

Construction de graphes avec Gephi pour *Crimes et Châtiment* de Fiodor Dostoïevski

Présenté par : **Alexandra DEMKOVIC**

Enseignante de M4SOL203 - Cognition, sentiment, opinion :
Mme Victoria EYHARABIDE

Présenté le 26/02/2024

INTRODUCTION

Utilisation de Gephi :

- outil d'analyse et de visualisation de réseaux par des graphes personnalisables ;
- Créé en 2008 ;
- développé par Mathieu Bastian, Sébastien Heymann et Mathieu Jacomy à l'Université de Technologie de Compiègne en France.

Objectif :

- Créer un ou plusieurs graphes pour un roman donné ;
- Graphe(s) représentants les personnages et le nombre de fois qu'ils apparaissent ensemble.

PLAN

I

Présentation
du livre

II

Traitements
des données

III

Représentation
graphique avec
Gephi

IV

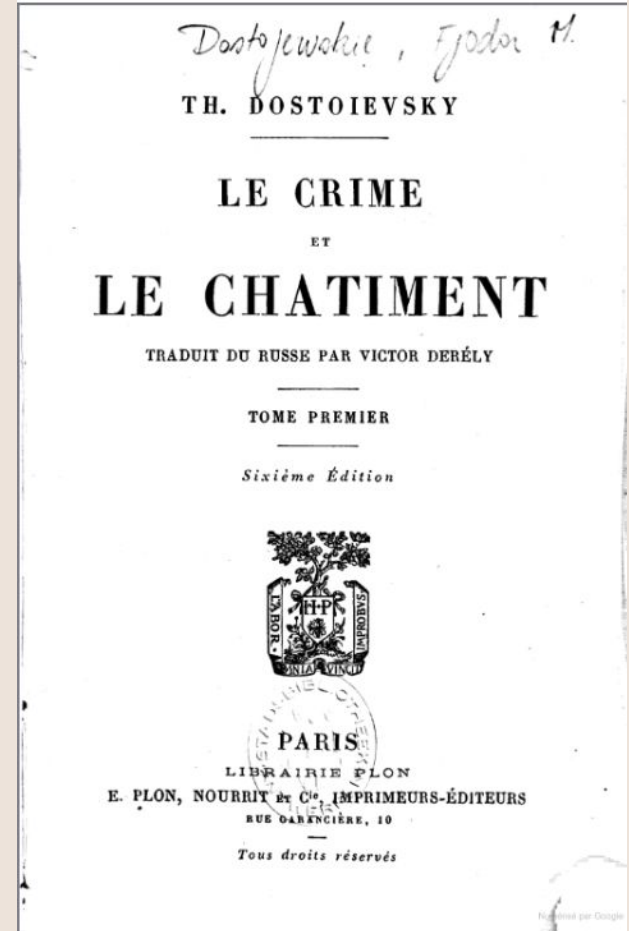
Analyses
graphiques des
données

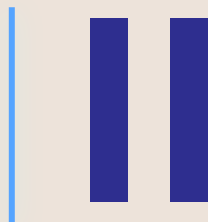


Présentation du livre

I. Présentation du livre

- ❖ **Titre** : *Crime et Châtiment*
- ❖ **Auteur** : Fiodor Mikhaïlovitch Dostoïevski
- ❖ **Date** : 1866-1867
- ❖ **Édition de référence** :
 - Le Livre de poche, Gallimard, 1950
 - En 2 volumes : 6 parties
- ❖ **Nb de pages** :
 - *Tome 1* : environ 660 pages
(*Dostoievski-crime-1.pdf*)
 - *Tome 2* : environ 780 pages
(*Dostoievski-crime-2.pdf*)
- ❖ Histoire qui comporte des **dialogues** et de **longues tirades** : intrigue psychologique et morale qui comportent beaucoup de réflexions des personnages





Traitements des données

II. Traitements des données 1/2

- **Traitement des données avec un code python :**
 - **Lecture** des PDFs (bibliothèque *PyPDF2*)
 - **Recueil** des personnages : à l'aide d'une liste déjà établie (*recherches internet*) : 29 personnages au total ;
- **Tri des relations en clé - valeur** : un nom de personnage et toutes ses appellations possibles
- **Association des relations** selon le nb d'apparitions des personnages ensemble dans un paragraphe donné
- **Création de 3 fichiers excel (bibliothèques *openpyxl* et *pandas*) :**
 - Identification des personnages : **id - label**
 - Un fichier excel pour chaque tome : ajout des poids (**weight**) pour chaque relation possible entre les personnages
(3 fichiers excel au total : IDpersonnages.xlsx, relationstome1.xlsx, relationstome2.xlsx)

II. Traitements des données 2/2

- **Association des relations** selon le nb d'apparitions des personnages ensemble dans une page donnée :
 - pas pertinent d'extraire par paragraphe :
 - Grande taille de la police d'écriture, pas pertinent d'extraire par paragraphe,
 - bcp de pages vides de personnages ;
 - Pour chaque personnage apparu ensemble dans une page donnée, ajouter un point (librairie itertools).
- **Nb de relations obtenu :**
 - *Tome 1* : 28 paires, 249 relations
 - *Tome 2* : 21 paires, 129 relations

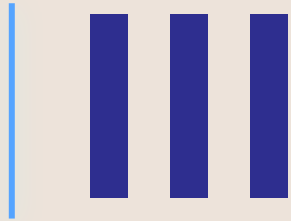
*Image à droite : extrait du tome 1 Partie I, page 6
(Dostoievski-crime-1.pdf)*

I

Par une soirée extrêmement chaude du début de juillet, un jeune homme sortit de la toute petite chambre qu'il louait dans la ruelle S... et se dirigea d'un pas indécis et lent, vers le pont K...

Il eut la chance de ne pas rencontrer sa propriétaire dans l'escalier.

Sa mansarde se trouvait sous le toit d'une grande maison à cinq étages et ressemblait plutôt à un placard qu'à une pièce. Quant à la logeuse qui lui louait la chambre avec le service et la pension, elle occupait un appartement à l'étage au-dessous, et le jeune homme, lorsqu'il sortait, était obligé, de passer devant la porte de sa cuisine, la plupart du temps grande ouverte sur l'escalier. À chaque fois, il en éprouvait une sensation malade de vague effroi, qui l'humiliait, et son visage se renfrognait. Il était terriblement endetté auprès de sa logeuse et il



Représentation graphique avec Gephi

III. Représentation graphique avec Gephi

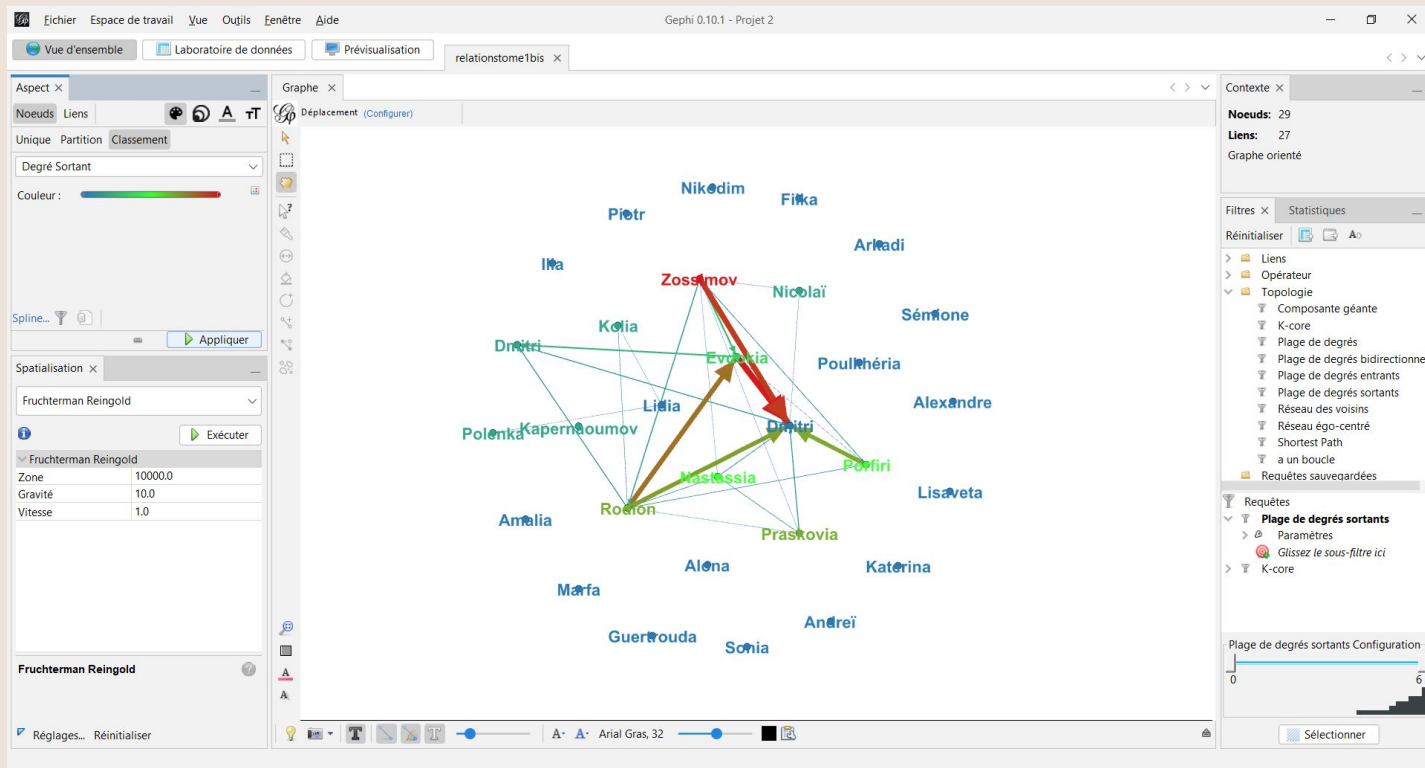


Image du graphe généré pour le tome 1

III. Représentation graphique avec Gephi

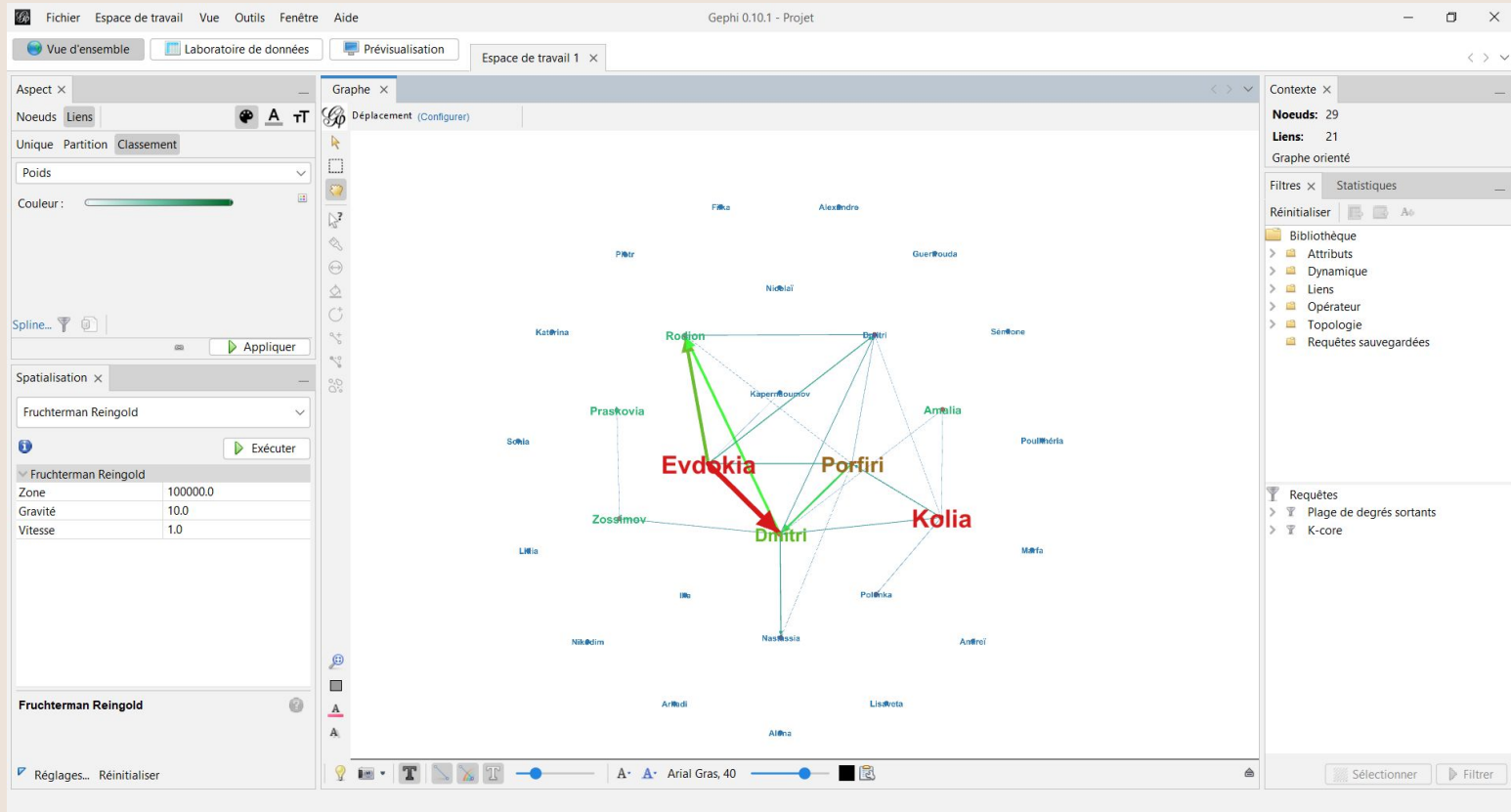


Image du graphe généré pour le tome 2

III. Représentation graphique avec Gephi 1/3

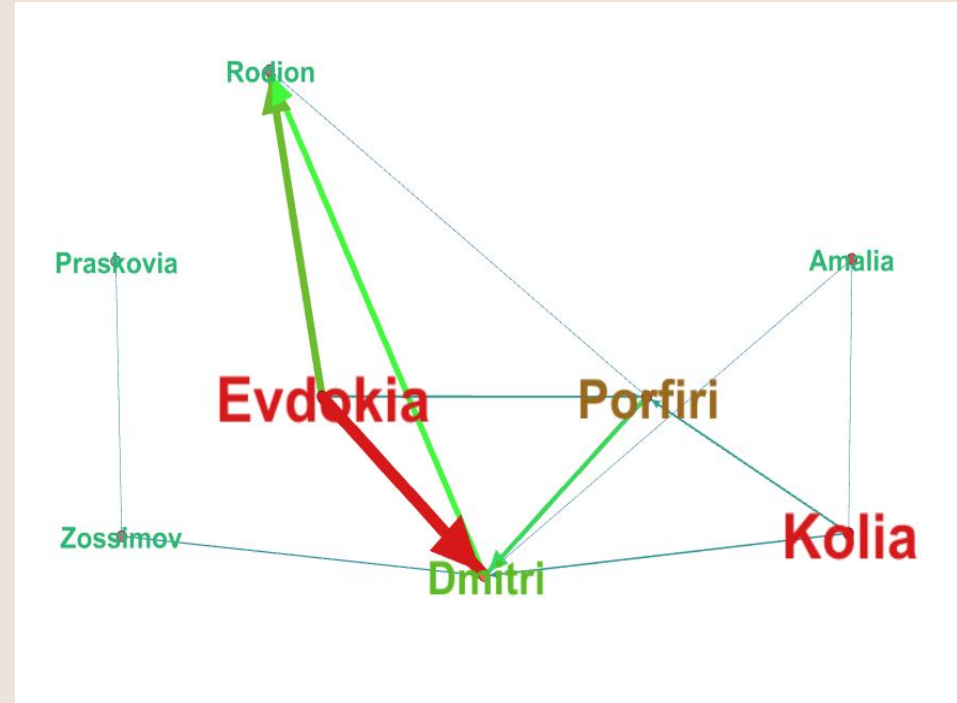
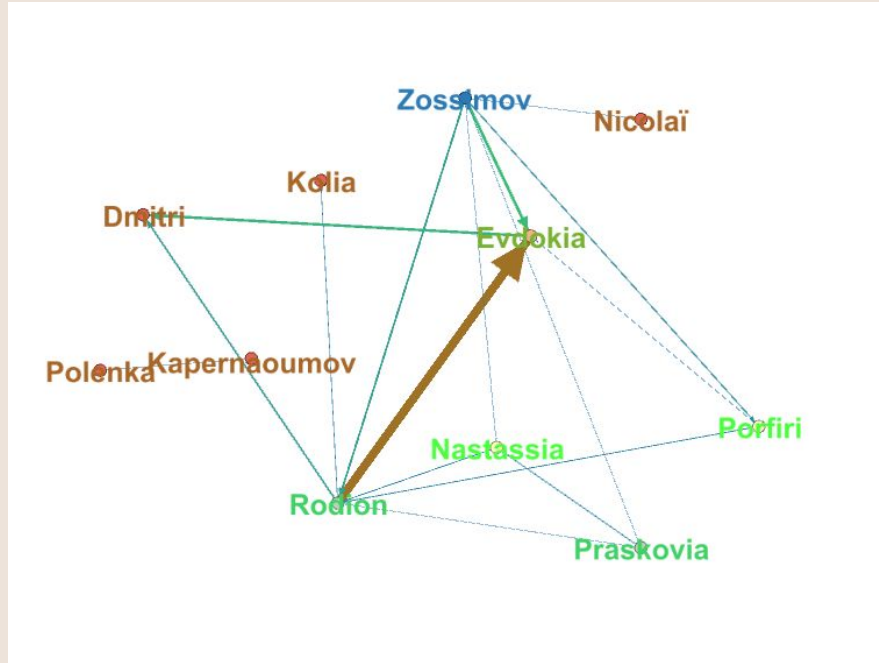
Personalisation du graphe :

- **Graphe dirigé** pour essayer d'obtenir des informations supplémentaires ;
- **Ajustement** des labels, de la taille des nœuds, etc ;
- **Déchevauchement** ;
- **Coloration dégradée** des relations orientées, des labels et des nœuds.

Observations :

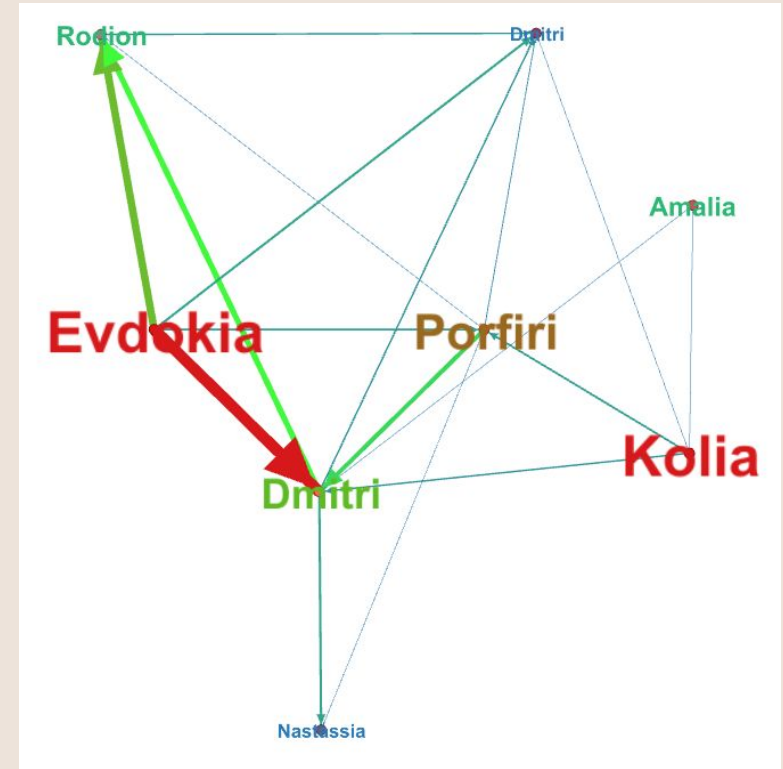
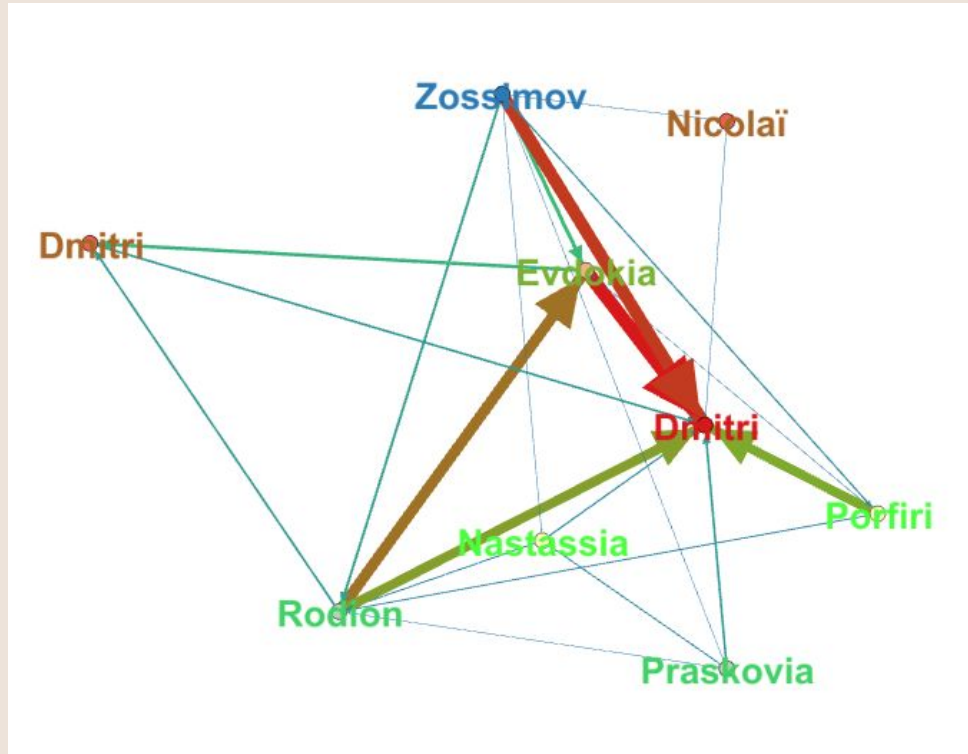
- K-core de 2
- Plage de degré sortant entre 1 et 5

III. Représentation graphique avec Gephi 2/3



Comparaison graphes des tomes 1 à droite et 2 à gauche : *plage de degré sortant entre 1 et 5*

III. Représentation graphique avec Gephi 3/3



Comparaison graphes des tomes 1 à droite et 2 à gauche : *k-core de 2*

The background is a light beige color with various abstract geometric elements. There are several circles of different sizes and colors, including purple, blue, and cyan. Some circles have a gradient. There are also dotted lines and small circles connected by lines, creating a network-like pattern. A solid blue vertical line is positioned to the left of the Roman numeral 'IV'.

IV

Analyses graphiques des données

IV. Analyses graphiques des données

- **Personnage principal Rodion :**
 - n'est pas le centre du graphe ;
 - n'a pas un fort poids dans les relations ;
- **Personnages les plus représentés :**
 - *Evdokia*, personnage mineur dans le roman, sœur du protagoniste Rodion ;
 - *Kolia* et *Dmitri Prokofitch Vrasoumikhine*, amis proches du protagoniste ;
 - Porfiri, l'enquêteur de police.
- Beaucoup de personnages mineurs sans relations ;

⇒ **L'histoire mentionne principalement les noms de certains principaux personnages qui sont proches du protagoniste**

CONCLUSION

- **Difficulté de recueillir les personnages du fait de la pluralité de leur dénomination** : il faut nécessairement connaître un minimum le livre et les personnages pour *éviter les doublons* notamment (limite de l'usage d'une regex) ;
- **Problème du format des pages** : problème de la taille de police d'écriture qui ne permet pas d'extraire les personnages et leurs relations correctement même par page ;
- **Trier les personnages** uniquement cités dans un paragraphe mais **non présents** car cela fausse le poids réel des relations ;
- **Problématique de l'usage de la première et troisième personnes** :
 - l'écrivain utilise ces pronoms personnels ;
 - impossibilité de relever le poids véritable des relations entre les personnages

BIBLIOGRAPHIE

Site de GEPHI : <https://gephi.org/>

Liste de personnages :

- Wikipédia : https://fr.wikipedia.org/wiki/Crime_et_Ch%C3%A2timent
- Academic.com : <https://fr-academic.com/dic.nsf/frwiki/467040#Personnages>
- Liste établie par le correcteur du groupe Ebooks : Crime_chatiment_FD_3tomesEN1.pdf