Rakenn			1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Alyse				1	1	2	2	2	2	0	0	0	0	0	0	0	0	0	0	0	0
Le Capitann					0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Gauve						2	2	2	2	0	0	0	0	0	0	0	0	0	0	0	0
Méric							2	2	2	0	0	0	0	0	0	0	0	0	0	0	0
Rashkang								2	2	0	0	0	0	0	0	0	0	0	0	0	0
Chéram									1	0	0	0	0	0	0	0	0	0	0	0	0
Lescantelles										1	1	0	0	0	0	0	0	0	0	0	0
La Rasinne											1	1	1	5	0	0	0	0	0	0	0
Werfam												1	1	5	0	0	1	0	0	0	0
Auguss													1	5	0	0	0	0	0	0	0
Le Lamineur														5	0	0	0	0	0	0	0
Nicol															0	0	0	0	0	0	0
Crépusculine																1	1	1	0	0	0
Auroraline																	1	1	0	0	0
King Saphir																		1	0	0	0
Thérald																			0	0	0
Malfred																				4	4
Mabine																					4
Érard																					

Application to a small dataset
(21 characters)



Application to a small dataset
(21 characters)

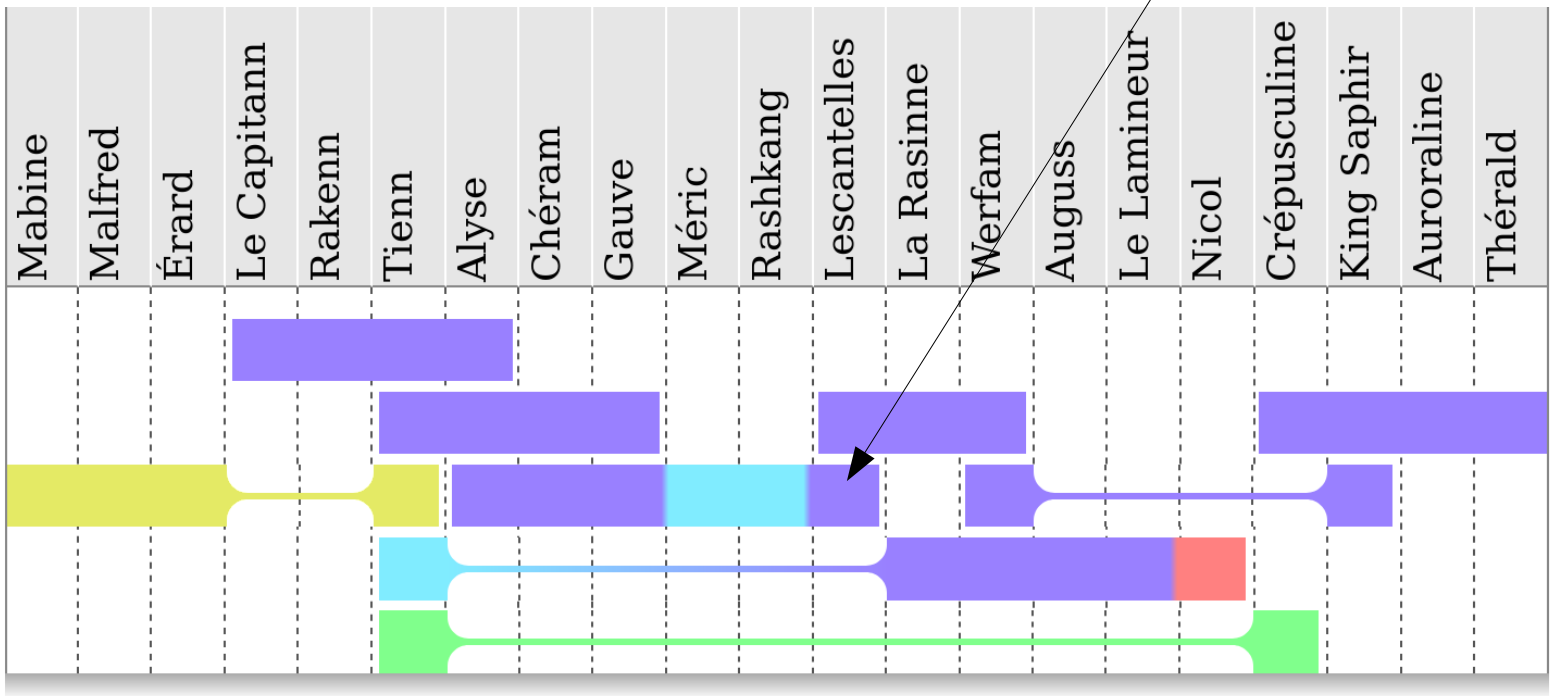
	Tienn	Rakenn	Alyse	Le Capitann	Gauve	Méric	Rashkang	Chéram	Lescantelles	La Rasinne	Werfam	Auguss	Le Lamineur	Nicol	Crépusculine	Auroraline	King Saphir	Thérald	Malfred	Mabine	Érard
Tienn		1	1	1	1	0	0	1	0	2	2	2	2	5	3	0	0	0	4	4	4
Rakenn			1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Alyse				1	1	2	2	2	2	0	0	0	0	0	0	0	0	0	0	0	0
Le Capitann					0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Gauve						2	2	2	2	0	0	0	0	0	0	0	0	0	0	0	0
Méric							2	2	2	0	0	0	0	0	0	0	0	0	0	0	0
Rashkang								2	2	0	0	0	0	0	0	0	0	0	0	0	0
Chéram									1	0	0	0	0	0	0	0	0	0	0	0	0
Lescantelles										1	1	0	0	0	0	0	0	0	0	0	0
La Rasinne											1	1	1	5	0	0	0	0	0	0	0
Werfam												1	1	5	0	0	1	0	0	0	0
Auguss													1	5	0	0	0	0	0	0	0
Le Lamineur														5	0	0	0	0	0	0	0
Nicol															0	0	0	0	0	0	0
Crépusculine																1	1	1	0	0	0
Auroraline																	1	1	0	0	0
King Saphir																		1	0	0	0
Thérald																			0	0	0
Malfred																				4	4
Mabine																					4
Érard																					



Before the story begins | Part #1 | Part #2 | Part #3 | Part #4

Application to a small dataset
(21 characters)

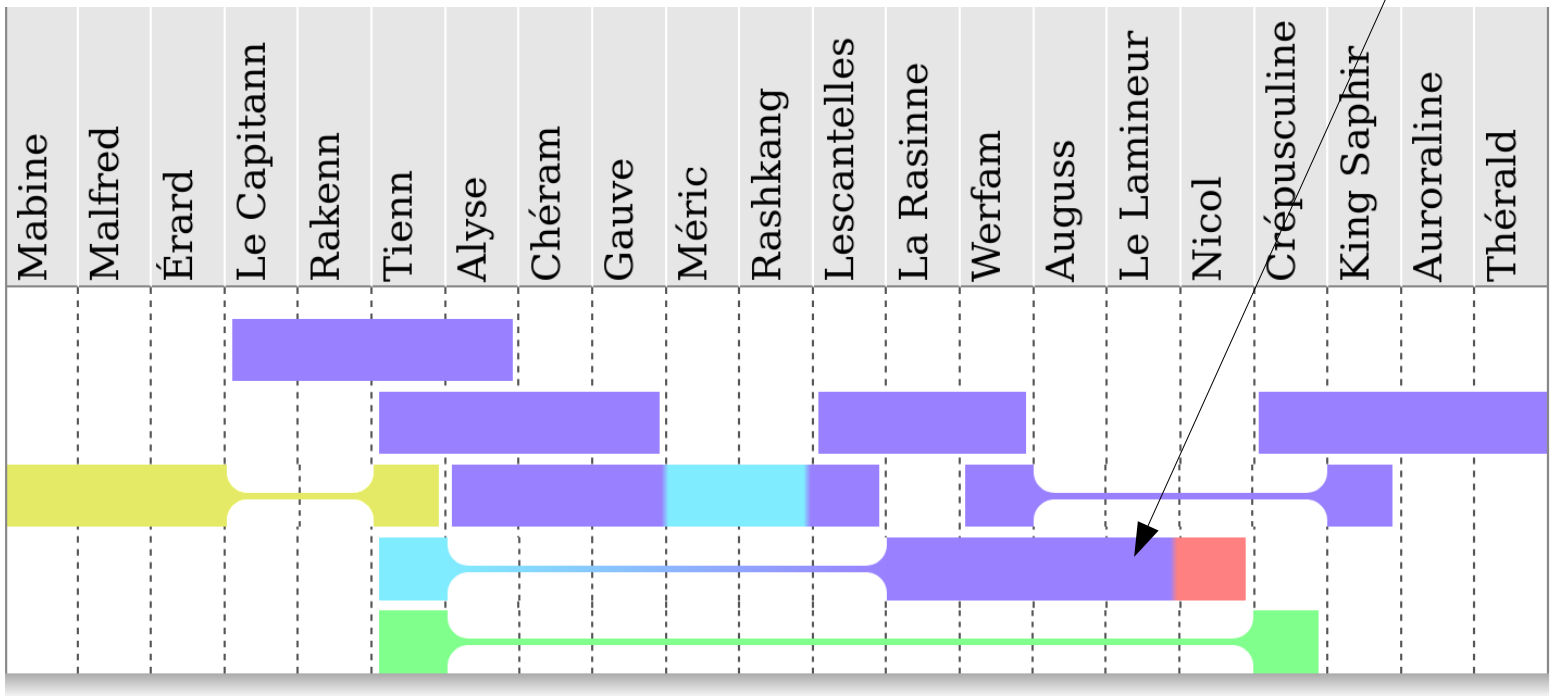
	Tienn	Rakenn	Alyse	Le Capitann	Gauve	Méric	Rashkang	Chéram	Lescantelles	La Rasinne	Werfam	Auguss	Le Lamineur	Nicol	Crépusculine	Auroraline	King Saphir	Thérald	Malfred	Mabine	Érard
Tienn		1	1	1	1	0	0	1	0	2	2	2	2	5	3	0	0	0	4	4	4
Rakenn			1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Alyse				1	1	2	2	2	2	0	0	0	0	0	0	0	0	0	0	0	0
Le Capitann					0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Gauve						2	2	2	2	0	0	0	0	0	0	0	0	0	0	0	0
Méric							2	2	2	0	0	0	0	0	0	0	0	0	0	0	0
Rashkang								2	2	0	0	0	0	0	0	0	0	0	0	0	0
Chéram									1	0	0	0	0	0	0	0	0	0	0	0	0
Lescantelles										1	1	0	0	0	0	0	0	0	0	0	0
La Rasinne											1	1	1	5	0	0	0	0	0	0	0
Werfam												1	1	5	0	0	1	0	0	0	0
Auguss													1	5	0	0	0	0	0	0	0
Le Lamineur														5	0	0	0	0	0	0	0
Nicol															0	0	0	0	0	0	0
Crépusculine																1	1	1	0	0	0
Auroraline																	1	1	0	0	0
King Saphir																		1	0	0	0
Thérald																			0	0	0
Malfred																				4	4
Mabine																					4
Érard																					



Before the story begins | Part #1 | Part #2 | Part #3 | Part #4

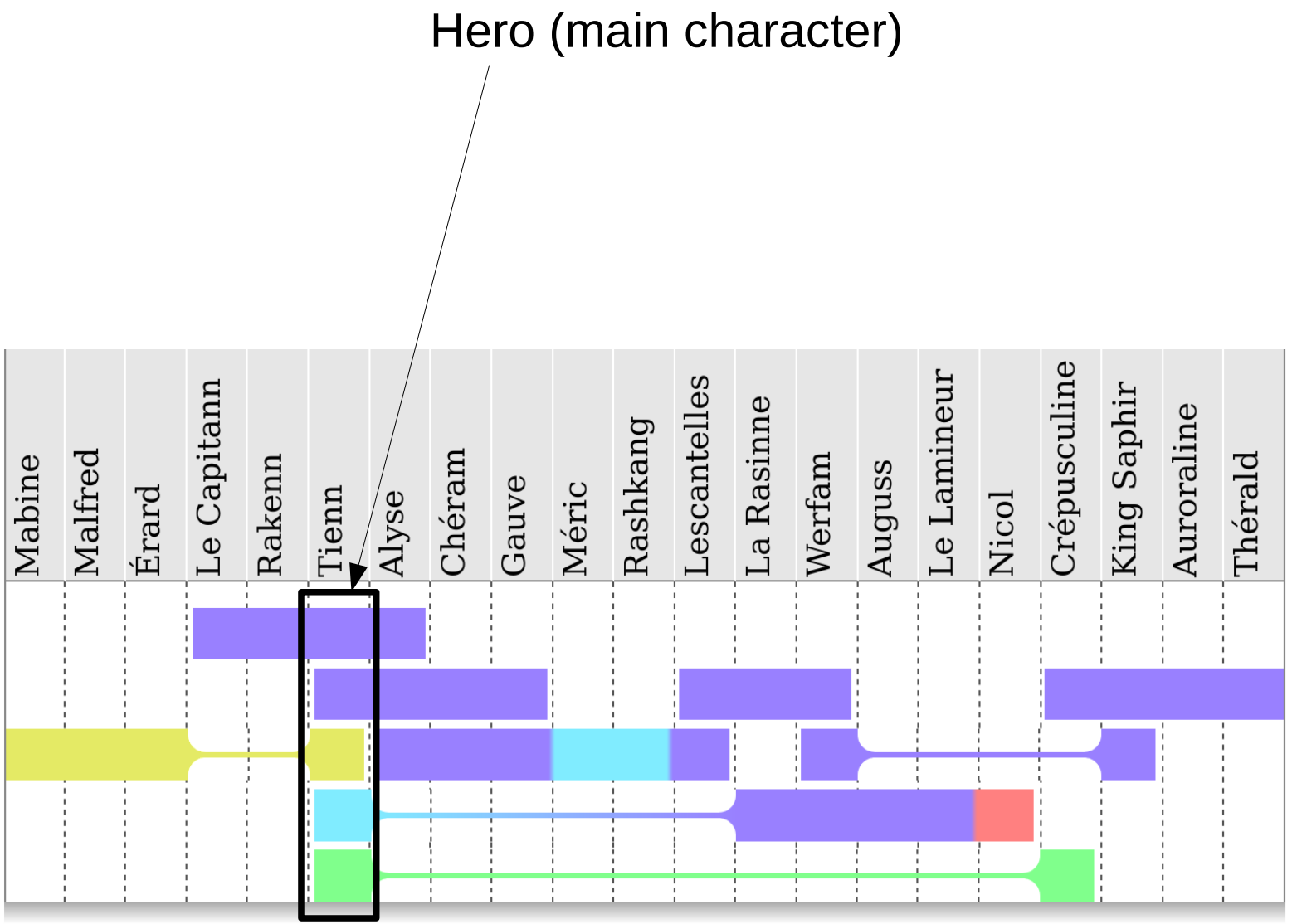
Application to a small dataset
(21 characters)

	Tienn	Rakenn	Alyse	Le Capitann	Gauve	Méric	Rashkang	Chéram	Lescantelles	La Rasinne	Werfam	Auguss	Le Lamineur	Nicol	Crépusculine	Auroraline	King Saphir	Thérald	Malfred	Mabine	Érard
Tienn		1	1	1	1	0	0	1	0	2	2	2	2	5	3	0	0	0	4	4	4
Rakenn			1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Alyse				1	1	2	2	2	2	0	0	0	0	0	0	0	0	0	0	0	0
Le Capitann					0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Gauve						2	2	2	2	0	0	0	0	0	0	0	0	0	0	0	0
Méric							2	2	2	0	0	0	0	0	0	0	0	0	0	0	0
Rashkang								2	2	0	0	0	0	0	0	0	0	0	0	0	0
Chéram									1	0	0	0	0	0	0	0	0	0	0	0	0
Lescantelles										1	1	0	0	0	0	0	0	0	0	0	0
La Rasinne											1	1	1	5	0	0	0	0	0	0	0
Werfam												1	1	5	0	0	1	0	0	0	0
Auguss													1	5	0	0	0	0	0	0	0
Le Lamineur														5	0	0	0	0	0	0	0
Nicol															0	0	0	0	0	0	0
Crépusculine																1	1	1	0	0	0
Auroraline																	1	1	0	0	0
King Saphir																		1	0	0	0
Thérald																			0	0	0
Malfred																				4	4
Mabine																					4
Érard																					

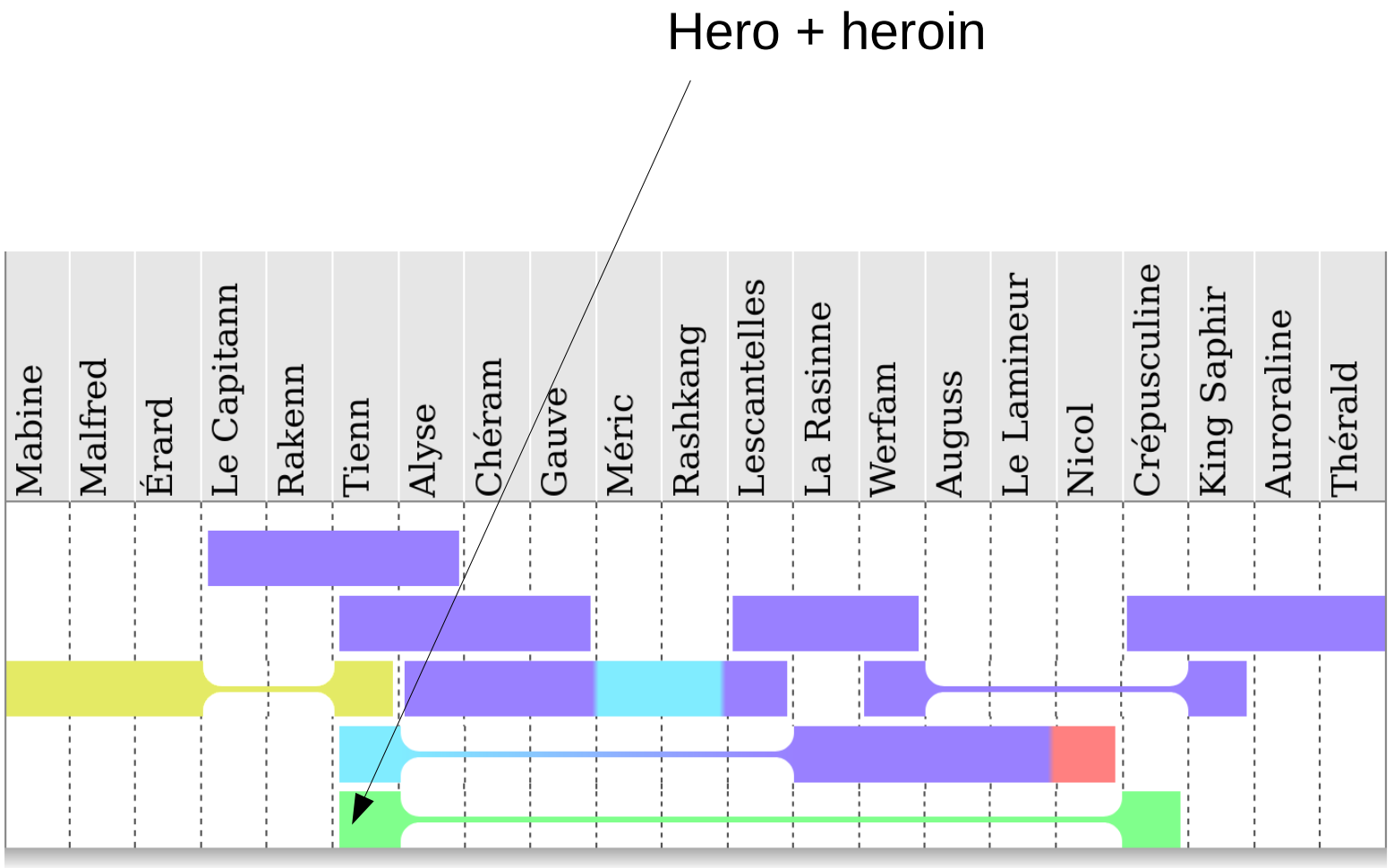


Before the story begins | Part #1 | Part #2 | Part #3 | Part #4

Application to a small dataset
(21 characters)

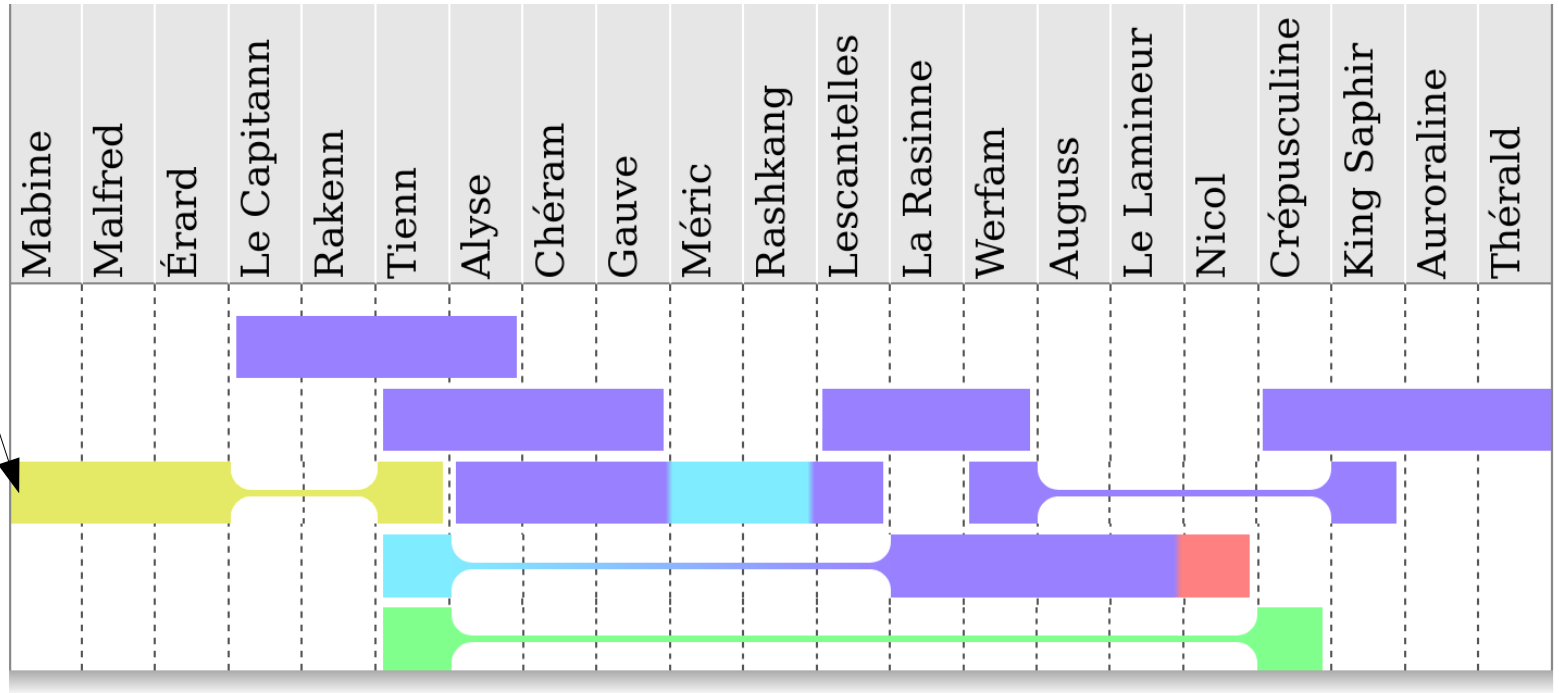


Application to a small dataset
(21 characters)



Application to a small dataset
(21 characters)

A group of isolated characters
(« ghetto »)

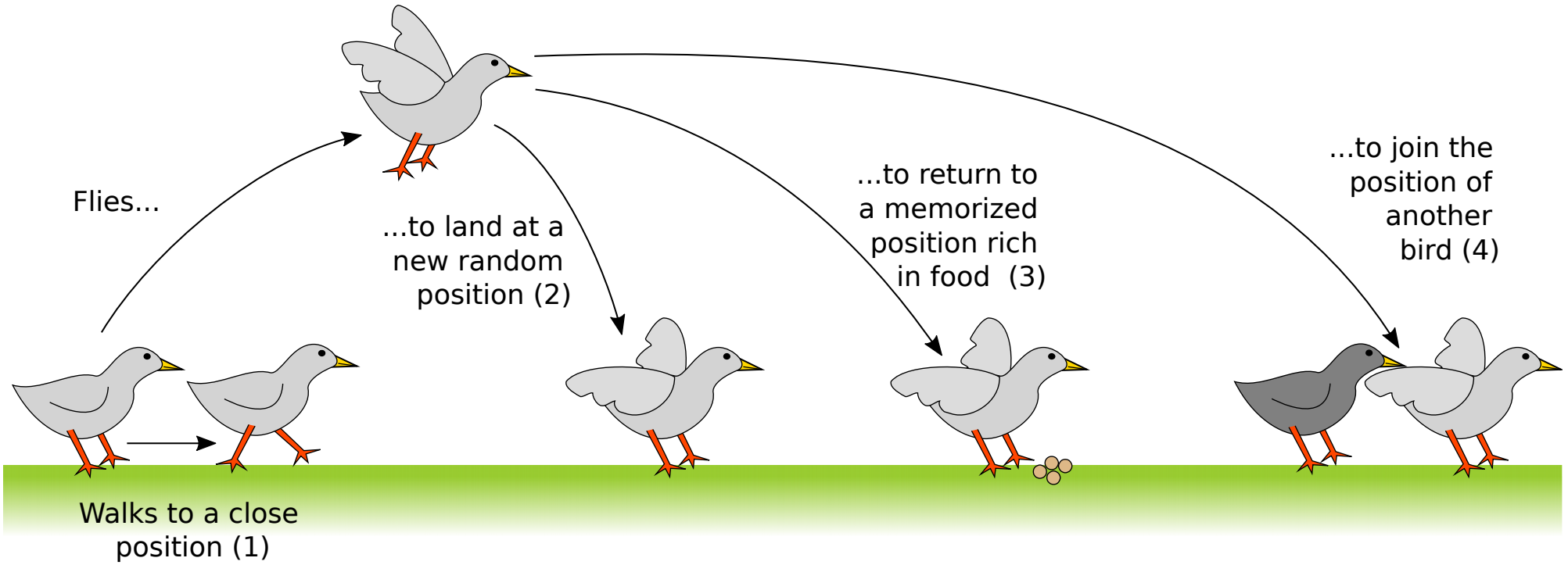


Application to a large dataset
(*Les Misérables*, 80 characters)

- **Problem:**
  - The optimization of the order of 80 columns
  - Factorial complexity
  - 80! ≈ 10e118
  - Previously published heuristic algorithm limited to 20-25 columns

AFB metaheuristic

Artificial Feeding Birds (AFB) [Springer]
Inspired by the behaviour of pigeons

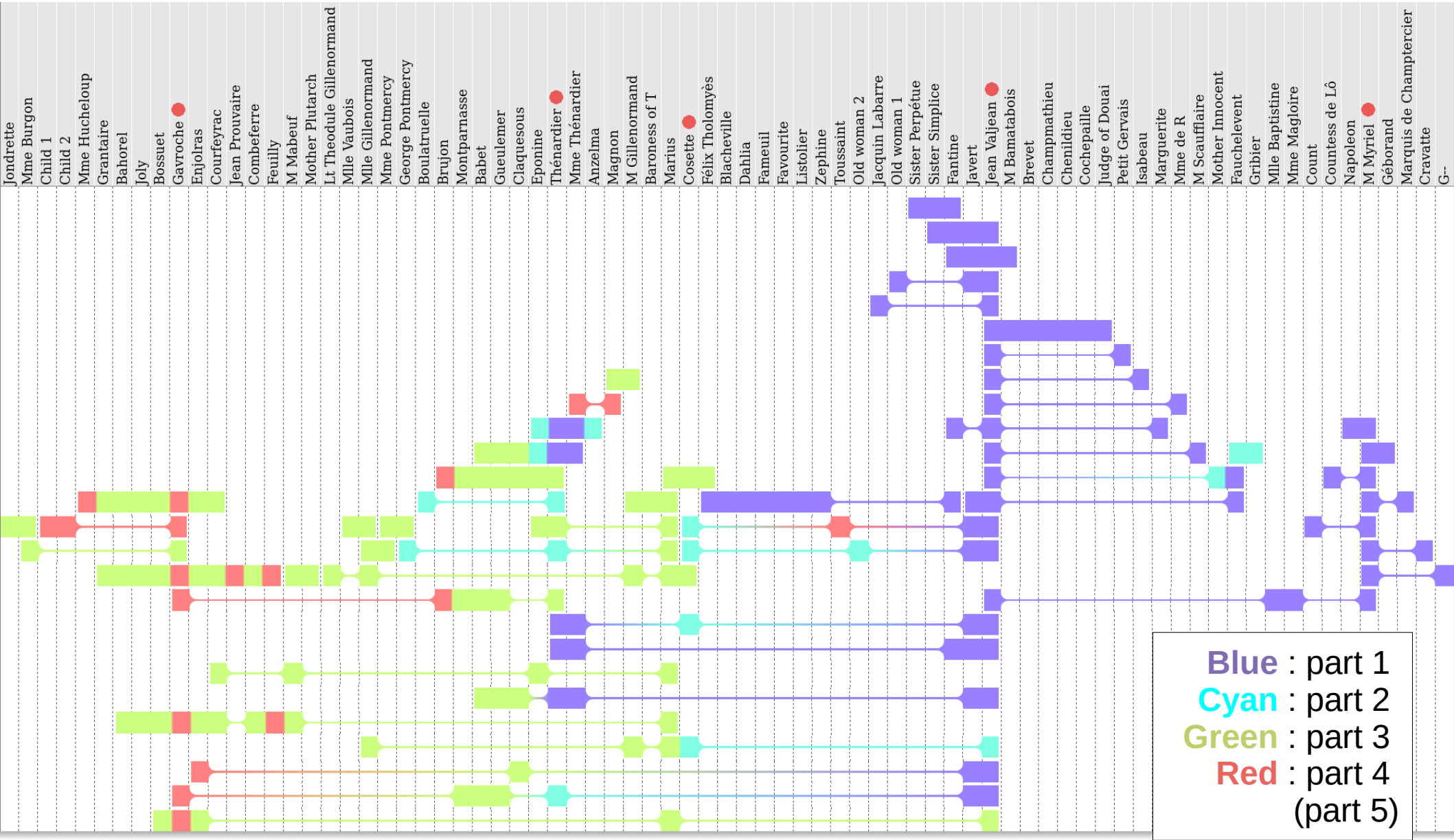


- **Simple**
- **Performant**
- **Generic**

An optimisation problem = A triplet (*coût*(), *vol*(), *marche*())

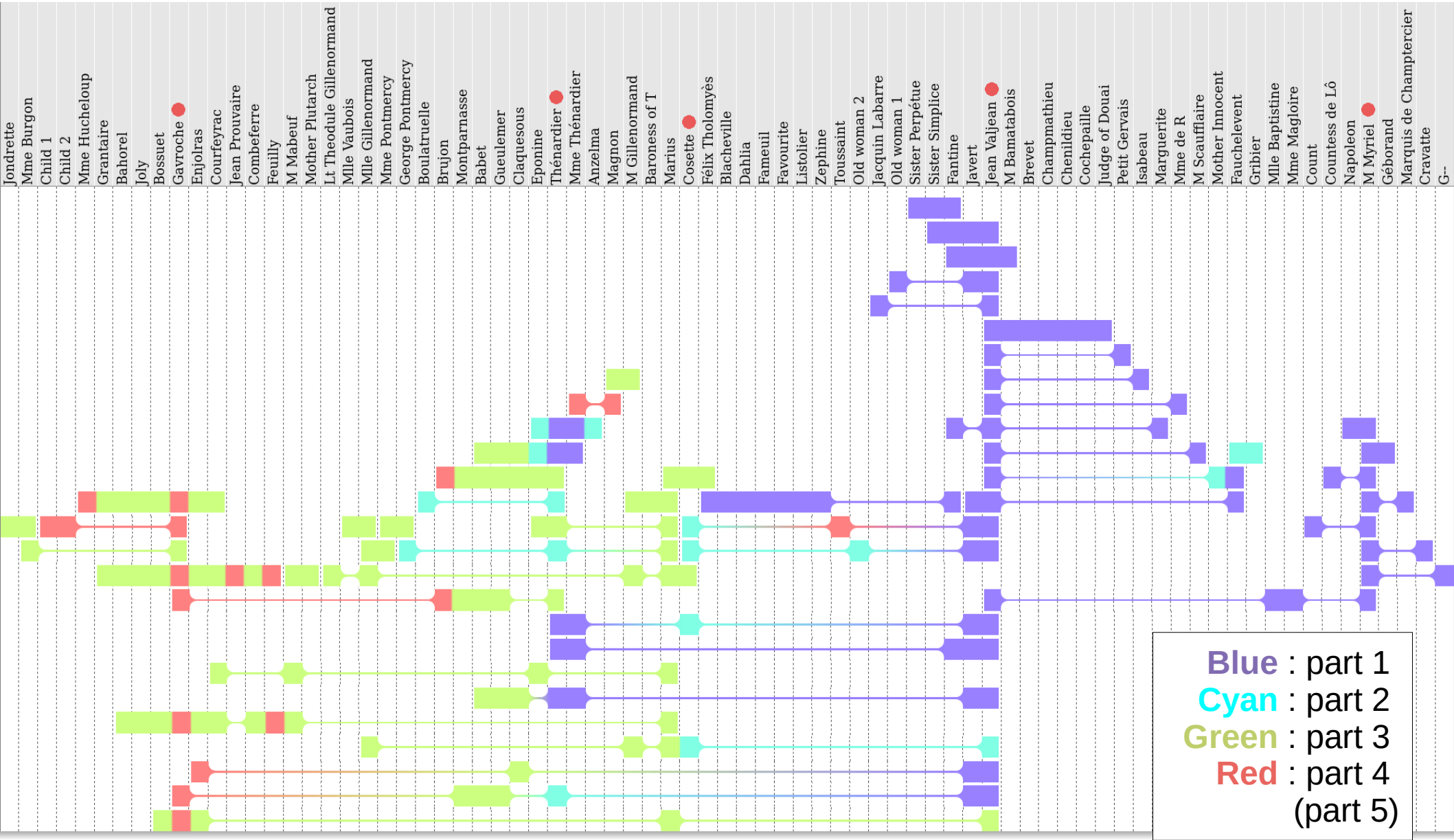
Application to a large dataset
(*Les Misérables*, 80 characters)

Same parameter functions as previously



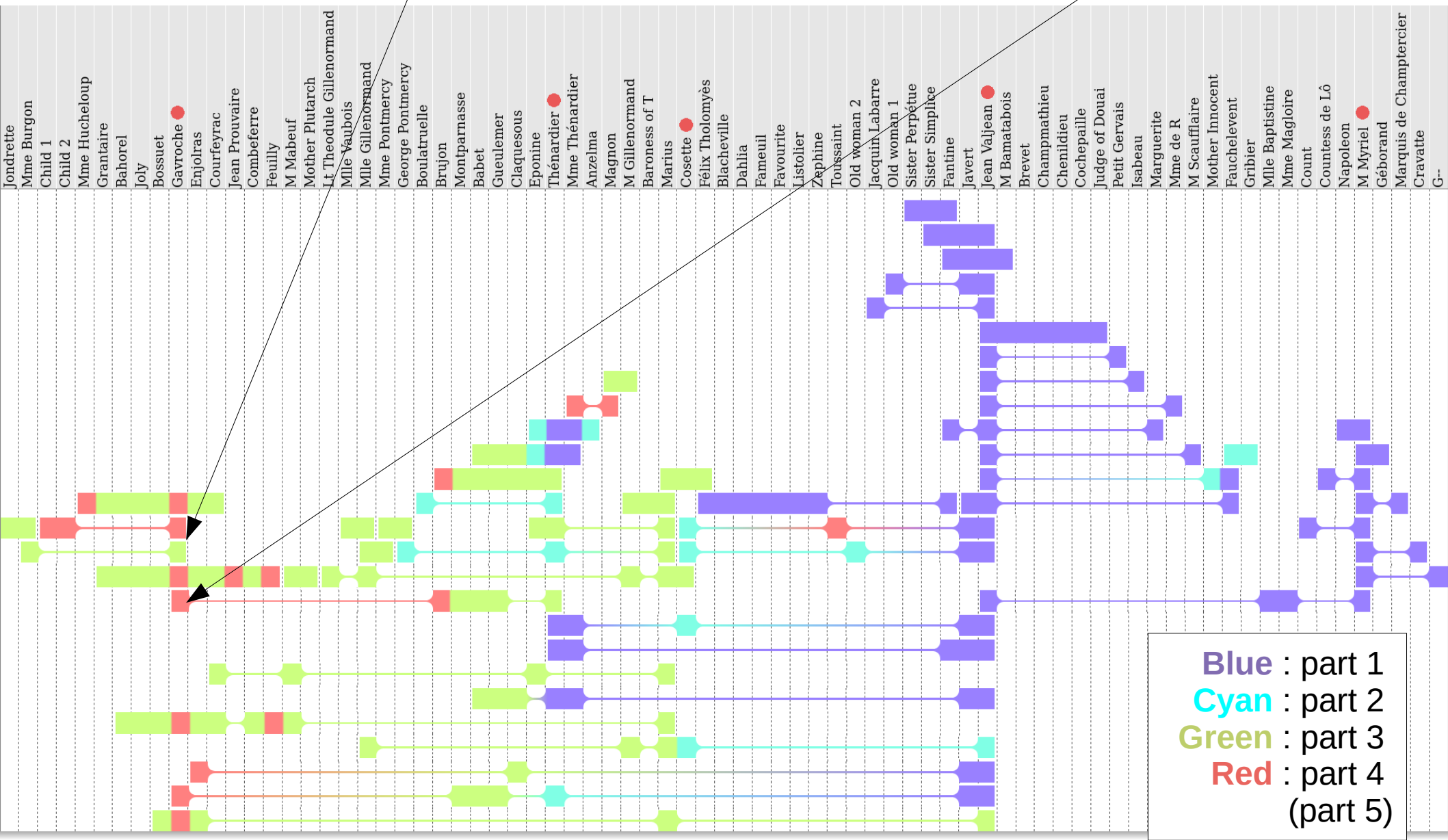
Application to a large dataset
(*Les Misérables*, 80 characters)

Most new encounters occur in parts 1 and 3, and seems well-separated



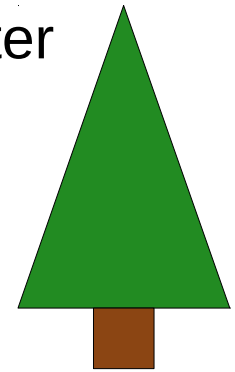
Application to a large dataset
(*Les Misérables*, 80 characters)

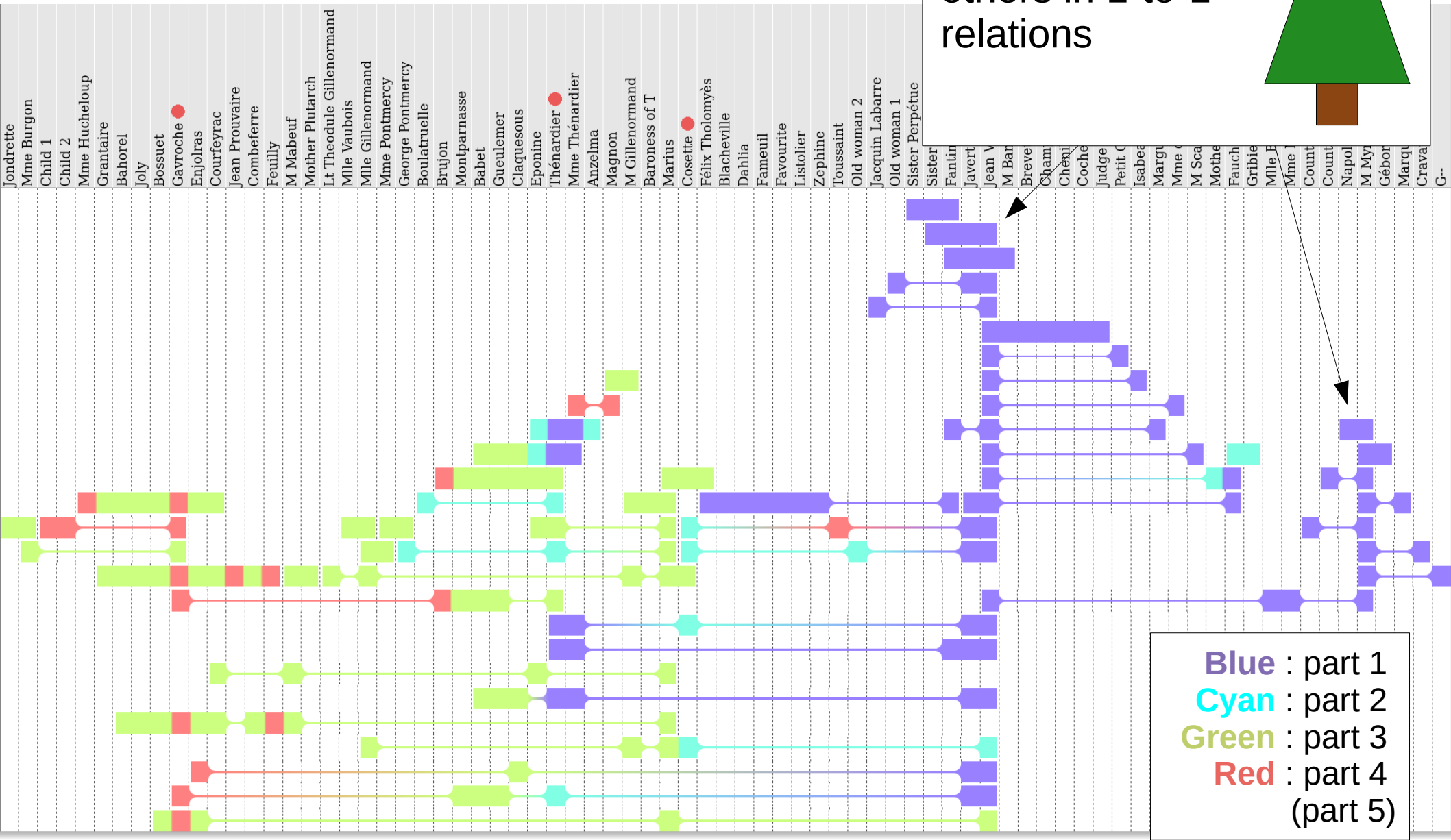
Gavroche appears in part 3 but encounters many more characters in part 4



Application to a large dataset
(*Les Misérables*, 80 cha

« Christmas tree » pattern: 1 character related to many others in 1-to-1 relations





Blue : part 1
Cyan : part 2
Green : part 3
Red : part 4
(part 5)

Application to a large dataset
(*Les Misérables*, 80 characters)

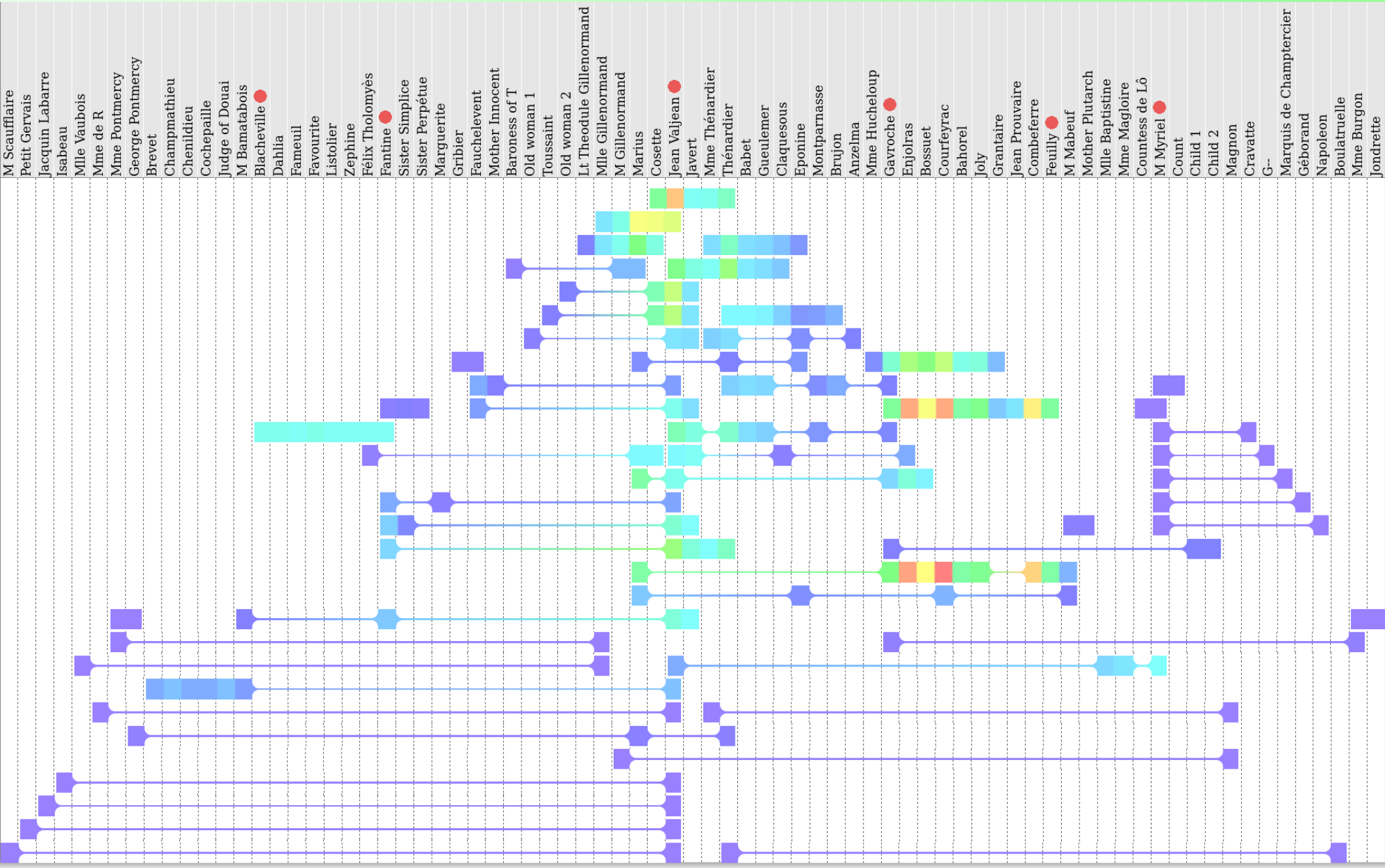
- **Another matrix**
  - Same novel and same characters
  - Numbers in the matrix are the number of times the characters meet
- **Different parameter functions**

$$select : p \mapsto True\, if\, p \neq 0, False\, otherwise$$

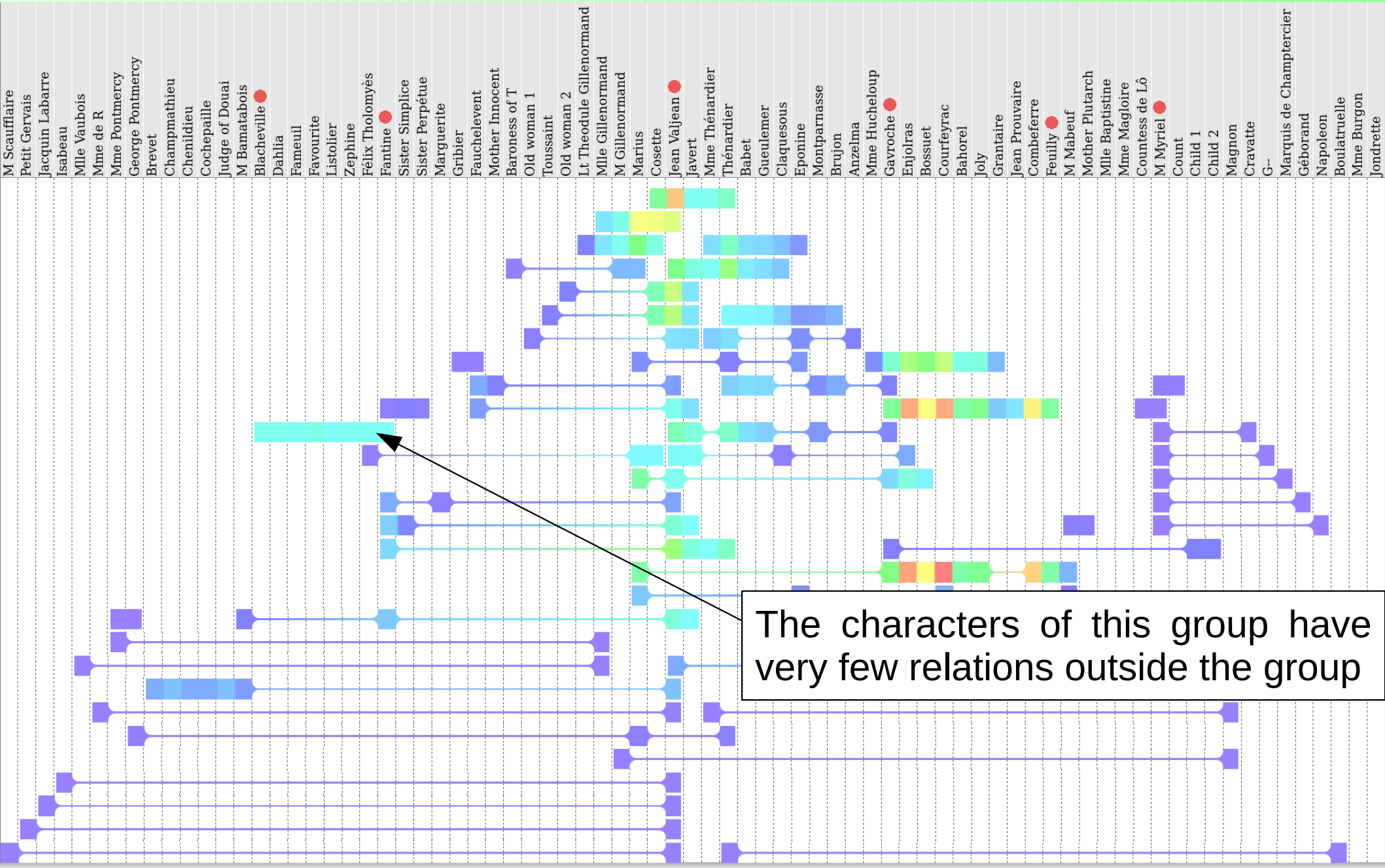
$$aggregate : c_1, c_2, ..., c_m \mapsto c_1 + c_2 + ... + c_m$$

$$colorize : c \mapsto \begin{cases} blue & if\ \frac{c}{c_{max}} = 0 \\ ... & \\ red & if\ \frac{c}{c_{max}} = 1 \end{cases}$$

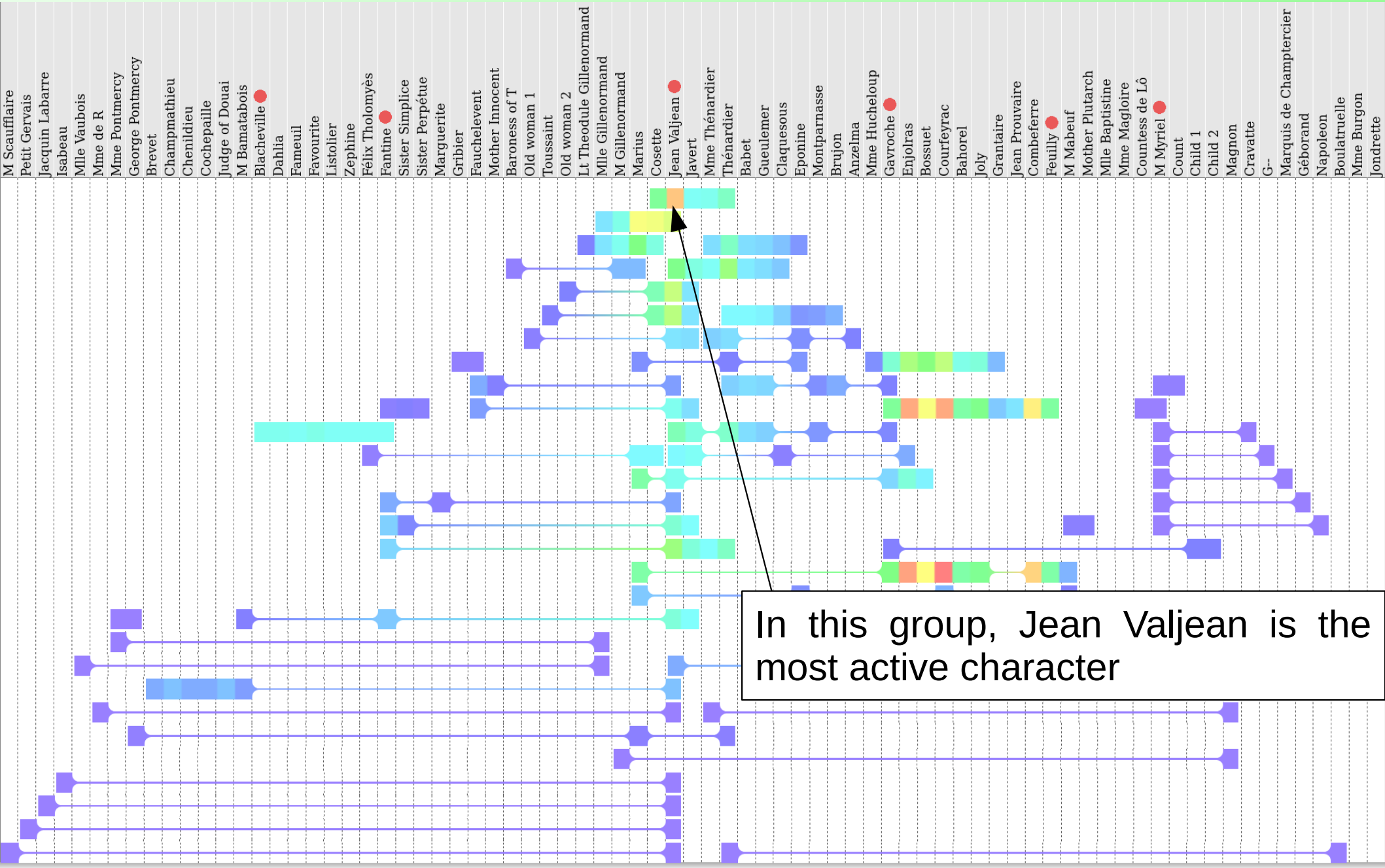
Application to a large dataset
(*Les Misérables*, 80 characters)



Application to a large dataset
(*Les Misérables*, 80 characters)



Application to a large dataset
(*Les Misérables*, 80 characters)



Discussion

- **Interesting for highly connected networks**
  - Rainbow boxes are more compact that matrices
- **We inferred groups of characters from pairwise relations**
  - A knows B, A knows C, B knows C
  - => A, B, C know each others
  - But is A aware that B knows C?
- **Perspectives**
  - Extend the method to non-symmetric matrices (directed graphs)
  - Adapt the method to other overlapping set visualization techniques
    - Venn diagrams,...
  - Apply the method to other domains
    - Social media (FOAF), bioinformatics (protein interaction matrices),...

References

[iV2016] : Lamy JB, Berthelot H, Favre M. Rainbow boxes: a technique for visualizing overlapping sets and an application to the comparison of drugs properties. International Conference Information Visualisation (iV) 2016;:253-260

[JVLC] : Lamy JB, Berthelot H, Capron C, Favre M. Rainbow boxes: a new technique for overlapping set visualization and two applications in the biomedical domain. Journal of Visual Language and Computing 2017;43:71-82

[Springer] : Lamy JB. Artificial Feeding Birds (AFB): a new metaheuristic inspired by the behavior of pigeons. Advances in nature-inspired computing and applications 2018;under press

[Leydesdorff] : Leydesdorff L and Vaughan L. Co-occurrence matrices and their applications in information science: Extending ACA to the Web environment. Journal of the association for information science and technology 2006;57(12):1616-1628

[Jänicke] : Jänicke S, Franzini G, Cheema MF, and Scheuermann G. Visual text analysis in digital humanities. 2016;226-250