

Rapport Textmining



Analyse du réseau social des sorciers

ANALYSE CENTRE SUR LE PREMIER VOLET D'HARRY POTTER
CELIAN RINGWALD | UNIVERSITE LYON 2

TABLE DES MATIERES

I.	INTRODUCTION.....	2
II.	DESCRIPTION DU PROCESSUS D'ANALYSE	2
A.	Le choix de la fenêtre d'analyse	2
B.	Création d'une liste de personnage	3
C.	Contextualisation des relations.....	4
III.	RESULTATS ET ANALYSE DES RESULTATS.....	6
1.	THE BOY WHO LIVED	6
2.	THE VANISHING GLASS.....	7
3.	LETTERS FROM NO ONE.....	8
4.	THE KEEPER OF THE KEYS	9
5.	DIAGON ALLEY	10
6.	THE JOURNEY FROM PLATFORM NINE AND THREE-QUARTERS.....	10
7.	THE SORTING HAT	11
8.	THE POTIONS MASTER.....	12
9.	THE MIDNIGHT DUEL	13
10.	HALLOWEEN	14
11.	QUIDDITCH	15
12.	THE MIRROR OF ERASED	16
13.	NICHOLAS FLAMEL	17
14.	NORBERT THE NORWEGIAN RIDGEBACK	18
15.	THE FORBIDDEN FOREST	19
16.	THROUGH THE TRAPDOOR.....	20
17.	THE MAN WITH TWO FACES.....	21
IV.	OBSERVATIONS FINALES	22
V.	CONCLUSIONS	25
VI.	SITO-GRAPHIE & REFERENCES	26



I. INTRODUCTION

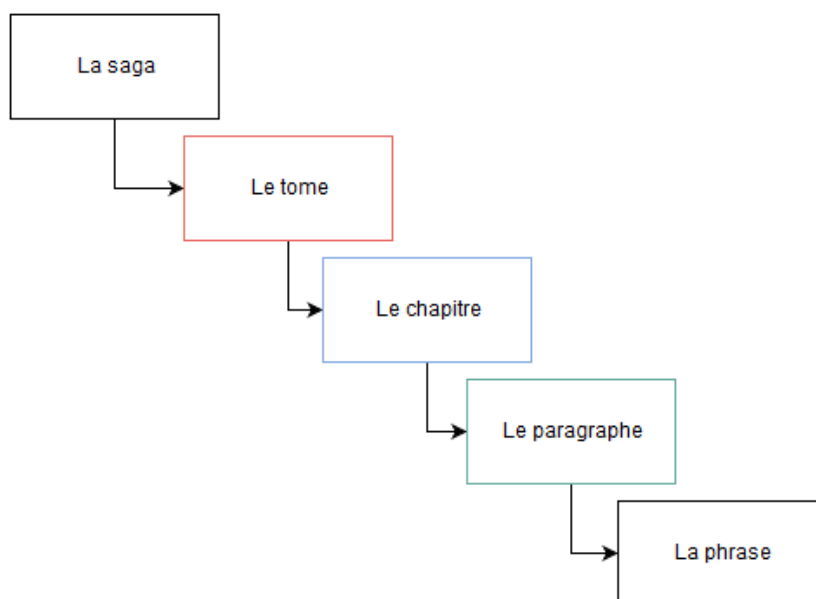


Ce rapport fait état du travail réalisé dans le cadre du cours de fouille de données textuelle. Centré sur l'étude de la série de J.K. Rollinks : Harry Potter. Disposant des 7 tomes de l'aventure de notre sorcier nous devons construire le graphe social permettant de comprendre la structure du réseau des personnages. C'est en utilisant les compétences vu en cours que nous avons pu réaliser ce travail tout en explorant différentes pistes innovantes comme l'analyse de sentiments et des outils d'analyse et visualisation de graphes. Ce travail est le résultat d'une analyse du premier roman et permet d'obtenir une cartographie des relations entre les principaux personnages par chapitre.

II. DESCRIPTION DU PROCESSUS D'ANALYSE

A. Le choix de la fenêtre d'analyse

Le premier point sur lequel il fallut se pencher était de savoir comment approcher le jeu de données en effet



l'analyse pouvait être réalisé à différents niveaux :

chapitre	nb token	nb mots moyen
chap.1	111	42.24324
chap.2	100	35.95000
chap.3	133	30.34586
chap.4	151	25.35762
chap.5	285	23.81053
chap.6	286	22.91259
chap.7	186	24.90323
chap.8	90	35.28889
chap.9	210	24.65714
chap.10	164	27.48780
chap.11	137	25.96350
chap.12	206	27.73786
chap.13	122	27.94262
chap.14	137	26.43796
chap.15	187	28.26203
chap.16	315	21.42540
chap.17	226	25.30973

Nous avons donc du extraire chaque tome des fichiers fournis, les découper par chapitre, et découper ceux-ci en token, représentant un paragraphe. Ceux-ci nous permettent donc d'obtenir des fenêtres de 25 à 30 mots en moyenne.





Les différents chapitres de ces livres étant généralement marqués par le titre de celui-ci nous avons dû aller le rechercher à l'aide d'une expression régulière :

```
#catch index of each chapter
chap_h1=as.data.frame(str_locate_all(h_1, "CHAPTER. [A-Z]{1,9}"))
chap_h2=as.data.frame(str_locate_all(h_2, "CHAPTER. [A-Z]{1,9}"))#18
chap_h3=as.data.frame(str_locate_all(h_3, "CHAPTER. [A-Z]{1,9}"))
chap_h4=as.data.frame(str_locate_all(h_4, "CHAPTER. [A-Z]{1,9}"))
chap_h5=as.data.frame(str_locate_all(h_5, "CHAPTER. [A-Z]{1,9}"))
chap_h6=as.data.frame(str_locate_all(h_6, "Chapter. {1,9}"))
chap_h7=as.data.frame(str_locate_all(h_7, "CHAPTER. {1,13}/n"))#36
```

Cependant le titre du chapitre n'étant pas assez normalisé nous n'avons pas pu obtenir un découpage correct pour les volumes 2 et 7 de la saga (moins de chapitres obtenus qu'attendu : 16 au lieu de 18 dans le volume 2 et 35 au lieu de 36 dans le dernier volume), c'est donc à la suite de ce problème que je me suis résigné travailler que sur le premier volume.

B. Création d'une liste de personnage



Il fallut dans un second temps trouver un moyen de déterminer un ensemble de personnages. Il y avait donc une première piste qui consistait à extraire toutes les mots commençant par une majuscule et qui n'était pas précédé d'un point. La seconde consistait à utiliser un tagger déjà implémenté nous permettant de repérer ce que l'on appelle des entités nommées. C'est donc ce que nous avons fait grâce à la librairie OpenNLP d'apache, une boîte à outil pour le TAL. Celle-ci nous permet de reconnaître plusieurs types d'entité : temporelle, des lieux ainsi que des personnages (basé sur un modèle Maximum-entropy Markov).

Grâce à cet outil nous avons pu obtenir deux listes l'une regroupant tous les personnages apparaissant dans le livre et l'autre les lieux cités dans celui-ci. Ces listes ne sont pas exactes dans le sens où certaines fois un prénom est reconnu comme un lieu, de plus certaines entités ne correspondent et d'autre fois une entité nommée.

	Vrai positif	Faux positifs	Total
Personnages	127	16	143
Lieux	21	21	42

En effet dans la liste de lieux nous obtenons beaucoup de noms, et y sont aussi insérés les noms des 4 maisons de Poudlard. Cet effet est peu marqué dans la liste des personnages, dans laquelle nous retrouvons cependant beaucoup de doublons, en effet certains personnages sont appelés sous différentes appellations, par exemple Ron est aussi appelé Ronald dans certains passages et dans d'autre l'on le retrouve sous l'appellation Ron Weasley, ou M. Weasley.



Nous avons donc créé regrouper toutes ces appellations et supprimer certains personnages comme les fantômes ou ceux cités qu'une seule fois dans le roman par exemple : Newt Scamander. A la suite de cette manœuvre nous obtenons une liste de 43 personnages.

Afin de repérer facilement dans le corpus les moments où chacun de ces personnages est cité nous avons donc dû remplacer toutes les appellations de celui-ci dans le roman par celui défini dans la liste.

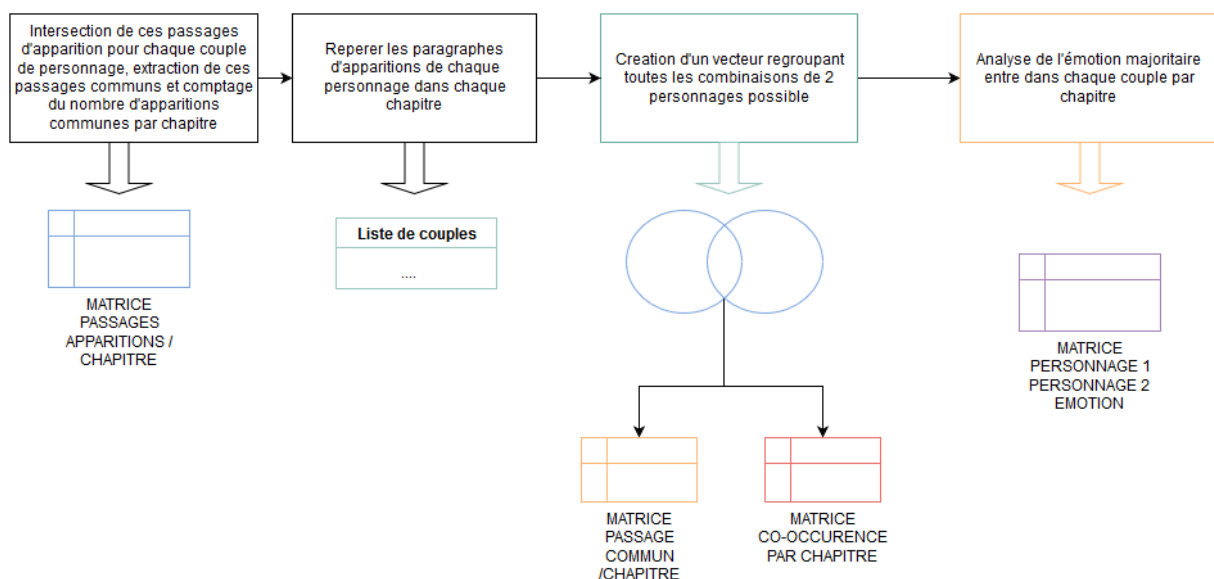
C. Contextualisation des relations



Afin de contextualiser les relations entre nos différents personnages nous avons plusieurs solutions :

- ⚠ s'axer sur des thématiques extraites via une technique de topic modeling
- ⚠ utiliser la liste de lieux fournie par le tagger utilisé par notre extracteur d'entités
- ⚠ réutiliser notre tagger pour définir la place du sujet dans la phrase

Mais nous pouvions simplement nous appuyer sur les tokens que nous avons extraits en effet si deux personnages se retrouvent dans le même paragraphe c'est qu'ils sont concernés par le même fil narratif. C'est donc sur ces séquences que nous nous sommes appuyé. Nous avons donc comme l'illustre le graphique ci-dessous repéré les paragraphes où apparaissait chaque personnage. En créant la liste de tous les couples de deux personnages possible nous avons ensuite pu réaliser l'intersection des passages apparaissant ceux-ci. C'est à la suite de cela que j'ai pu obtenir la matrice de cooccurrence (basé sur la somme du nombre de paragraphe commun à chaque couple par chapitre) ainsi que la matrice regroupant tous les passages communs de ce couple.



Ayant obtenu la liste de chaque passages communs à chaque couple je pouvais alors me permettre d'analyser le sentiment principal sous-jacent à ces passages. Nous avons donc opté pour une solution basé sur un Lexique et un model naïf bayésien : le package « sentiment » développé par Timothy P. Jurka.

Afin d'obtenir une vue d'ensemble la fonction `classify_emotion()` de ce package je suis donc entré dans le code afin d'en comprendre un peu la démarche. Dans un premier temps celui-ci charge un lexique où sont étiquetés 1541 mots selon 6 catégories : joie, dégoût, colère, la tristesse et surprise. Pour chacun des passages passés en entrée la fonction attribue alors un score calculé comme suit :

$$score(passage_i, categorie_j) = \log(nb.mots_{categorie.j} / nb.mots_{lexique}) + \sum_{k=1}^{mot.passage_k} \log(1 / nb.mots_{nb.mots_{categorie.j}})$$

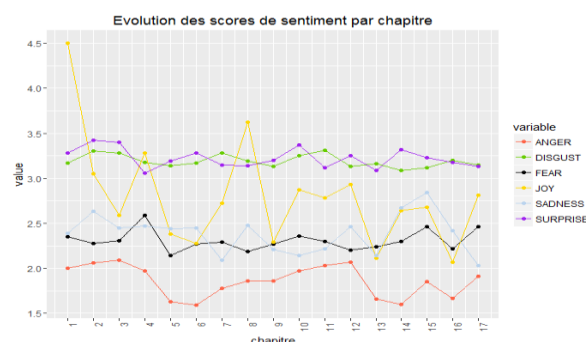
Afin d'obtenir un score cohérent l'auteur se permet de ramener à la valeur absolue ce score. Nous noterons un biais dû à la pauvreté du champ lexical du dégout fourni dans la base, la fonction aura tendance à surestimer le pouvoir de ce sentiment dans le passage :

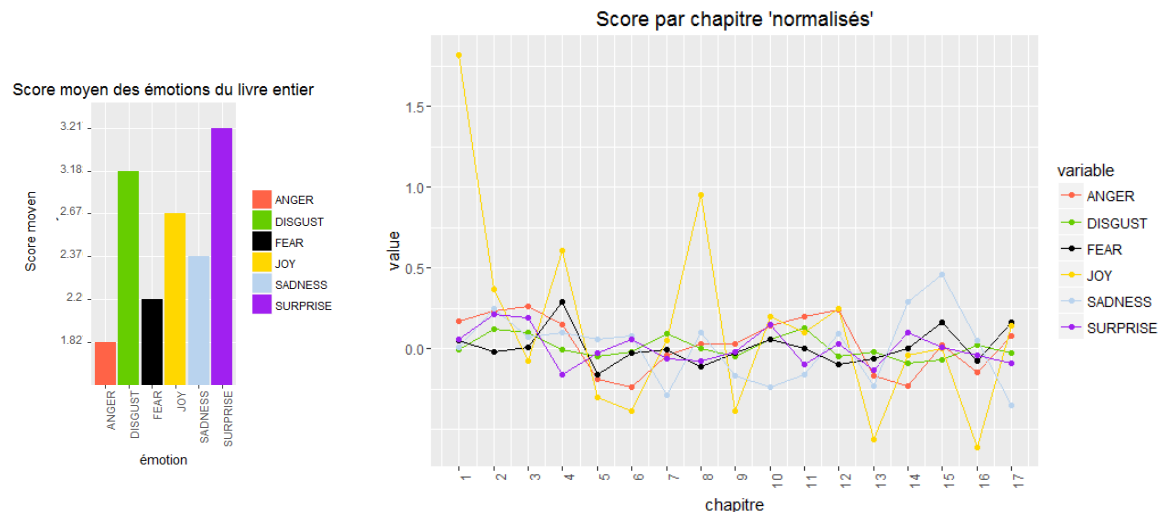
	nb.mots	log(nb.mots/total_mots)
anger	355	-1.46871776464786
disgust	70	-3.09234031207392
fear	195	-2.06783599555953
joy	553	-1.02547755260094
sadness	274	-1.7277074477352
surprise	95	-2.78695866252273

En effet l'on remarque que sans qu'un mot lié au dégout apparaisse l'on obtient un fort score de base pour le dégoût ! Ce que l'auteur s'est permis de corriger, dans le cas où le dégout est définis comme sous-jacent à un texte (meilleur score), celui-ci vérifie si cela n'est pas dû à une valeur trop élevé du score de ce sentiment et le définit alors comme un sentiment non défini.

C'est donc en utilisant cette fonction, maintenant connue que j'ai pu identifier l'**émotion sous-jacente au passage commun** entre chaque couple de personnage par chapitre.

Nous traçons les scores de chaque sentiment sur les 17 chapitre de notre livre :





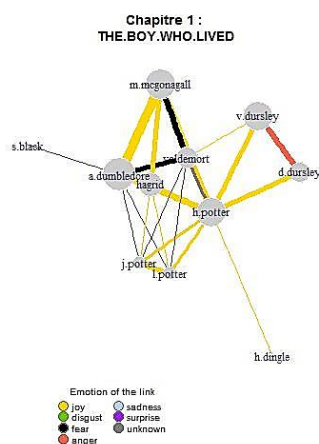
Afin de vérifier si l'écriture de l'auteur possède un registre marqué par une émotion, nous traçons l'histogramme des scores moyen à tous les tokens du livre. Nous observons que la surprise est fortement présente, ainsi que le dégoût (peut-être dû à la pauvreté du champ lexical associé donnant un fort poids à cette émotion), et la joie.

III. RESULTATS ET ANALYSE DES RESULTATS

Dans cette partie nous réalisons une étude poussée du réseau social, obtenu via le package « igraph » pour chacun des chapitres du premier volet de la saga. Grâce aux matrices précédemment construites nous avons pu faire apparaître ces relations sous forme de réseaux sociaux. Ainsi nous avons pu faire apparaître :

- ⚠ le nombre de cooccurrence commune entre deux personnages via l'épaisseur du lien
- ⚠ l'importance d'un personnage (le nombre d'apparitions de celui-ci dans un chapitre), via la grosseur du nœud
- ⚠ le sentiment émergeant des passages communs à ces deux personnages via la couleur du lien
- ⚠ Une analyse du réseau : le personnage central, ses communautés, les personnages ayant le rôle de relais...

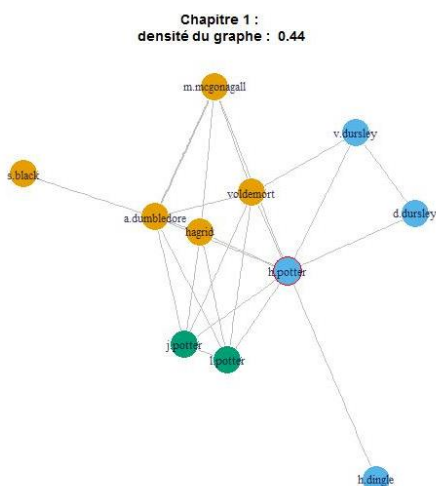
1. THE BOY WHO LIVED



Le premier chapitre du roman présente l'arrivée chez les Dursley de notre héros. L'émotion sous-jacente principale à ce chapitre est la joie suivie et la colère ce que nous retrouvons sur le premier réseau de gauche. Nous remarquons de plus une émotion de peur entre Dumbledore et Voldemort de même entre McGonagall et



Voldemort. Nous notons une forte colère dans les passages parlant de M. Dursley et son fils. En effet l'aventure commence et l'on comprend vite que notre héros n'est pas à sa place dans ce cadre. Lorsque nous nous concentrons sur une analyse plus poussée au niveau du réseau nous observons de un qu'Harry est le personnage central de celui-ci, que Dumbledore, Hagrid, Voldemort et les parents d'Harry (James et Lilly) jouent un fort rôle de relai dans le graphe (au point de vue de la centralité d'intermédiarité). Après un regroupement de ce beau monde via l'algorithme du fastGreedy (graphe ci-dessous) nous remarquons 4 communautés : celle formée par Harry et sa famille adoptive, celle de ses parents, et celle formée par les autres. Enfin nous noterons une grande clique par le groupe : Harry, ses parents Dumbledore et Voldemort. Il s'agit sûrement du passage où Dumbledore expose la miraculeuse survie de notre héros à l'attaque de celui-dont –on-doit-pas-prononcer-le-nom. Nous noterons aussi la première référence à Sirius Black, ainsi que l'apparition de M. Dingle un autre magicien accompagnant Dumbledore, McGonagall et Hagrid.

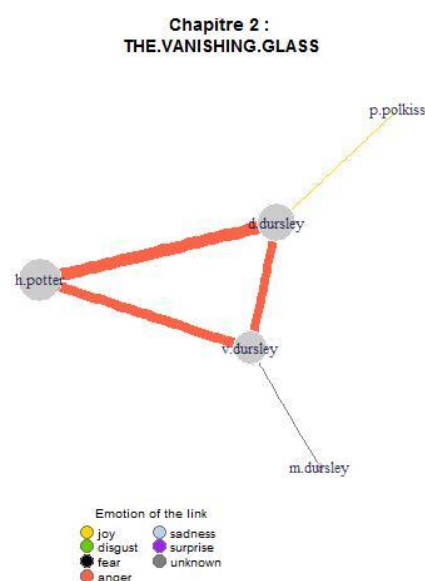


	nombre de degré	centralité d'inter
a.dumbledore	7	9.33
d.dursley	2	0
hagrid	5	6.13
h.dingle	1	0
h.potter	9	16
j.potter	5	9.6
l.potter	5	9.6
m.mcgonagall	4	0
s.black	1	0
v.dursley	3	0
voldemort	6	7.93

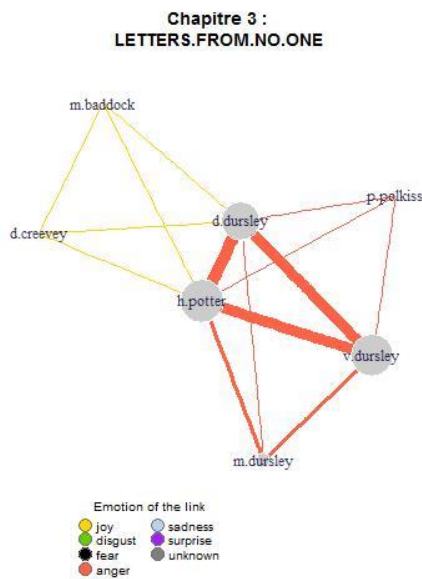
2. THE VANISHING GLASS

La vitre disparue est le second chapitre du roman, durant celui-ci une forte colère apparaît dans la clique formée par M. Dursley son fils et Harry. En effet notre héros crée le scandale en faisant disparaître une vitrine sur laquelle était appuyé son cousin, le jour de son anniversaire ! Ce chapitre ne fait apparaître que 5 personnages.

	nombre de degré	centralité d'inter
d.dursley	3	3
h.potter	2	0
m.dursley	1	0
p.polkiss	1	0
v.dursley	3	3



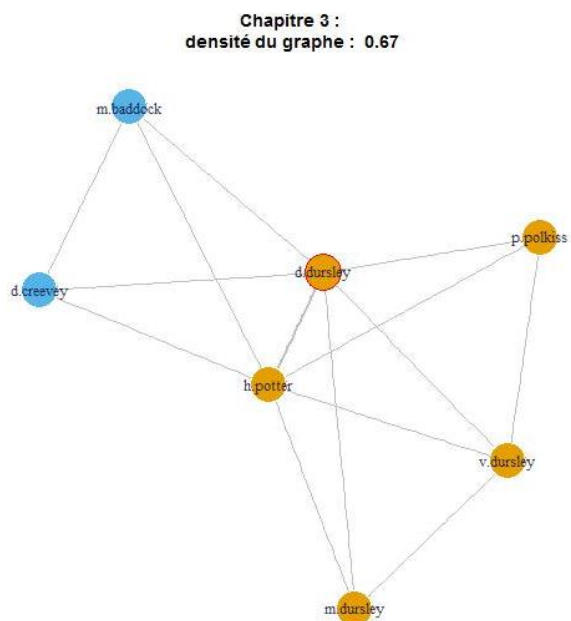
3. LETTERS FROM NO ONE



Le troisième chapitre du roman est celui durant lequel Harry reçoit les lettres d'inscription pour Poudlard, cependant son Oncle fait tout pour l'empêcher de lire ces lettres. Nous remarquons une aberrance sur ce graphe... En effet Malcom Baddock et Denis Creevey futurs camarades d'Harry ne peuvent pas être présent dans ce chapitre... Après vérification il semblerait que nous ayons remplacé par mégarde Dennis et Malcom deux amis de Dudley qui martyrisent notre ami à lunettes. Nous remarquons une forte colère dans les passages liant

Harry et les Dursley, contrecoup de ce qu'il s'est passé au Zoo et conséquence du fait une ceux-ci ne permettent pas à Harry de lire ces fameuses lettres. Le personnage de ce chapitre est cette fois-ci Dudley, notons que Piers Polkiss (un autre amis de Dudley) tiens le rôle d'un véritable relai dans ce passage, bien devant Harry.

Enfin nous pouvons observer une clique composée de 4 personnes : Dursley, son père, Piers et Harry, et nous pouvons observer deux communautés sur le graphe suivant :

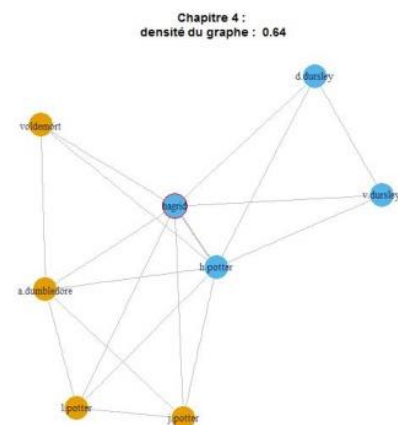
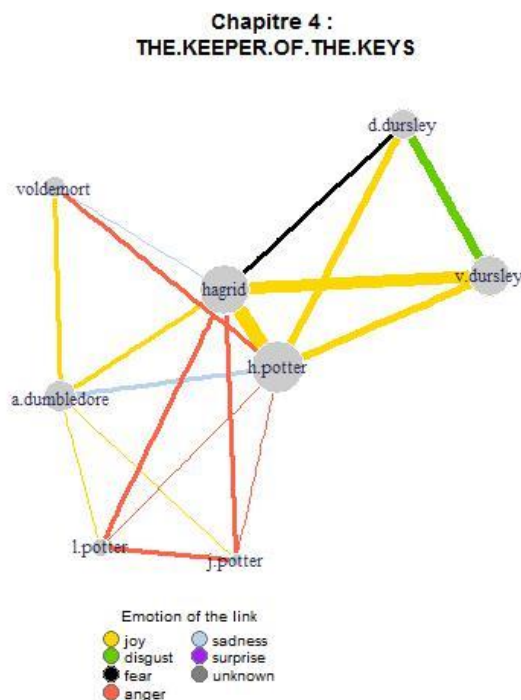


	nombre de degré	centralité d'inter
d.creevey	3	0.33
d.dursley	6	5
h.potter	6	2
m.baddock	3	0.33
m.dursley	3	0
p.polkiss	3	4.33
v.dursley	4	0



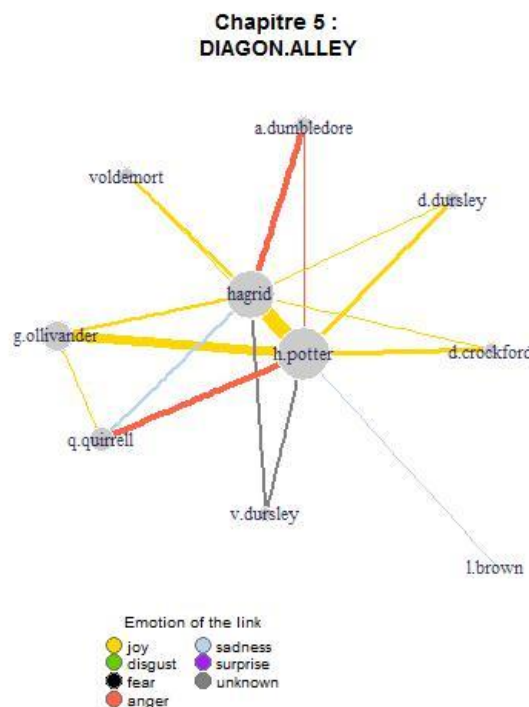
4. THE KEEPER OF THE KEYS

Le chapitre 4 marque l'arrivée d'Hagrid, qui vient chercher Harry afin de lui remettre en main propre la lettre d'admission à Poudlard que son oncle et sa tante s'étaient évertués à lui cacher. Ce qui se ressent sur le graphe, car Hagrid en est le personnage central de ce réseau. Nous notons la grande joie d'Harry en présence de ce personnage barbu, de la peur entre Hagrid et Dudley : en effet Hagrid fait pousser une queue de cochon à celui-ci, ce qui explique le dégoût que peut ressentir Vernon Dursley. C'est à ce moment qu'Hagrid dévoile à Harry ses origines (ce qui crée la grande clique Harry/Hagrid /Les parents d'Harry/Dumbledore), d'où la colère qu'éprouve Harry pour Voldemort, ainsi que la colère d'Hagrid vis-à-vis des parents d'Harry. Nous notons deux communautés, celles des personnages présents physiquement dans ce chapitre et ceux qui y sont simplement évoqués.



5. DIAGON ALLEY

Durant le chapitre 5, Hagrid emmène Harry au chemin de traverse (The Diagon Alley) afin qu'il puisse y trouver son matériel de cours de sorcier, ils passent alors par le Chaudron Baveur. Harry personnage principal de ce chapitre y rencontre quelques sorciers que sont Quirrell (une rencontre marqué par de la colère pour Harry et de la tristesse pour Hagrid), ainsi que le vendeur de baguette M. Ollivander. Voldemort y est évoqué à plusieurs reprises. Nous ne pouvons observer de communauté sur notre graphe. Nous notons enfin une tristesse sous-jacente au passage où Hagrid évoque Dumbledore.

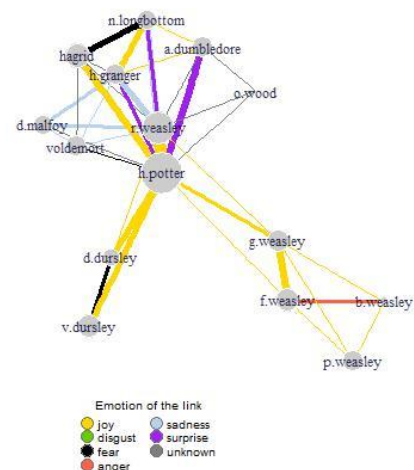


	nombre de degré	centralité d'inter
a.dumbledore	2	0
d.crockford	2	0
d.dursley	2	0
g.ollivander	3	0
hagrid	8	17.67
h.potter	9	14.67
l.brown	1	0
q.quirrell	3	0
v.dursley	2	0
voldemort	2	6.33

6. THE JOURNEY FROM PLATFORM NINE AND THREE-QUARTERS

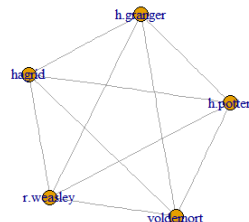
Il est temps pour Harry de partir pour Poudlard, pour cela Hagrid laisse Harry à la gare de Londre où celui-ci doit trouver la plateforme 9 ³/₄. C'est à l'aide de la famille Weasley (communauté verte) que celui-ci réussit à prendre son train. Il y rencontre Hermione, Drago et y évoque Voldemort (ces trois personnages apparaissent dans le roman dans des passages associés à Ron classés comme tristes). Nous noterons que

**Chapitre 6 :
THE.JOURNEY.FROM.PLATFORM.NINE.AND.THREE.QUARTERS**

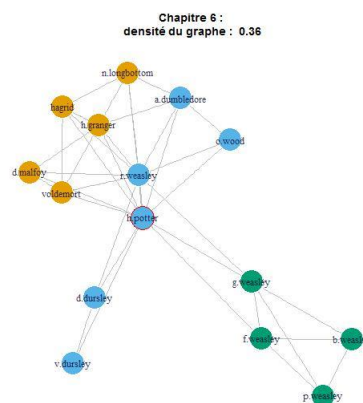


Harry lui paraît surpris dans les passages associés à Dumbledore notamment lorsque notre héros ouvre sa choco-grenouille, puis avec Hermione qui à la recherche du crapaud de Neuville.

Nous noterons une clique formée par Harry, Hagrid, Hermione, Ron et Voldemort.

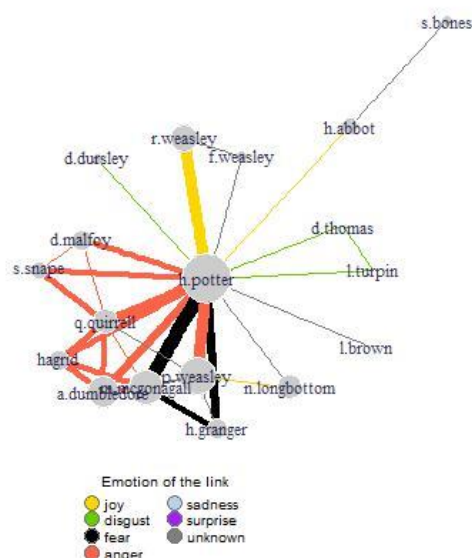


	nombre de degré	centralité d'inter
a.dumbledore	5	10.52
b.weasley	3	0
d.malfoy	4	0
d.dursley	3	8.1
f.weasley	4	4.97
g.weasley	5	17.62
hagrid	5	1.73
h.potter	12	17.6
h.granger	7	2.23
n.longbottom	5	0
o.wood	3	4.74
p.weasley	3	2.62
r.weasley	10	40.75
v.dursley	2	0
voldemort	5	10.36



7. THE SORTING HAT

Chapitre 7 : THE.SORTING.HAT

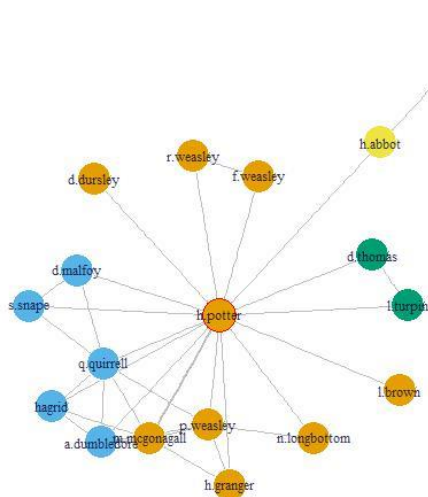


L'arrivée à Poudlard, marque la rencontre de nombre de nouveaux personnages, en effet celui-ci compte près de 18 personnages. Toujours tourné vers Harry nous identifions 4 communautés formées en orange par les élèves de Poudlard + Le professeur McGonagall, bleu les autres professeurs + Malfoy, en vert et jaune des élèves moins présents. Une fois arrivés les élèves sont alors rangés par maisons, c'est alors qu'Harry espère être destiné pour Gryffondor. Nous noterons de



la colère dominante dans les passages liant Harry au cluster bleu, ainsi que de la per envers Hermione et McGonagall et de la joie envers Ron.

Chapitre 7 :
densité du graphe : 0.22

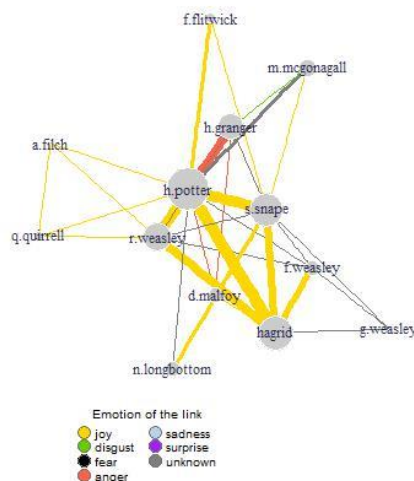


	nombre de degré	centralité d'inter
a.dumbledore	5	0.58
d.thomas	2	0
d.malfoy	3	7.33
d.dursley	1	0
f.weasley	2	16
hagrid	4	0
h.abbot	2	16
h.potter	16	99.83
h.granger	3	0
l.brown	1	0
l.turpin	2	0
m.mcgonagall	6	1.83
n.longbottom	2	31.5
p.weasley	6	33.25
q.quirrell	7	10.75
r.weasley	2	0
s.snape	3	0
s.bones	1	0

8. THE POTIONS MASTER

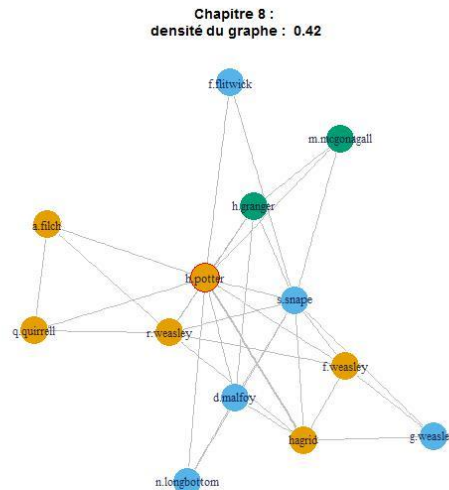
Les cours commencent alors pour nos trois amis, viens l'heure du premier cours de potion permet de avec Severus Snape et des premiers tours avec S. Flitwick. Hermione se fait alors remarqué comme une bonne élève et semble proche de McGonagall, Drago attise la colère.

Chapitre 8 :
THE.POTIONS.MASTER



	nombre de degré	centralité d'inter
a.filch	3	0.53
d.malfoy	5	8.7
f.flitwick	2	0
f.weasley	5	5.03
g.weasley	3	4.17
hagrid	6	1.33
h.potter	11	13.43
h.granger	5	5.37
m.mcgonagall	3	0
n.longbottom	3	0
q.quirrell	3	0.53
r.weasley	7	9.17
s.snape	10	15.7

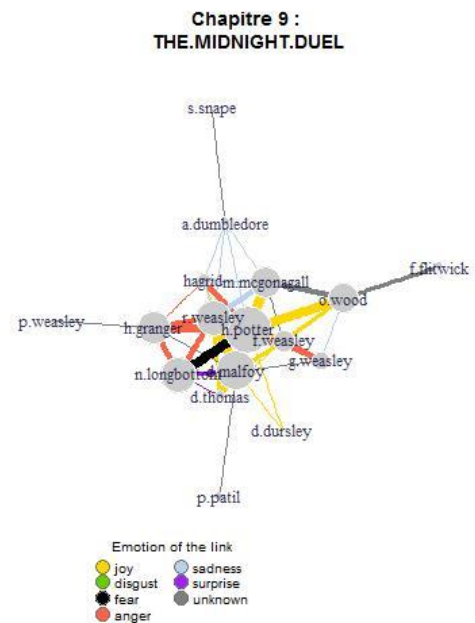
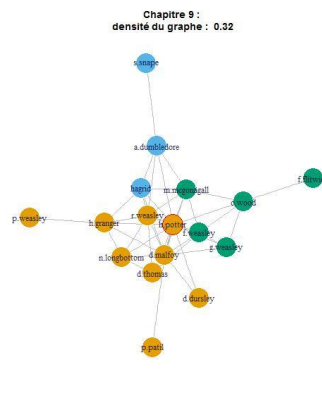




9. THE MIDNIGHT DUEL

Harry toujours central à notre réseau, récupère sur son balai, poussé par Malfoy, l'objet de Neuville (surprise de Neuville, peur d'Harry) que celui-ci avait mis sur le toit de l'école durant le cour de balais. McGonagall repère alors le garçon et le présente à Oliver Wood par le biais de McGonagall, ce qui lui permet de rejoindre l'équipe de Quidditch de Gryffondor où jouent Fred et Georges, les frères jumeaux de Ron. Après avoir provoqué en duel Harry Malfoy leur donne rendez-vous avec Ron mais appel M. Filch afin de faire renvoyer les trois amis, à qui ils échape de justesse et découvrent une pièce avec le fameux monstre évoqué par Dumbledore que possède Hagrid. Il semblerait que celui-ci protégeait quelque chose...

	nombre de degré	centralité d'int
a.dumbledore	5	20.83
d.thomas	5	6.33
d.malfoy	11	47.05
d.dursley	2	3
f.flitwick	1	0
f.weasley	7	12.93
g.weasley	4	16.6
hagrid	6	22.13
h.potter	12	5.97
h.granger	6	15
m.mcgonagall	7	5.18
n.longbottom	5	0
o.wood	5	15.33
p.patil	1	0
p.weasley	1	0
r.weasley	9	5.8
s.snape	1	0

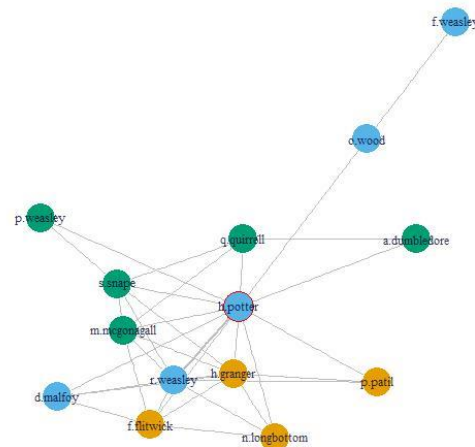


10. HALLOWEEN

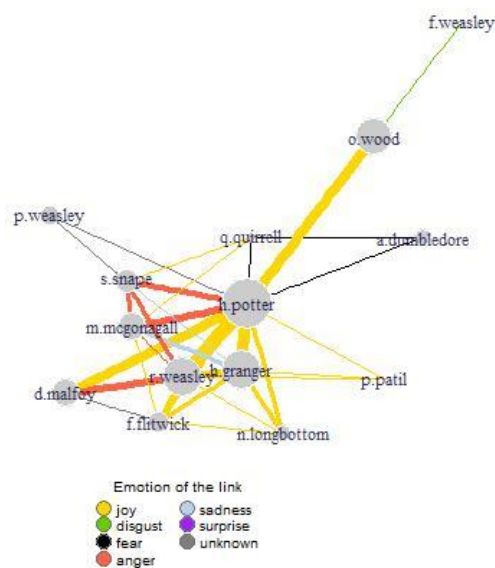
Harry reçoit son Nimbus 2000 et apprend les règles de Quidditch. Le jour d'Halloween, Hermione maîtrise parfaitement le sortilège de Lévitation durant le cours de sortilèges mais la remarque d'un Ron jaloux l'envoie pleurer dans les toilettes des filles où elle est piégée par un troll. Harry et Ron s'empressent d'aller la secourir et vainquent le troll. Les professeurs les retrouvent alors avec le troll assommé, d'où la colère envers nos héros Hermione dit un mensonge pour leur éviter les ennuis et ils deviennent tous trois amis.

	nombre de degré	centralité d'inter
a.dumbledore	2	0
d.malfoy	4	0.92
f.flitwick	6	6.08
f.weasley	1	0
h.potter	12	32.83
h.granger	8	8.08
m.mcgonagall	6	5.67
n.longbottom	4	0.75
o.wood	2	12
p.patil	3	10.5
p.weasley	2	1.5
q.quirrell	4	9.92
r.weasley	8	3.83
s.snape	6	5.92

Chapitre 10 :
densité du graphe : 0.37

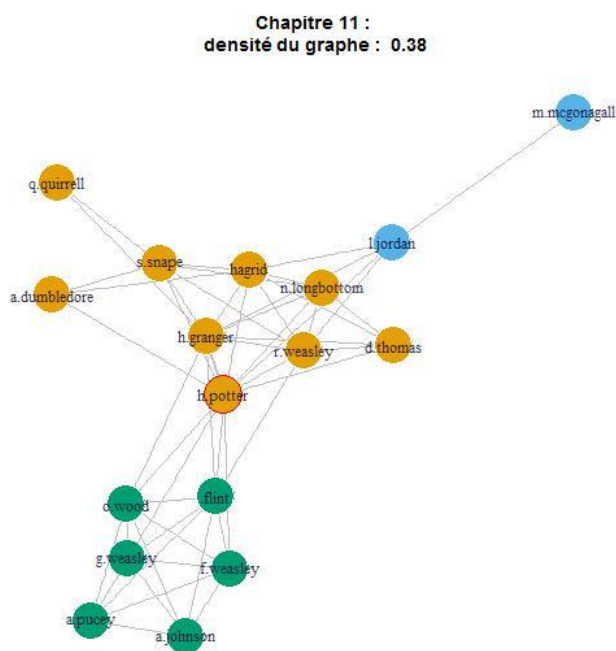
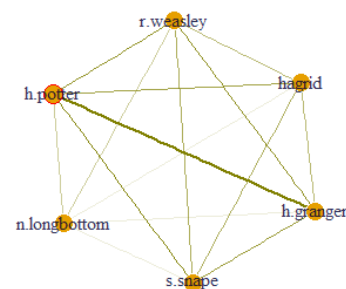
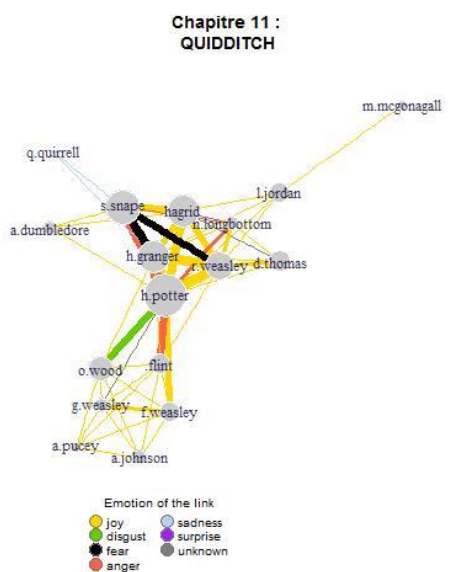


Chapitre 10 :
HALLOWEEN



11. QUIDDITCH

Harry aperçoit que Severus Snape est blessé à la jambe, celui-ci provoque peur et colère auprès du groupe d'amis. Harry joue son premier match de Quidditch contre Serpentard durant lequel, Hermione et Ron pense que Severus a ensorcelé son balai. Hermione met feu à la tribune de Quirell et Severus pour arranger la chose. Oliver Wood est blessé, mais Harry parvient néanmoins à attraper le Vif d'or et fait remporter le match à son équipe (communauté verte). Lee Jordan joue durant ce match un fort rôle de relai d'information dans le réseau, en effet celle-ci commente le match. Enfin nous remarquons une grosse clique de 6 personnes.



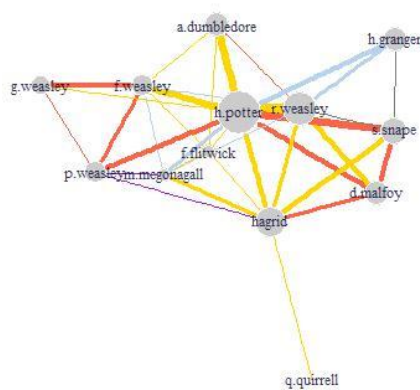
	nombre de degré	centralité d'inter
a.pucey	5	0.53
a.dumbledore	3	5.1
a.johnson	5	0.53
d.thomas	5	4.54
f.weasley	6	0
g.weasley	6	12.33
hagrid	8	5.4
h.potter	12	17.52
h.granger	10	38.81
l.jordan	5	23.38
flint	8	19.45
m.mcgonagall	1	0
n.longbottom	6	8.39
o.wood	7	11.45
q.quirrell	2	3.34
r.weasley	8	3.26
s.snape	7	3.02



12. THE MIRROR OF ERASED

Nos trois amis commencent leurs recherches sur Flamel dans la bibliothèque, l'école se vide pour les vacances de Noël. Ron (forte centralité d'intermédiarité) apprend à Harry les échecs version sorcier, Harry reçoit la cape d'invisibilité de son père comme cadeau de Noël, l'essaye durant la nuit et découvre le Miroir de Risé, y voit sa famille et est mis en garde par Dumbledore. Nous notons 3 communautés l'un e composé d'Hermione, Harry, Ron et Dumbledore, une autre comptant Hagrid (forte centralité d'intermédiarité dans ce groupe), Malfoy, Severus et Quirrell. Enfin un dernier composé des autres frères Weasley, Flitwick et McGonagall (personnages secondaires à ce chapitre).

Chapitre 12 :
THE.MIRROR.OF.ERISED

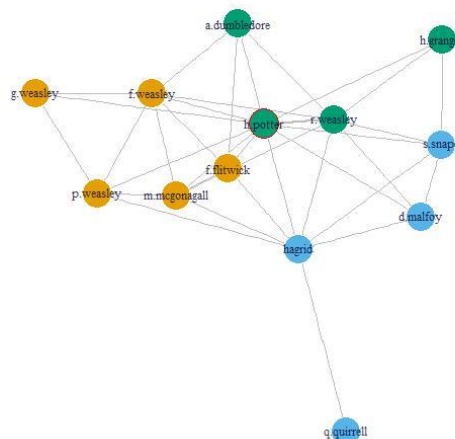


Emotion of the link

- joy
- disgust
- fear
- anger
- sadness
- surprise
- unknown

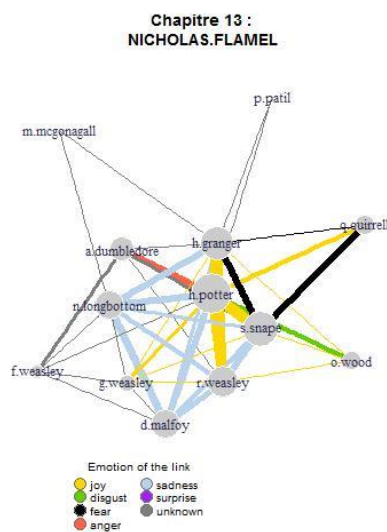
	nombre de degré	centralité d'inter
a.dumbledore	4	1.4
d.malfoy	4	0
f.flitwick	5	13.75
f.weasley	7	2.14
g.weasley	3	1
hagrid	8	15.95
h.potter	11	3.89
h.granger	3	0.2
m.mcgonagall	6	8.7
p.weasley	5	6.92
q.quirrell	1	0
r.weasley	8	15.77
s.snape	5	5.26

Chapitre 12 :
densité du graphe : 0.45

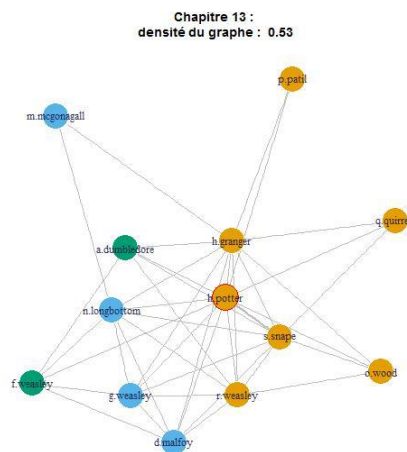


13. NICHOLAS FLAMEL

Gryffondor gagne le match contre Poufsouffle lorsqu'Harry attrape le Vif d'or après seulement cinq minutes (forte intermédierité de plusieurs membres de l'équipe). Harry suit ensuite Rogue dans la Forêt interdite où ce dernier rencontre Quirrell (peur envers Hermione, Severus Snape et Quirrell). Nos trois héros en savent plus sur Nicholas Flamel l'inventeur de la Pierre Philosophale. Une clique de 7 personnages apparait aussi dans ce réseau : Snape, Potter, Gringer, Ron et Georges, Neuville et Drago.

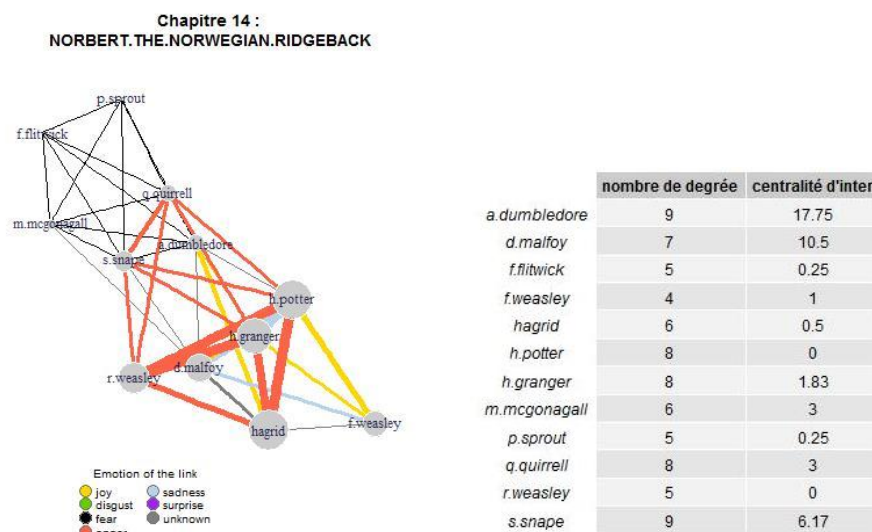


	nombre de degré	centralité d'inter
a.dumbledore	5	1.25
d.malfoy	7	0
f.weasley	5	6.42
g.weasley	7	20.52
h.potter	11	1.67
h.granger	11	27.15
m.mcgonagall	2	1.78
n.longbottom	8	3.25
o.wood	4	3.5
p.patil	2	3
q.quirrell	3	0
r.weasley	8	2.62
s.snape	9	1.32

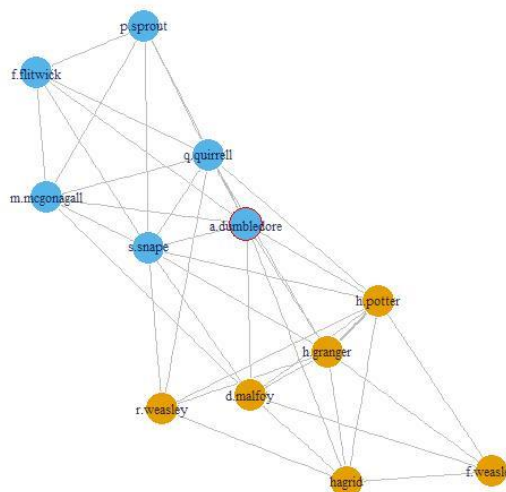


14. NORBERT THE NORWEGIAN RIDGEBACK

Nos trois amis rendent visite à Hagrid, découvrent qu'il possède un œuf de dragon et plus tard le regardent éclore. Le dragon grandissant, il devient de plus en plus difficile pour Hagrid de dissimuler cet animal de compagnie illégal, Dumbledore (personnage étant le plus connecté dur réseaux) demande à ce que celui-ci soit renvoyé en Roumanie. Sur le graphique nous avons deux communauté d'un côté les enseignants (clique de 6 personnages) de l'autre Hagrid et les élèves.

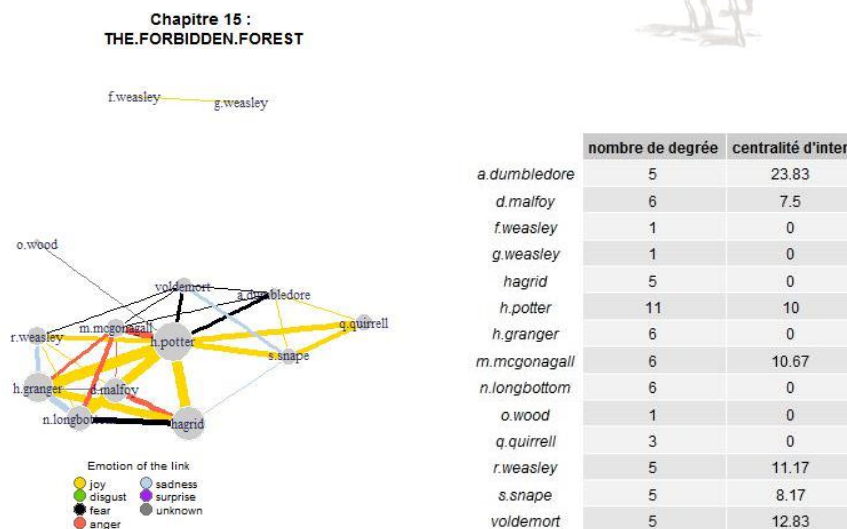


Chapitre 14 :
densité du graphe : 0.61

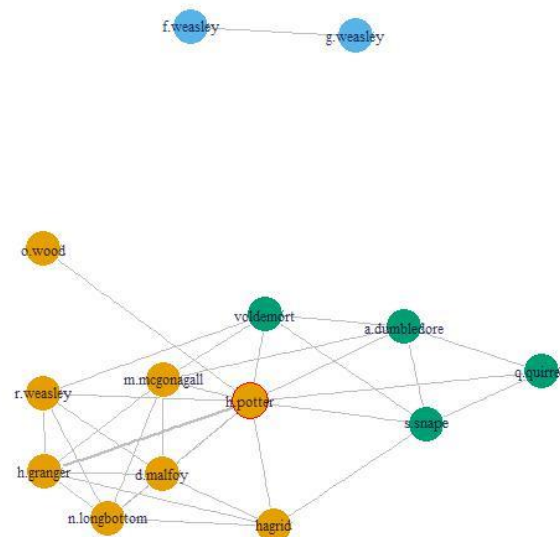


1.5. THE FORBIDDEN FOREST

Drago (colère de celui-ci) et Harry ont une retenue à effectuer avec Hagrid, qui les emmène dans la forêt interdite à la recherche d'une licorne blessée qu'ils trouvent morte. Harry rencontre alors Voldemort mais celui-ci s'enfuit à l'arrivée d'un centaure. Ron, Hermione et Neville sont liés par une certaine tristesse. Nous remarquons une clique à part entre Georges et Fred les deux frères et une de 6 élèves dans la communauté jaune. Nous remarquons encore une forte centralité d'intermédiarité autour de Dumbledore, Voldemort et McGonagall.

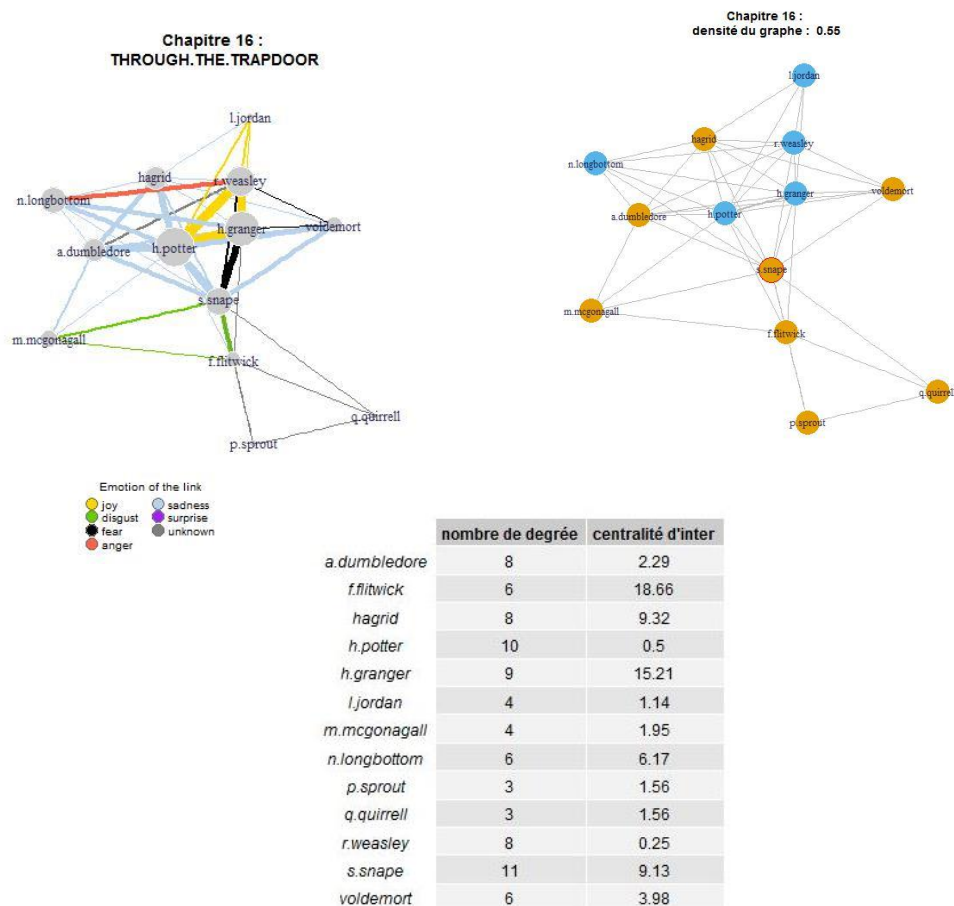


**Chapitre 15 :
densité du graphe : 0.36**



16. THROUGH THE TRAPDOOR

Hagrid leur apprend par mégarde comment contrôler Touffu, Nos trois amis tentent d'avertir Dumbledore (émergence d'une tristesse à son égard) qui est malheureusement absent. Craignant que Snape (fort peur de la peur d'Hermione et Ron en son rencontre ce qui fait de Snape le personnage le plus connecté de ce graphe) soit sur le point de s'emparer de la Pierre Philosophale, ils d'entrer par la trappe. Ils devront affronter plusieurs épreuves, dont de dangereuses plantes, un échiquier géant et des potions. Un fort dégoût apparait entre les professeurs Snape, MacGonagall et Flitwick. Nous observons une séparation nette entre deux communautés d'élèves et de professeurs. Nous observons enfin une clique d'intérêt reliant les 7 personnages principaux de ce chapitre entre Dumbledore, Snape, Harry, Ron, Hermione, Neville et Hagrid .

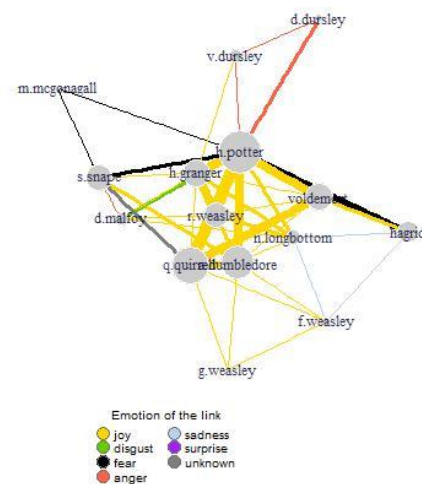


17. THE MAN WITH TWO FACES

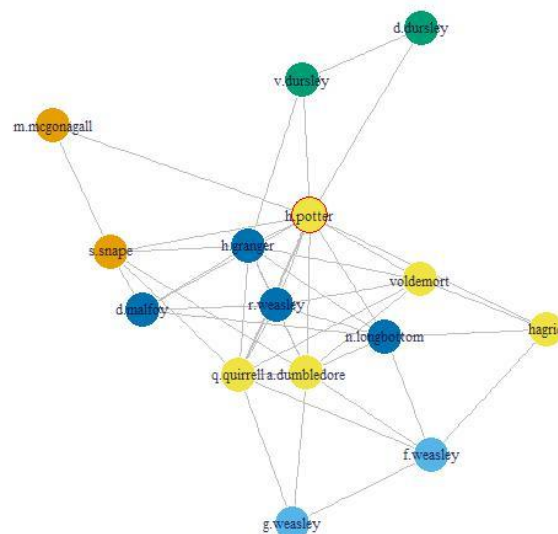
Harry se retrouve seul avec Quirrell qui tente de voler la Pierre pour Voldemort, et non pas Snape. Quirrell tente de faire usage du Miroir de Risé pour trouver la Pierre qui apparaît finalement dans la poche d'Harry. Voldemort, qui "fait partie" de Quirrell, s'en rend compte et ordonne à Quirrell de tuer Harry (peur envers Voldemort et Snape), qui s'évanouit durant la lutte. Il se réveille à l'infirmerie et y retrouve Dumbledore. Gryffondor gagne alors la Coupe des Quatre Maisons (dégout de Malfoy), durant le banquet de fin d'année, grâce au courage de nos trois héros. Harry rentre alors chez les Dursley pour les vacances (communauté verte et colère d'Harry).

Chapitre 17 :
THE.MAN.WITH.TWO.FACES

	nombre de degré	centralité d'inter
a.dumbledore	9	10.79
d.malfoy	5	9.33
d.dursley	2	0
f.weasley	5	5.9
g.weasley	3	0.33
hagrid	4	0
h.potter	12	4.21
h.granger	9	30.42
m.mcgonagall	2	0.5
n.longbottom	7	11.13
q.quirrell	8	9.07
r.weasley	7	1.7
s.snape	6	8.83
v.dursley	3	15.67
voldemort	6	2.18

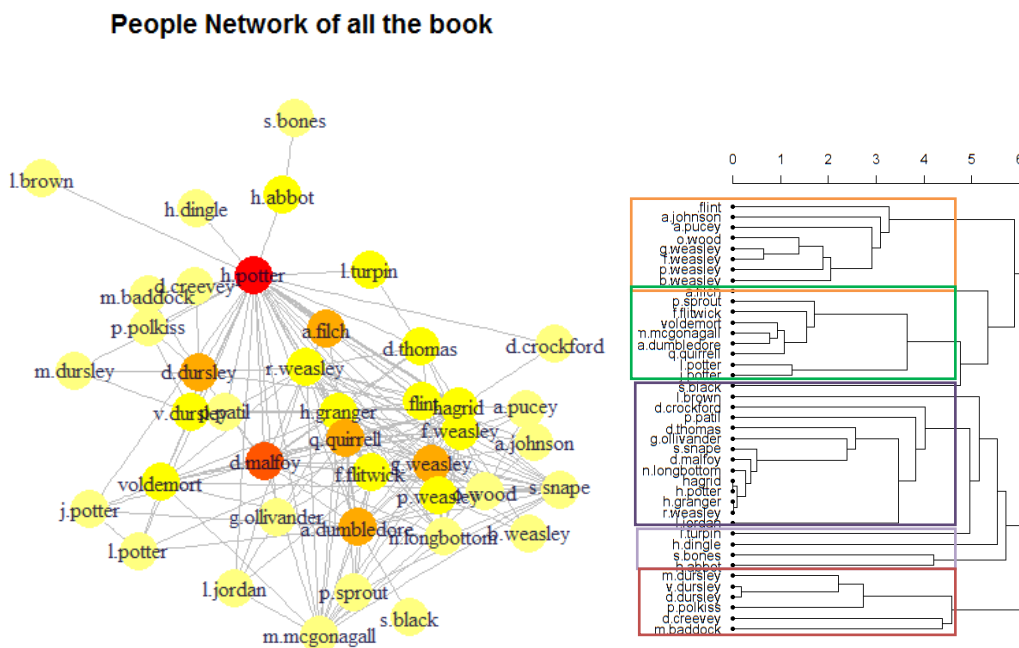


Chapitre 17 :
densité du graphe : 0.42

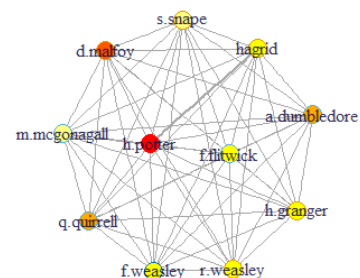


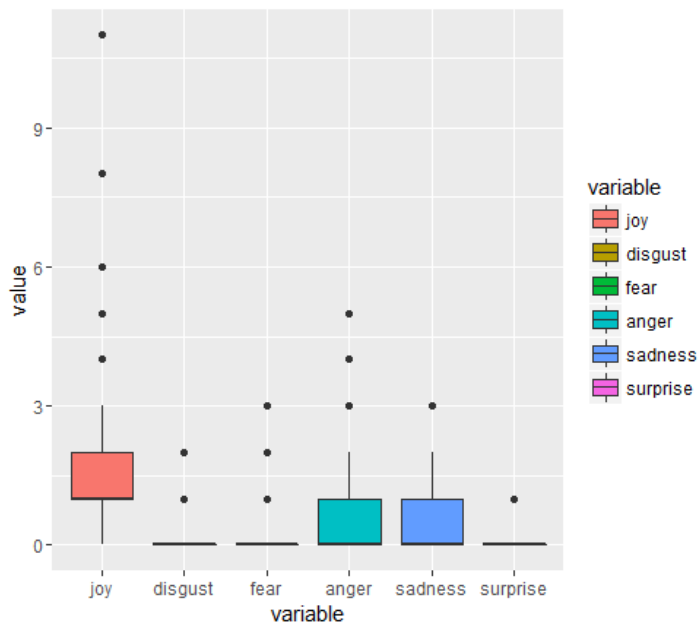
IV. OBSERVATIONS FINALES

Afin de clôturer notre analyse nous observons le livre sous sa forme entière en traçant le graphe des relations nous colorons en les nœuds sur une échelle de couleur entre rouge et jaune fade en fonction de leur centralités, nous observons qu'Harry en personnage central, Malfoy occupe un rôle relais dans ce graphe, tout comme Dudley, Filch, Dumbledore, Quirrell, et Georges Weasley. Nous repérons 5 communautés, les Weasleys, Les professeurs, les parents d'Harry, et Voldemort, les élèves de poudlard, ainsi que des Dursley et les amis de Dudley.



Nous observons de plus une clique de 11 personnages, dont Harry, Draco, Hermione et Ron, Hagrid, Dumbledore, Severus et McGonagall.





L'on que l'on observe les boxplot des relations par humeurs nous remarquons 10% de personnages ayant plus de trois chapitre où leurs relations sont joyeuses, 10% ont au moins deux relation de dégout, 10% ont entre 2 et 3 relations de peur. 10% ont au moins trois relations contre lesquelles elles sont en colère dans l'histoire, et 10% ont 3 relations de tristesse et 10% ont au moins 1 relation associée à la surprise.

Ci-dessous les relations fortement marqué par un sentiment :

JOY > 3



- △ "a.dumbledore & **h.potter**"
- △ "d.malfoy & **h.potter**"
- △ "d.malfoy & r.weasley"
- △ "d.dursley & **h.potter**",
- △ "f.weasley & g.weasley"
- △ "f.weasley & **h.potter**"
- △ "hagrid & **h.potter**"

- △ "**h.potter** & q.quirrell"
- △ "**h.potter** & r.weasley"
- △ "h.granger & n.longbottom"
- △ "h.granger & r.weasley"
- △ "n.longbottom & r.weasley"
- △ "**h.potter** & h.granger"

DISGUT > =1



- △ "f.flitwick & m.mcgonagall"
- △ "f.flitwick & s.snape"
- △ "d.thomas & h.potter"
- △ "d.thomas & l.turpin"
- △ "d.malfoy & **h.potter**"
- △ "d.dursley & **h.potter**"
- △ "d.dursley & v.dursley"

- △ "f.weasley & o.wood"
- △ "**h.potter** & l.turpin"
- △ "**h.potter** & o.wood"
- △ "h.granger & m.mcgonagall"
- △ "m.mcgonagall & s.snape"



FEAR >1

- △ "a.dumbledore & h.potter"
- △ "a.dumbledore & m.mcgonagall"
- △ "a.dumbledore & q.quirrell"
- △ "a.dumbledore & voldemort"
- △ "hagrid & n.longbottom"
- △ "h.potter & m.mcgonagall"
- △ "h.potter & voldemort"
- △ "h.granger & s.snape"
- △ "m.mcgonagall & s.snape"
- △ "m.mcgonagall & voldemort"
- △ "r.weasley & s.snape"
- △ "r.weasley & voldemort"



ANGER >2

- △ "a.dumbledore & h.potter"
- △ "d.malfoy & h.potter"
- △ "d.malfoy & s.snape"
- △ "d.dursley & h.potter"
- △ "d.dursley & v.dursley"
- △ "hagrid & h.potter"
- △ "h.potter & q.quirrell"
- △ "h.potter & s.snape"
- △ "h.potter & v.dursley"



SURPRISE >0

- △ "a.dumbledore & h.potter"
- △ "d.thomas & n.longbottom"
- △ "d.malfoy & n.longbottom"
- △ "hagrid & p.weasley"
- △ "h.potter & h.granger"
- △ "m.mcgonagall & p.weasley"
- △ "n.longbottom & r.weasley"



SADNESS>2

- △ "a.dumbledore & hagrid"
- △ "a.dumbledore & h.potter"
- △ "d.malfoy & h.granger"
- △ "h.potter & n.longbottom"
- △ "h.granger & n.longbottom"
- △ "h.granger & r.weasley"



V. CONCLUSIONS



Ce projet nous aura permis d'approcher de nouveaux outils d'analyse textuelle : tagger, de tester le modèle lda, de toucher du doigt une technique d'analyse d'opinion et de nous mettre à l'analyse de graphe.

L'analyse se base sur un script qu'il serait facile à déployer sur les autres ouvrages, et permet de réaliser une cartographie des relations des personnages, de leurs relation et dans quels contextes ceux-ci interagissent, il serait judicieux de garder les scores fournis par notre classifieur bayésien afin d'en évoluer l'évolution durant les 7 tomes.

Pourrait aussi être menée une analyse avec un contexte grammaticale afin de connaitre qui est tributaire de l'émotion. Un graphe dirigé du réseau pourrait être dessiné.



VI. SITO-GRAPHIE & REFERENCES



Tutoriels en ligne :

- △ <http://www.rdatamining.com/examples/social-network-analysis>
- △ <http://www.r-bloggers.com/summary-of-community-detection-algorithms-in-igraph-0-6/>
- △ <http://kateto.net/network-visualization>
- △ <http://stackoverflow.com/questions/29918017/how-to-use-sentiment-package-in-r-3-2-0>
- △ <http://francescopochetti.com/community-detection-social-networks/>
- △ <https://elementr.hypotheses.org/tag/igraph>
- △ http://edutechwiki.unige.ch/fr/Analyse_de_sentiments_en_text_mining#M.C3.A9thode_non-supervis.C3.A9e:_Information_ponctuelle_mutuelle

Supports de cours :

- △ Cours de Nicolas Carayol (*Analyse et Modélisation des Réseaux* | Université Bordeaux IV)
- △ Cours d'introduction à l'analyse des réseaux de Dominique BEYNIER
- △ Cours de Rémi Bachelet (*Réseaux sociaux* | Centrale Lille)
- △ Cours de CHLOÉ CLAVEL (*Analyse des opinions sur le web et Opinion Mining* | TELECOM-PARISTECH)
- △ Cours de Xavier Tannier (*Extraction d'Information et Structure de Documents : Reconnaissance des Entités Nommées* | Université Paris Sud)

Publications scientifiques :

- △ Yayoi Nakamura-Delloye, Rosa Stern. *Extraction de relations et de patrons de relations entre entités nommées en vue de l'enrichissement d'une ontologie.*
- △ David Combe. *Détection de communautés dans les réseaux d'information utilisant liens et attributs. Intelligence artificielle.*

Crédit image :

- △ Illustrations by Mary Grandpré (<http://www.pojo.com/harrypotter/books>)
- △ Icons design by Polina Flegontovna

