

# Refining Character Relationships in a Textual Narrative using Embeddings of Interactions

---

Guillaume Gux

University of Lausanne

# Table of contents

---

1. Introduction

2. Correspondence Analysis (CA)

# Introduction

---

# The construction of a character network

The construction of a character network takes generally 3 steps [Labatut and Bost, 2019]:

- Identification of **characters**
- Detection of **interactions**
- Construction of the **graph**

# The construction of a character network

When dealing with **textual narrative**, a frequently used method is to count **character co-occurrences** in **textual units** (e.g., [Elsner, 2012, Rochat and Kaplan, 2014]).

⋮

... cet affaiblissement de la maladie qui ressemble à l'enfance, afin que, la voyant si paisible, on ne fit pas difficulté de lui amener Cosette. Cependant, tout en se contenant, elle ne pouvait s'empêcher d'adresser à M. Madeleine mille questions

⋮

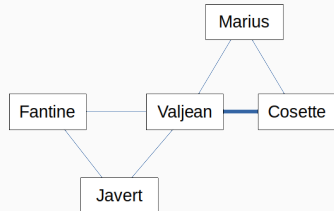
Lorsque Jean Valjean, dans la nuit même du jour où Javert l'arrêta près du lit de mort de Fantine, s'échappa de la prison municipale de Montreuil-sur-Mer,...

⋮

Cosette, un peu moins rêveuse que Marius, était gaie, et cela suffisait à Jean Valjean pour être heureux. Les pensées que Cosette avait, ses préoccupations tendres, l'image de Marius qui lui remplissait l'âme,...

⋮

Character 1	Character 2	Co-occurrences
Cosette	Marius	1
Cosette	Valjean	2
Fantine	Javert	1
Fantine	Valjean	1
Javert	Valjean	1
Marius	Valjean	1



# Dataset

In this context, the **dataset** used to construct the character network has the following form

tome	chapitre	livre	text	Azelma	Babet	Bahorel	Barnabois	Baptistine	Basque	...	Montparnasse	Myriel	Nicolette	Pontmercy	Prouvaire	Simplice	Tholomyès	Toussaint	Valjean	Zéphine
1	1	1	Monsieur Myriel	0	0	0	0	0	0	...	0	1	0	0	0	0	0	0	0	0
1	1	1	En 1815, M. Charles-François-Bienvenu Myriel était évêque de Digne.	0	0	0	0	0	0	...	0	1	0	0	0	0	0	0	0	0
1	1	1	Quelque ce détail ne touche en aucune manière au fond même de ce que	0	0	0	0	0	0	...	0	4	0	0	0	0	0	0	0	0
1	1	1	En 1804, M. Myriel était curé de Brignoles. Il était déjà vieux, et	0	0	0	0	0	0	...	0	1	0	0	0	0	0	0	0	0
1	1	1	Vers l'époque du couronnement, une petite affaire de sa cure, on ne sait	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
1	1	1	—Quel est ce bonhomme qui me regarde?	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
1	1	1	—Sire, dit M. Myriel, vous regardez un bonhomme, et moi je regarde un	0	0	0	0	0	0	...	0	1	0	0	0	0	0	0	0	0
1	1	1	L'empereur, le soir même, demanda au cardinal le nom de ce curé, et	0	0	0	0	0	0	...	0	1	0	0	0	0	0	0	0	0
1	1	1	Qu'il avait-il de vrai, du reste, dans les récits qu'on faisait sur la	0	0	0	0	0	0	...	0	2	0	0	0	0	0	0	0	0
1	1	1	M. Myriel devait subir le sort de tout nouveau venu dans une petite	0	0	0	0	0	0	...	0	1	0	0	0	0	0	0	0	0
1	1	1	Quoi qu'il en fût, après neuf ans d'épiscopat et de résidence à Digne,	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
1	1	1	M. Myriel était arrivé à Digne accompagné d'une vieille fille,	0	0	0	0	1	0	...	0	1	0	0	0	0	0	0	0	0
1	1	1	Ils avaient pour tout domestique une servante du même âge que	0	0	0	0	1	0	...	0	0	0	0	0	0	0	0	0	0
1	1	1	Mademoiselle Baptistine était une personne longue, pâle, mince, douce,	0	0	0	0	1	0	...	0	0	0	0	0	0	0	0	0	0
1	1	1	Madame Magloire était une petite vieille, blanche, grasse, replète,	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
1	1	1	À son arrivée, on installa M. Myriel en son palais épiscopal avec les	0	0	0	0	0	0	...	0	1	0	0	0	0	0	0	0	0

However, apart from a few exceptions [Nalisnick and Baird, 2013, Trovati and Brady, 2014, Min and Park, 2019], the content of the **text column** is not used. **Edges in the resulting graph aggregate blindly various kind of interactions.**

# Approaches

In this presentation, we propose to **refine character relationships** by using the **textual data**. Quantities of approaches can be undertaken, but we focus on:

- **Bag-of-paths** approaches, a corpus is represented by an **unit-term matrix**.

	abaisser	abandonner	abbé	abominable	aboutir	absence	absolu	absolument	absorber	abîme	accablement	accabler	accent
1	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	0	2	0	0	0	1	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	1	1	1	0	2	0
5	0	0	0	0	0	0	0	0	0	0	0	0	0
6	0	0	4	0	0	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	1	0	0	0	0	0
8	0	0	0	0	0	0	0	0	0	0	0	0	0
9	0	1	0	0	0	0	0	0	0	0	0	0	0
10	0	0	0	0	0	0	0	0	1	0	0	0	3

- **Embeddings of textual units**. The unit-term matrix is used to construct **vectors** representing units.
- **Embeddings of character relationships**, which derives from the embedding of textual unit.

More specifically, three approaches are made:

- **Correspondence Analysis (CA).**
- **Pre-trained Word Embeddings (WE).**
- “Topic” vectors from **Non-negative Matrix Factorization (NMF).**



# Correspondence Analysis (CA)

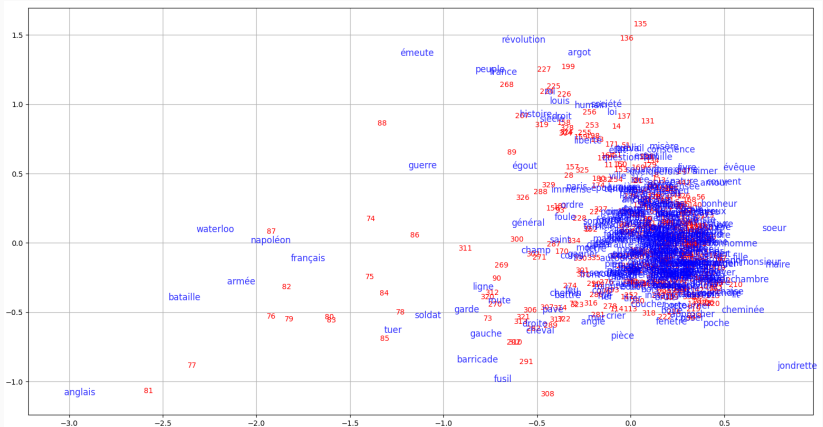
---

Correspondence Analysis is the natural tool for analyzing textual resources if data are organized in a  $(n \times p)$  **document-term matrix** [Lebart et al., 2019] (here, documents are our textual units). It gives:





- $\min(n, p) - 1$  **factorial axes**, by decreasing order of importance, which can be interpreted as latent variables.
- **Coordinates of each document** along these factors, where proximity can be interpreted as similar profile in term of words.
- **Coordinates of each word** along the same factors, where proximity can be interpreted as similar profile in term of documents.
- **Affinities between a document and a term**, which is computed by the **scalar product** between their vectors.

# Principles

If we plot units and words on the first two axes, we get the usual **biplot**:



# References

-  Elsner, M. (2012).  
**Character-based kernels for novelistic plot structure.**  
In *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics*, pages 634–644, Avignon, France. Association for Computational Linguistics.
-  Labatut, V. and Bost, X. (2019).  
**Extraction and Analysis of Fictional Character Networks: A Survey.**  
*ACM Computing Surveys*, 52(5):89:1–89:40.
-  Lebart, L., Pincemin, B., and Poudat, C. (2019).  
**Analyse des données textuelles.**  
Number 11 in *Mesure et évaluation*. Presses de l'Université du Québec, Québec.
-  Min, S. and Park, J. (2019).  
**Modeling narrative structure and dynamics with networks, sentiment analysis, and topic modeling.**  
*PLOS ONE*, 14(12):e0226025.
-  Nalisnick, E. T. and Baird, H. S. (2013).  
**Character-to-Character Sentiment Analysis in Shakespeare's Plays.**  
In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 479–483, Sofia, Bulgaria. Association for Computational Linguistics.
-  Rochat, Y. and Kaplan, F. (2014).  
**Analyse des réseaux de personnages dans Les Confessions de Jean-Jacques Rousseau.**  
*Les Cahiers du numérique*, 10(3):109–133.  
Place: Cachan Publisher: Lavoisier.
-  Trovati, M. and Brady, J. (2014).  
**Towards an Automated Approach to Extract and Compare Fictional Networks: An Initial Evaluation.**  
In *2014 25th International Workshop on Database and Expert Systems Applications*, pages 246–250, Munich, Germany. IEEE.