

LES PERSONNAGES DE LA SAGA HARRY POTTER

Du texte au graphe de relations

28 mai 2019

Soutenance MODAL INF473G

Arnaud RIPPOL
Hector₁SAINVET

Les personnages de la saga Harry Potter

1. Objectifs et démarches
2. Extraction du texte
3. Analyse des personnages
4. Confrontation aux sentiments
5. Discussions, critiques et conclusion

1. Objectifs et démarches

- Pourquoi de l'analyse de texte ?
- Pourquoi « Harry Potter » ?
- Concevoir les étapes du projet



Quels outils ?

- Données initiales:
 - Textes bruts en .txt en anglais en tome
 - Harry Potter API
- Deux méthodes d'extraction:
 1. Méthode directe : reconnaissance des noms avec majuscules
 2. Stanford NLP
 - A. Name Entity Recognition
 - B. POS et ne garde que NNP
- Croisement avec l'API

"That's really nice of her," said Harry, trying the fudge, which was very tasty.

His next present also contained candy – a large box of Choco-late Frogs from Hermione.

This only left one parcel. Harry picked it up and felt it. It was very light. He unwrapped it.

Something fluid and silvery gray went slithering to the floor where it lay in gleaming folds. Ron gasped.



2. Extraction du texte

Etapas progressives:

1. Importer le texte
2. Lire le texte
3. Découper le texte
4. Reconnaître les noms :
 - A. Include
 - B. Noms et alias
 - C. Jaro ?

- Croiser les bases de données API JSON

```
Potter - Harry Potter - 0.4166666666666667  
Potter - James Potter I - 0.41269841269841273  
Potter - Lily J. Potter - 0.3383838383838384
```

Exemple de confusion de détection avec Jaro

3. Analyses des personnages

Définition de la mesure :

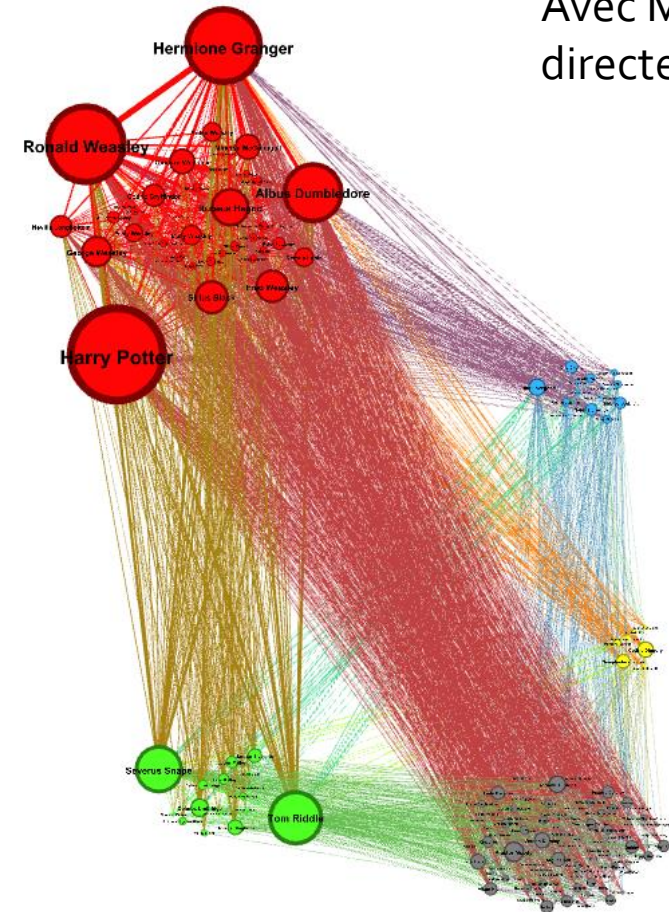
- Taille nœud : Nombre de chapitres présents
→ Avec un seuil sur un chapitre donné
- Poids lien : Nombre chapitre en commun

Critiques de la mesure :

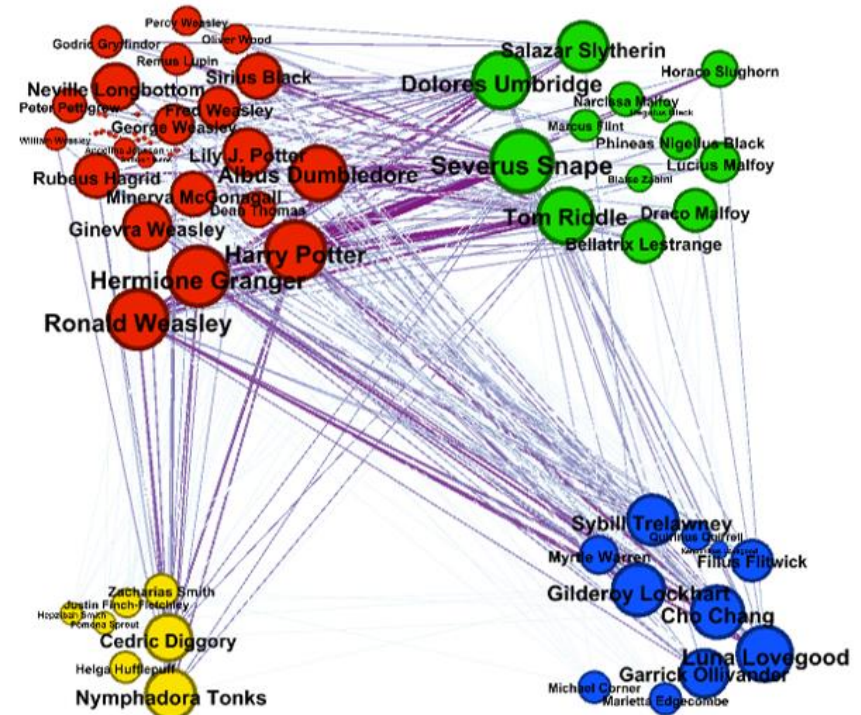
Notion de présence

dépendance entre nœud et liens

Tous livres confondus
Avec Méthode
directe

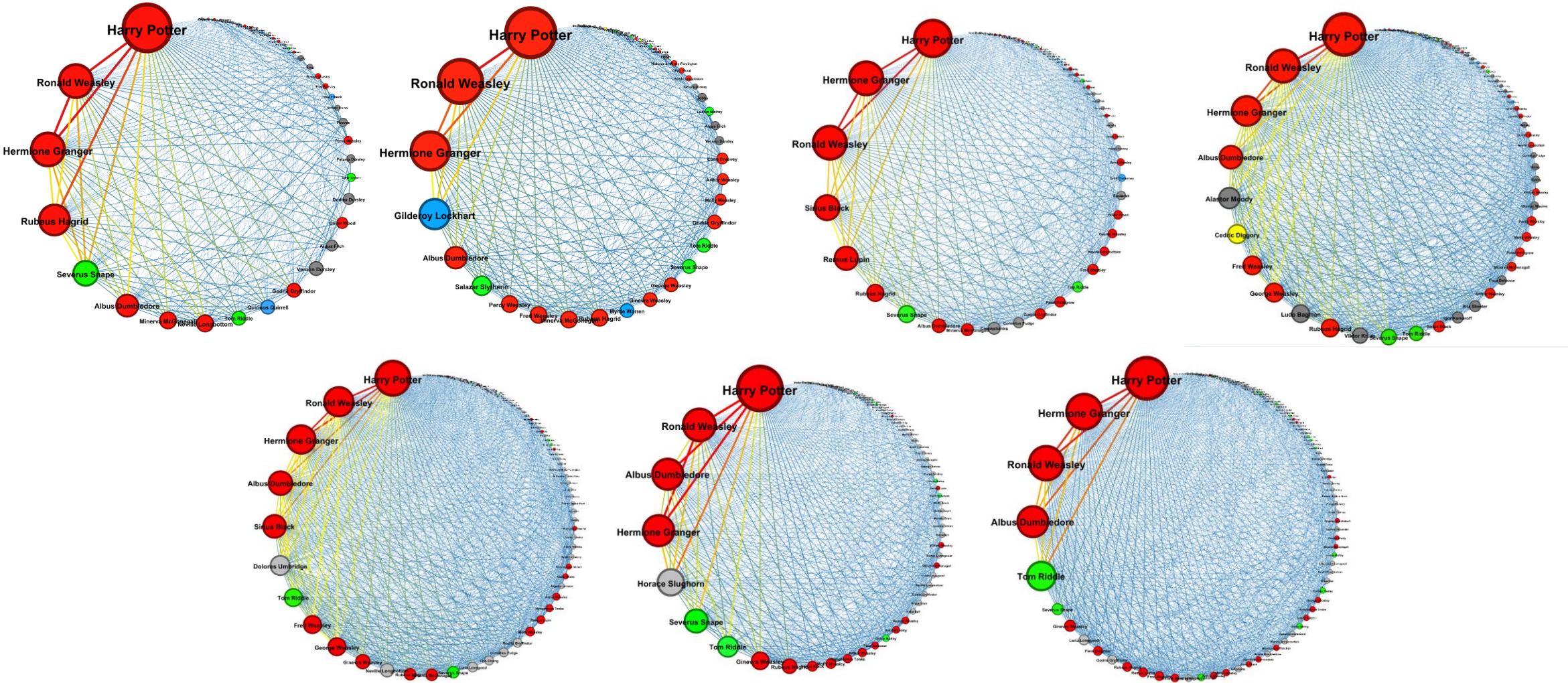


Tous livres confondus Avec Stanford



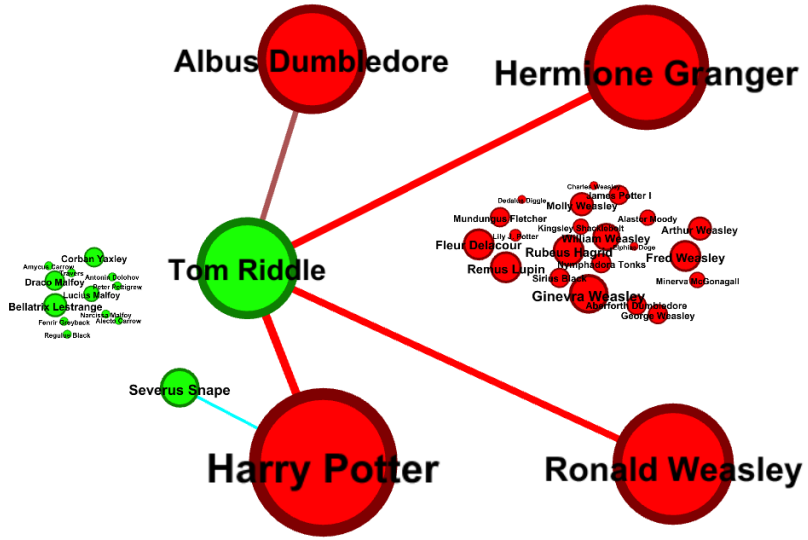
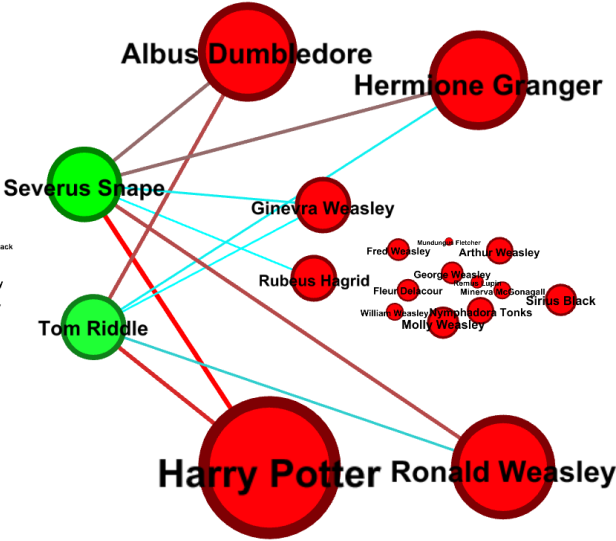
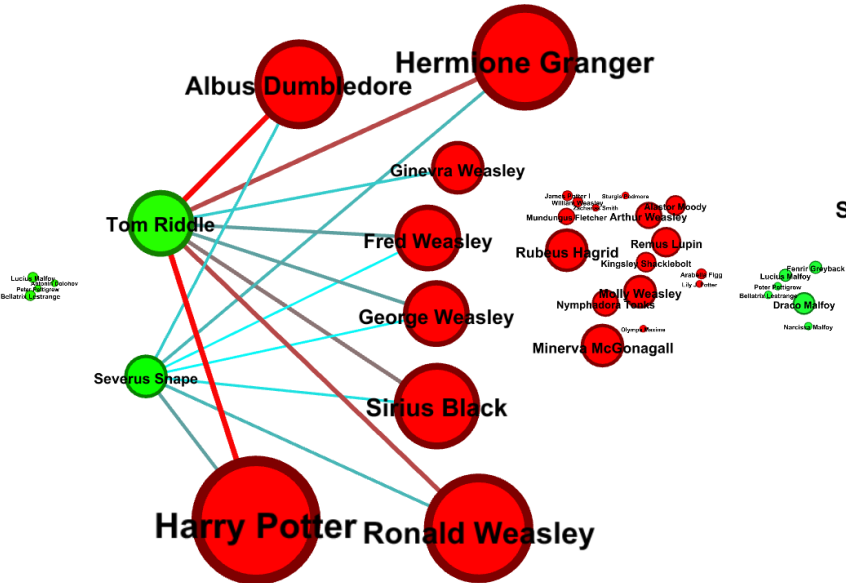
- Par tome : suivre un personnage ex: Voldemort

1	2	3	4
5	6	7	



- Par factions : ordre du phénix/mange-morts

5	6	7
---	---	---



Requête

Question :

- Dans quelle mesure les maisons interagissent entre elles plutôt qu'avec les autres ?

→ Degré d'introversion

→ Pour une maison donnée, quel est son rapport d'interaction interne/externe ?

```
1 MATCH (q:Perso)<-[]-(s:Perso)-[]->(t:Perso)
2 WHERE s.House = t.House AND s.House = "Gryffindor"
3 WITH s,COUNT(DISTINCT t) AS maison, COUNT(DISTINCT q) AS tous
4 WITH s,maison*1.0/tous AS rapport
5 WITH sum(rapport) AS tot, count(s) AS n
6 RETURN tot*1.0/n AS Rapport_Gryffindor
```



Rapport_Gryffindor
0.4562710228015806
Rapport_Slytherin
0.13507225543515863
Rapport_Ravenclaw
0.043478260869565216

4. Confrontation aux sentiments

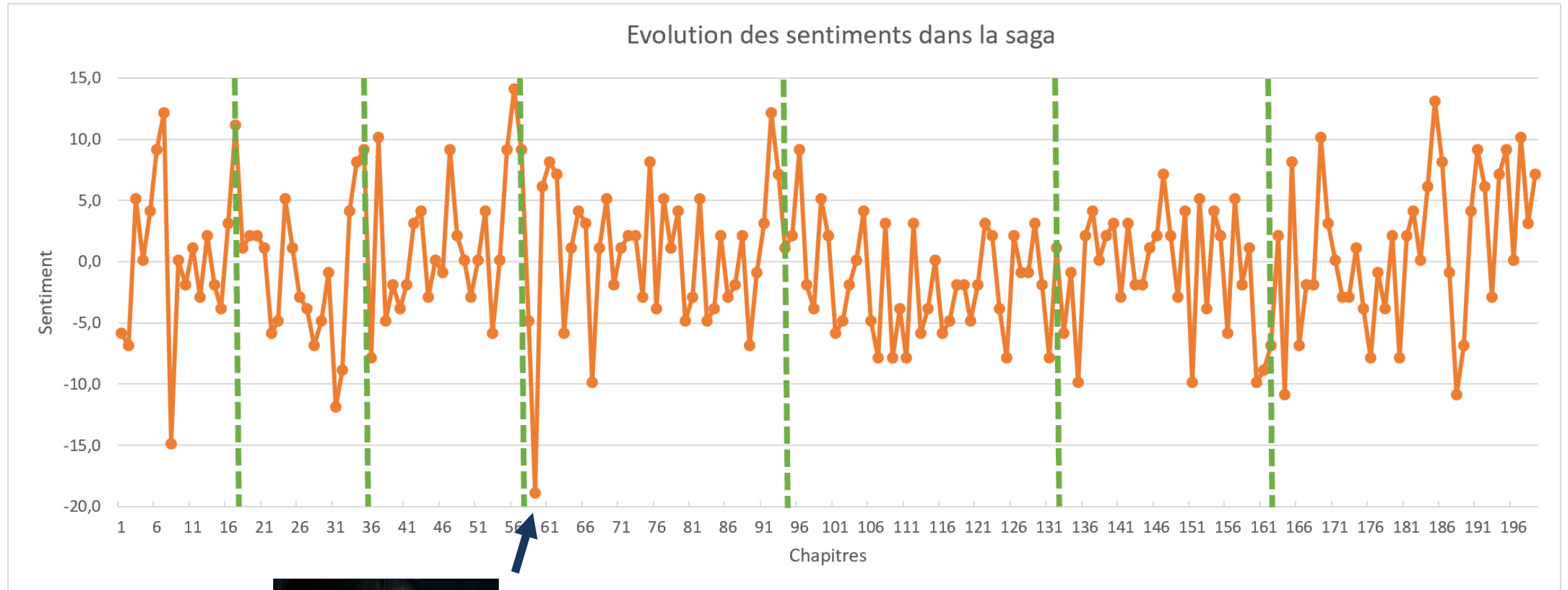
- Pourquoi les sentiments ?
- Comment les détecter ? Stanford NLP
 1. Tree
 2. Sentiment

→ Comptage du nombre de phrases Pos, Neu, Neg

→ On définit un score de sentiment :

1. $\text{Pos} - \text{Neg}$
2. $\text{Pos} + \text{Neu} - \text{Neg}$
3. $3 * \text{Pos} - \text{Neg}$

Par chapitre



Tome 1



Apparté curieuse

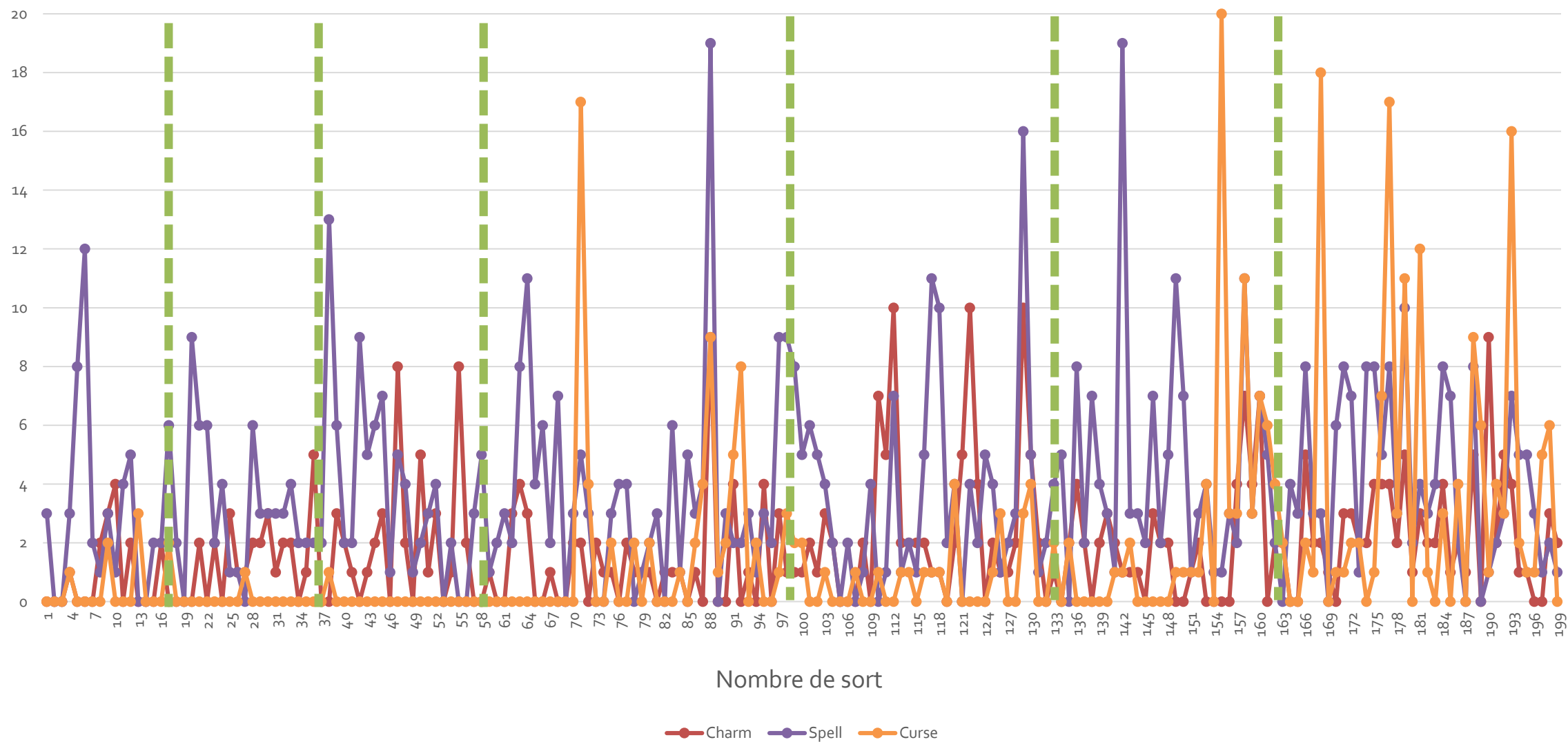
- Comparaison avec les sorts

- 6 types :

Charm, Spell, Curse, Enchantment, Hex, Jinx

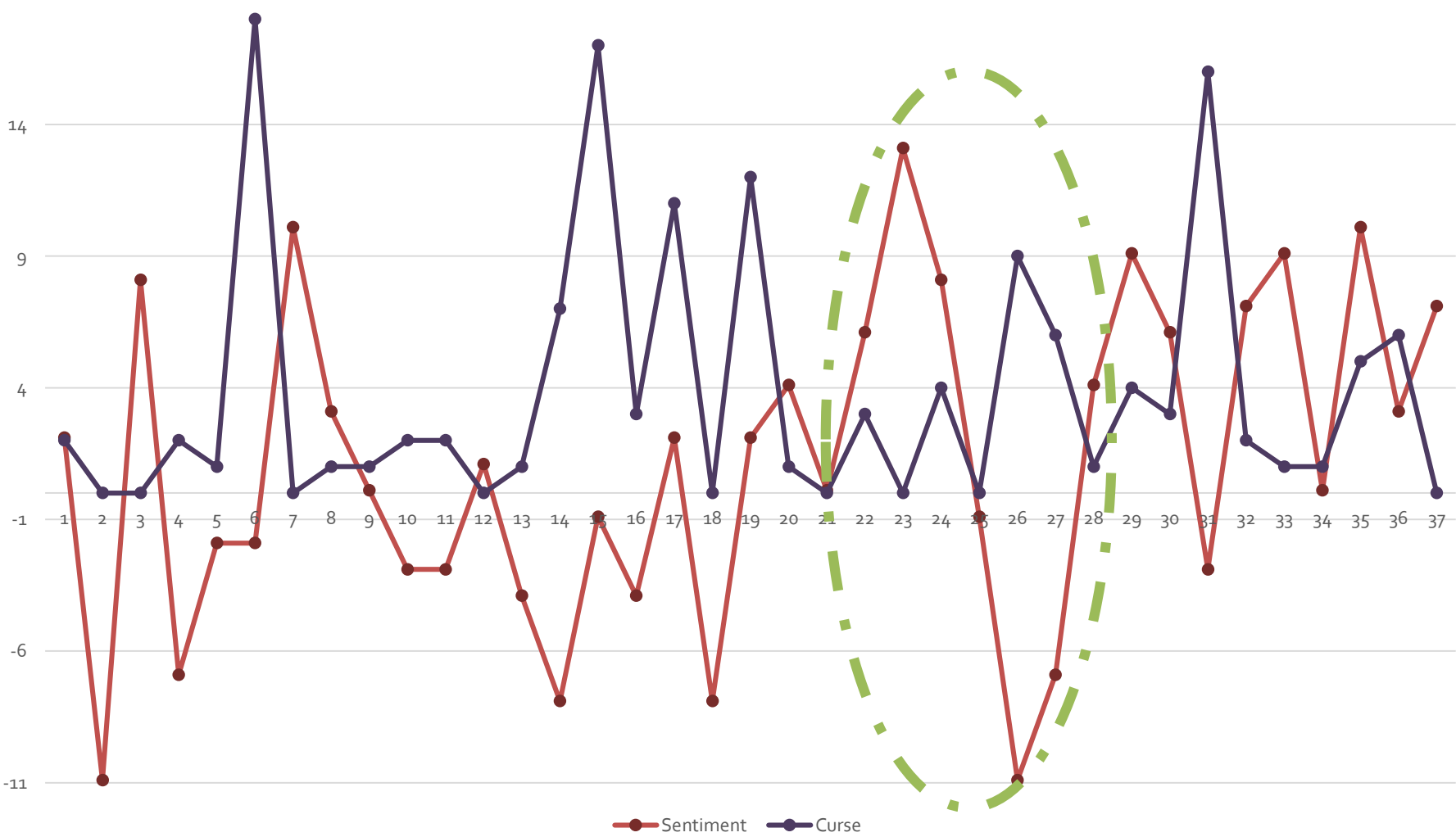
→ Intuition testée : Les sorts indiquent l'évolution des sentiments





Nombre de sort

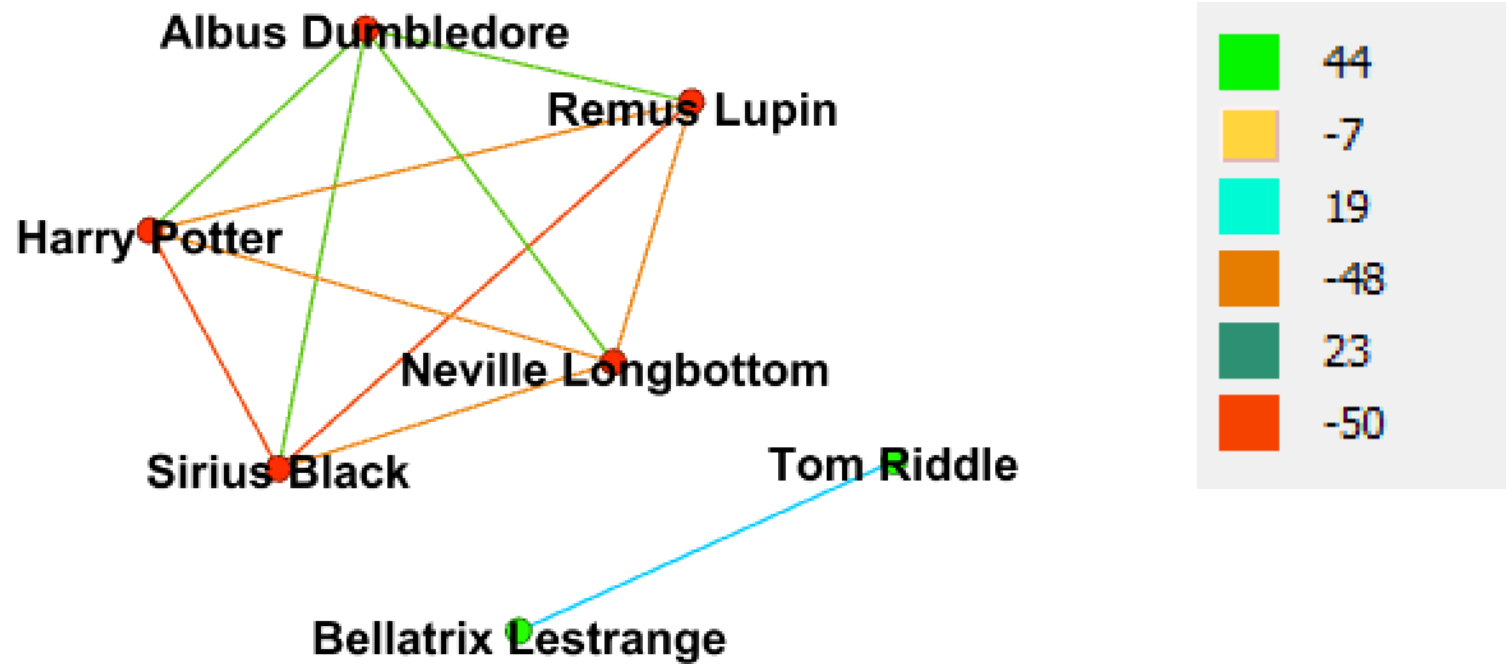
Tome 7 : Comparaison Sentiment/Curse



Rapporter les sentiments aux personnages?

- On définit la « positivité »:
 - Pour un chapitre : score-sentiment (pos – neg)
 - Pour un personnage : moyenne des positivités sur les chapitres où il est présent
 - Pour un couple de personnage : moyenne des positivités des chapitres en commun
- Résultats peu cohérents par rapport à l'œuvre
 - Ex: Harry Potter est neutre; Draco Malefoy est assez positif (comme sa famille)

Livre 5 Chapitre 36



5. Discussions, critiques et conclusion

- La méthode d'extraction
- Le choix de la mesure :
 - Ce que signifie la présence dans le livre
- Le choix des logiciels
- Très ludique