

THESIS / THÈSE

MASTER EN INGÉNIEUR DE GESTION À FINALITÉ SPÉCIALISÉE EN ANALYTICS & DIGITAL BUSINESS

Les techniques de process mining dans l'analyse des réseaux sociaux

« La communication sur Twitter dans la promotions de films suit-elle des processus identifiables ? »

GRAILLY, Manon

Award date:
2022

Awarding institution:
Universite de Namur

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.



Les techniques de process mining dans l'analyse des réseaux sociaux :
« La communication sur Twitter dans la promotions de films suit-elle des processus
identifiables ? »

Manon GRAILLY

Directeur: Prof. I. LINDEN

Mémoire présenté
en vue de l'obtention du titre de
Master 120 en ingénieur de gestion, à finalité spécialisée
en Analytics & Digital Business

ANNEE ACADEMIQUE 2021-2022

Avant-Propos

Ce mémoire représente l'aboutissement de mon parcours universitaire à Namur. Celui-ci m'a permis de me surpasser et d'élargir mes compétences dans le monde de l'I.T.. Je tiens à remercier toutes les personnes de près ou de loin qui m'ont aidé à l'élaboration de ce travail.

Un remerciement tout particulier va à ma promotrice, Isabelle Linden, pour son accompagnement et sa patience durant toute la réalisation de ce mémoire.

Je souhaiterais également remercier Yannis Bertrand, doctorant en process mining, pour ses précieux conseils, sa disponibilité et son temps. Son aide contribue à la réussite de ce travail.

Enfin, je tiens à remercier de tout cœur ma famille et mes amis qui n'ont cessé de me rappeler l'importance de persévérer et leur soutien durant ce travail mais également durant mes années universitaires. J'aimerais remercier plus spécialement Clémence Griffé pour les relectures entreprises lors de la rédaction de ce mémoire.

Résumé

L'accroissement des réseaux sociaux génère un nombre considérable de données. Les utilisateurs de plus en plus actifs sur les réseaux expriment leur opinion au reste du monde sur différents sujets. De plus, ces données sont facilement exploitables et accessibles. En parallèle, durant les deux dernières décennies, le process mining n'a cessé de se développer. Ces techniques se concentrent sur l'analyse des processus et créent de nombreuses possibilités. L'analyse des réseaux sociaux en utilisant les techniques du process mining est une recherche récente qui est, pour le moment, peu abordée dans le monde scientifique. Cela crée une magnifique opportunité de recherche et d'analyse. Ce mémoire proposera une méthodologie en prenant comme étude de cas l'industrie du cinéma francophone.

Abstract

The growth of social networks generates huge amounts of data. Users who are increasingly active on networks express their opinions on various topics to the rest of the world. Moreover, data is easily exploitable and accessible. At the same time process mining has continued to develop during these past two decades. These techniques focus on process analysis, offering many possibilities. Analysis of social networks using process mining techniques is a recent research that is, for the time being, little addressed in the scientific world. This creates great opportunity for research and analysis. This thesis will propose a methodology using the French-speaking film industry as a case study.

Table des matières

Chapitre I. Introduction	1
1.1. Introduction	1
1.2. Problématique.....	2
Chapitre II. Revue de la littérature	3
2.1. Process mining	3
2.2. Réseaux sociaux	4
2.2.1. Le cinéma et les réseaux sociaux.....	4
2.3. Le process mining et les réseaux sociaux.....	4
Chapitre III. Méthodologie.....	6
Chapitre IV. Préparation des données	8
4.1. Extraction et pré-filtrage des données	8
Chapitre V. Analyse	9
5.1. Analyse du comportement de la communauté.....	9
5.1.1. Analyse descriptive	9
5.1.2. Construction des event log	11
5.1.3. Processus par film.....	14
5.1.4. Généralisation.....	20
Chapitre VI. Conclusion.....	23
6.1. Apports méthodologique et analytiques	23
6.1.1. Apport méthodologique.....	23
6.1.2. Apport analytiques.....	23
6.2. Limites du travail.....	24
6.3. Pistes travaux futures.....	24
Références.....	26
Annexes	28
Annexe 1 – code R	28
1.1. Boite Noire	28
1.2. Illusions perdues.....	29
1.3. Benedetta	29
1.4. L'évènement.....	29
1.5. Les olympiades.....	29
1.6. Titane.....	30
Annexe 2 – Graphiques - Autres films	31
2.1. Illusions Perdues.....	31

2.2.	Benedetta	32
2.3.	L'évènement	32
2.4.	Les Olympiades	33
2.5.	Titane	34
Annexe 3 – Graphiques tweets en fonction des catégories		35
3.1.	Illusions perdues	35
3.1.	Benedetta	35
3.2.	L'évènement	36
3.3.	Les Olympiades	36
3.4.	Titane	37
Annexe 4 – Les différentes catégories pour les cases_id dans le cadre du film Boite noire		38
Annexe 5 – Code PM4Py		39
Annexe 6 – Processus Boite Noire		43
6.1.	Processus parties prenantes Boite noire	43
6.2.	Tableau des variants	44
6.3.	Processus utilisateurs	45
Annexe 7 – Processus média		46
7.1.	Illusions perdues	46
7.2.	Benedetta	47
7.3.	L'évènement	48
7.4.	Les olympiades	49
7.5.	Titane	50

Chapitre I. Introduction

1.1. Introduction

Les réseaux sociaux sont une source de données considérables. Aujourd'hui, 4,5 milliards de personnes utilisent un réseau social (Facebook, Instagram, Twitter, ..), ce qui représente environ 54% de la population mondiale¹. Tous ces utilisateurs génèrent de plus en plus de données. Nous pouvons observer l'ampleur de ce phénomène sur la figure 1.1. à droite.

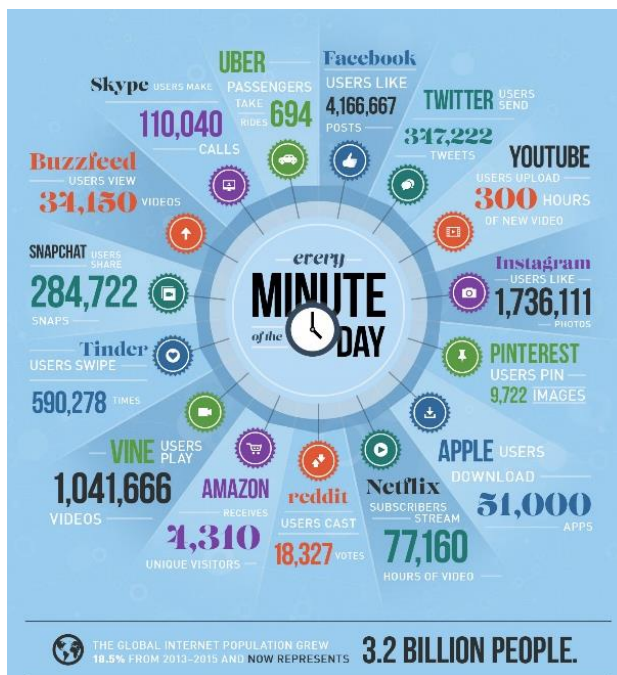


Figure 1.1. <https://toolkit.loomly.com/one-minute-of-social-networks/>

Au vu du succès des réseaux et de l'accessibilité des données de ceux-ci, de nombreuses recherches et analyses sur les réseaux sociaux sur ont vu le jour. Ces

travaux montrent différentes techniques d'analyse comme, par exemple, le data mining qui permet d'analyser les données sous différentes perspectives et de les transformer en informations utiles en établissant des relations entre elles.

En parallèle, depuis quelques années, de nouvelles techniques d'analyse sont apparus, notamment, le process mining. Cette approche propose *un ensemble de techniques largement utilisées pour extraire des informations à partir d'event log générés par des système d'information* (Li, Murillas, Carvalho, & Van Der Aalst, 2018). Cette technique se concentre sur l'analyse des processus.

Ce travail explore le potentiel de l'utilisation des techniques de process mining en vue d'analyser des données extraites du réseau social Twitter. Dans un premier temps, une problématique cible de ce travail est posée : la communication autour des films dans le cinéma français afin de comprendre la base et l'enjeu du sujet. Ensuite, la revue de la

¹ <https://fr.oberlo.be/blog/chiffres-reseaux-sociaux#:~:text=Aujourd'hui%2C%20la%20France%20compte,de%2076%20%25%20de%20la%20population.>

littérature présentera le contexte de manière détaillée. Le chapitre III sera consacré à l'explication de la méthodologie utilisée pour ce travail. Il sera suivi d'une analyse descriptive des données afin d'obtenir une vision plus globale de celles-ci. Après une compréhension plus approfondie des données, ce travail tentera de trouver des processus pertinents, de les analyser et de les comparer en utilisant diverses techniques de process mining. Enfin, le chapitre VI consistera en la conclusion générale et inclura les apports du travail, ses limites et ses perspectives futures.

1.2. Problématique

Le travail se centrera sur l'analyse du cinéma français sur Twitter.

Comme énoncé précédemment, Le taux d'utilisateurs de réseaux sociaux croît au quotidien et s'accompagne d'une évolution des possibilités liées à ceux-ci. Ceux-ci sont notamment largement exploités par les entreprises qui s'en servent comme d'un support marketing pour réaliser la promotion de leurs biens et services.

La contribution de ce travail vise donc à répondre à la question de recherche suivante :

« La communication sur Twitter dans la promotions de films suit-elle des processus identifiables ? »

L'étude analyse les films Boite noire, Benedetta, Illusions Perdues, Les Olympiades, L'évènement et Titane.

Certaines étapes dans la promotion d'un film sont incontournables, le dévoilement de l'affiche, de la bande annonce, ... Est-il possible de retrouver ces étapes dans le contenu des utilisateurs quant ils parlent d'un film ?

Ce mémoire a pour ambition d'apporter une nouvelle manière d'analyser les comportements sur les réseaux sociaux.

Chapitre II. Revue de la littérature

Ce chapitre commencera par expliciter les bases du process mining : les différentes perspectives, les algorithmes utilisés ainsi que les outils.

Ensuite, cette revue de la littérature abordera les recherches sur les analyses des réseaux sociaux avant de faire le lien avec l'industrie du cinéma.

Pour finir, cette recherche passera en revue les travaux combinant le process mining et les réseaux sociaux.

2.1. Process mining

Le process mining n'a cessé de se développer durant ces deux dernières décennies. Ce domaine de recherche se concentre sur l'analyse des processus (Van Der Aalst, 2012). Les techniques du process mining permettent de « *découvrir, surveiller, et améliorer les processus dans une variété de domaines d'application* » (Aalst, et al., 2011). Dans la pratique, il construit des processus à l'aide de données appelées « event log ».

Plusieurs perspectives peuvent être distinguées. D'après Dustdar et al., il y en a 3 : la perspective processus, la perspective organisationnel et la perspective des cas. Ce mémoire se penchera sur la première. Le but étant de trouver une représentation de tous les chemins possible au sein du processus (Turner, Tiwari, Olaiya, & Xu, 2012).

Pour réaliser les processus, plusieurs outils sont disponibles. Chaque outil utilise l'un ou l'autre algorithme. Cela signifie que le résultat peut être différent en fonction de l'algorithme utilisé.

L'algorithme « Inductive Miner » est une approche pour découvrir les arbres de processus des event log. L'avantage de cet algorithme est qu'il fournit toujours un processus (Ghawi, 2016). Cet algorithme peut être retrouvé dans le logiciel ProM ou encore PM4Py. Cependant les processus de la vie réelle sont moins structurés et forment souvent des processus dit « en spaghetti » (Gupta, 2014). Ces processus sont difficilement compréhensibles, ne faisant pas la distinction entre ce qui est important ou non. L'algorithme « Fuzzy Miner » permet de travailler avec ce genre de processus et une quantité importante de données (Gupta, 2014). L'outil Disco travaille avec cet algorithme. Ce logiciel a été conçu pour être un outil rapide et facile d'utilisation.

2.2. Réseaux sociaux

La recherche dans le domaine des réseaux sociaux n'est pas nouvelle. Il existe des techniques de data mining et de nombreux travaux et méthodes ont déjà vu le jour. Dans l'ouvrage de référence G. Barbier et H. Liu expliquent les bases du data mining dans l'analyse des réseaux sociaux. (Barbier & Liu, 2011).

2.2.1. Le cinéma et les réseaux sociaux

Des articles plus spécifiques au marketing peuvent aussi être trouvés. Les réseaux sociaux sont très importants pour les entreprises : Avoir une communauté active augmente le taux de consommation de leurs produits et de leurs services. L'industrie du cinéma n'y échappe pas. Les firmes utilisent les réseaux sociaux pour se connecter avec leur clientèle (Oh, Roumani, Nwankpa, & Hu, 2017). Les réseaux sociaux et plus principalement twitter sont les « bouches à oreilles en ligne ». Un lien significatif a été trouvé entre le fait que ce soit une personne ayant un nombre de followers conséquent et les ventes en salle d'un film. De plus, l'effet augmente si l'auteur exprime son intention d'achat, son intention d'aller voir le film. Dans leur article, Huaxia Rui, Yizao Liu et Andrew Whinston proposent une sensitive analysis, les tweets sont divisés en 4 catégories: intentionnel, positif, négatif et neutre (Rui, Liu, & Whinston, 2013).

2.3. Le process mining et les réseaux sociaux

Lorsque le process mining fait référence aux communautés, c'est généralement dans le but de construire un réseau à partir d'event log (Van der Aalst & Song, 2004) (Li, Murillas, Carvalho, & Van Der Aalst, 2018). En effet, les event log regorgent d'informations. Au sujet, par exemple, de l'individu qui effectue une tâche précise. Il est donc possible de créer un réseau grâce aux relations que les personnes ont en fonction de la tâche qu'ils effectuent. C'est ce que démontre Wil M.P. van der Aalst et Minseok Song dans leur article (Van der Aalst & Song, 2004).

Un article de 2019 de Guangming Lo et Renata Medeiros De Carvalho tente d'utiliser les techniques de process mining pour analyser un réseau social (Li & De Carvalho, Process

mining in social media: applying object-centric behavioral constraint models, 2019). Les auteurs tentent d'analyser le comportement des consommateurs sur le réseau *Stack exchange*. Pour ce faire, ils considèrent les données comme des données orientées objet, c'est-à-dire des base de données qui représentent un ensemble de données très proche où chaque donnée est représentée par classe. En effet, une notion de relation *one to many* et *many to many* est présente dans le cas des réseaux sociaux. Dans le contexte de Twitter, un utilisateur peut écrire plusieurs tweets. Par contre, un tweet ne fait référence qu'à un seul utilisateur.

Une autre approche qui combine les réseaux sociaux et le process mining est développée par C. Diamantini, L. Ganga, F. Marozzo, D. Potena et P. Trunfio (Diamantini, Genga, Marozzo, Potena, & Trunfio, 2017). Cet article ne considère pas les données des réseaux sociaux comme étant objet-centrée.

Chapitre III. Méthodologie

Comme énoncé dans le premier chapitre, l'ambition du mémoire est d'analyser le comportement des utilisateurs sur les réseaux sociaux, plus précisément Twitter, à l'aide de techniques de process mining. Cette recherche s'appuie sur la méthodologie développée dans l'article de C. Diamantini, L. Ganga, F. Marozzo, D. Potena et P. Trunfio (2017)

La première partie de ce travail concerne l'extraction des données. A l'aide de l'API (Application Programming Interface, soit l'interface de programmation) de Twitter, il est possible d'extraire certaines informations du réseau social, y compris les tweets. Afin de récupérer un maximum des tweets pertinents, des critères de sélection sont appliqués : les hashtags utilisés, la langue et la période durant laquelle les tweets sont publiés. Ensuite, un filtrage des tweets est réalisé dans le but de ne conserver que les tweets pertinents, c'est-à-dire, les tweets dont le contenu est en lien avec le sujet. Une fois que les données sont propres, l'étape suivante peut commencer ; une première analyse descriptive est nécessaire pour se familiariser avec les données et en avoir un premier aperçu. Cette analyse est réalisée grâce au logiciel R Studio. Une bonne compréhension des données est nécessaire afin de faciliter les analyses ultérieures.

Après avoir acquis une bonne connaissance des données, la transformation de celles-ci en event log est réalisable. Plusieurs éléments doivent se retrouver dans celui-ci: une activité, un timestamp et une instance de processus (case_id). La manière de les construire influence fortement sur le modèle final, c'est une partie cruciale de la méthodologie.

Afin d'obtenir une activité, les tweets sont catégorisés en fonction de leur contenu. Pour se faire, une connaissance du sujet est nécessaire. Il s'agit de trouver des catégories pertinentes en fonction du sujet analysé. Une fois les catégories déterminées, les tweets peuvent être classés dans celles-ci.

Le timestamp représente l'horodatage, c'est-à-dire l'heure à laquelle le tweet ou le retweet a été effectué. L'instance de processus quant à lui fait référence à l'utilisateur qui a publié un ou plusieurs tweets ou retweets. L'utilisateur est identifiable grâce à son ID unique.

Dans un premier temps, une analyse individuelle des films est effectuée. L'enjeu est d'identifier la manière dont les utilisateurs s'expriment à propos de ceux-ci. Dans un second temps, une agrégation de toutes les données de chacun des films est réalisée. Le but est de générer un processus général.

Lorsque les données sont correctement transformé en event log, il est possible de créer des processus. Les outils utilisés sont [PM4Py](#) et le logiciel [Disco](#).

PM4Py est un logiciel créé par *Fraunhofer Institute for Applied Information Technology* (FIT) et plus précisément par le groupe de recherche de Wil van der Aalst. Le but de ce logiciel est de mettre à disposition les techniques du process mining à un large public. Il consiste à un package Python. Ce travail l'utilisera sur l'environnement Jupyter. PM4Py sera utilisé pour réaliser les premières analyses sur les films. Comme par exemple, pour identifier les premières et dernières activités des internautes pour chacun des films. L'algorithme utilisé est l'algorithme «inductive miner». Disco, quant à lui, est un logiciel qui permet la création de processus et de visuels de manière rapide. La prise en main est facile et intuitive. C'est avec ce logiciel que les processus seront générés et analysés plus en détails. Il utilise l'algorithme « fuzzy miner ».

Afin d'avoir d'obtenir des processus plus précis et compréhensible, il est possible d'effectuer du trace clustering. « *Les event log sont divisés en sous partie homogènes et un processus est créé pour chaque sous ensemble* », (Song, Günther, & Van der Aalst, 2008).

Chapitre IV. Préparation des données

4.1. Extraction et pré-filtrage des données

De nombreux films sortent chaque mois dans les salles de cinéma et génèrent de nombreuses réactions sur la plateforme Twitter. Afin d'obtenir une quantité de données raisonnable et plus facilement exploitable, certains critères ont été sélectionnés dans le choix des films analysés. Ceux-ci proviennent de l'industrie du cinéma français et sont sortis en salle entre le 1^{er} juillet 2021 et le 30 novembre 2021. Pour garantir une comparaison pertinente, tous les films sont des drames ou thrillers. De plus, tous les tweets récoltés sont uniquement en français.

Pour chaque tweet, les informations suivantes ont été récoltées : son contenu, la date et l'heure de création, les public metrics (nombre de likes, réponses et de retweets (RT)), l'id du tweet, l'id de l'auteur du tweet, le nom de l'utilisateur associé à l'id et la description de celui-ci. Les données ont été récoltées entre le début du tournage et le 30 avril 2022.

Les données ont été récolté grâce à l'API de Twitter. Toutes les données sont publiques récoltées, c'est-à-dire que c'est les utilisateurs qui ont librement décidé de partager leur contenu au monde.

Après l'extraction de tous les tweets, un filtrage des données a été effectué.

En guise d'exemple, en ce qui concerne le film *Boite*, les tweets comportant les mots « voiture » ou « top chef » ont été supprimé. En effet, d'une part, une nouvelle réforme en France est en cours de discussion et concerne le fait d'imposer l'installation de boites noires dans les voitures. D'autre part, une épreuve de l'émission *Top chef* qui a eu lieu en mai 2022, était intitulée « Boite Noire ». De plus, les tweets issus du média *Canal +* ont demandé une analyse plus approfondie en amont. En effet, la chaîne médiatique possède une émission qui, malgré le fait qu'elle n'ait aucun rapport avec le film, se nomme *Boîte noire*. Des cas similaires ont pu être observés pour les autres films. Par exemple pour *Illusions perdues*, des tweets au sujet de Jean-Luc Mélenchon ont dû être supprimés. Cette étape de filtrage des données est importante pour s'assurer de conserver uniquement les tweets qui sont pertinents pour le travail.

Chapitre V. Analyse

5.1. Analyse du comportement de la communauté

Pour une meilleure compréhension de l'analyse, le film Boite noire sera le film de référence. Le but de ces analyses est de comprendre comment les utilisateurs s'expriment au sujet d'un film et si il est possible d'en découler des processus.

5.1.1. Analyse descriptive

Le code et les réponses du terminal effectués pour cette analyse est disponible en [annexe 1](#).

Les figures 5.1. et 5.2. ci-dessous montrent l'évolution du nombre de tweets en fonction du temps. Une hausse en septembre 2021 est très nettement observable. Cette augmentation quantitative correspond en réalité à la sortie du film dans les salles de cinéma en France le 8 septembre 2021.

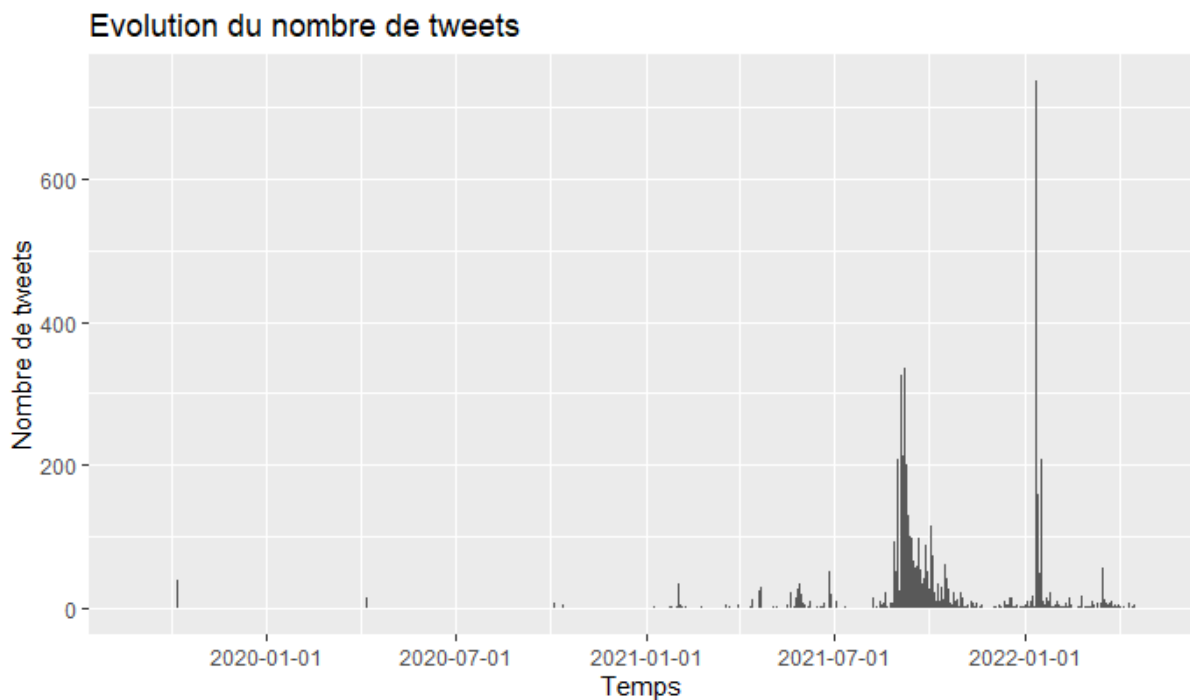


Figure 5.1. Evolution du nombre de tweets par jour du film Boite noire de septembre 2019 à avril 2022

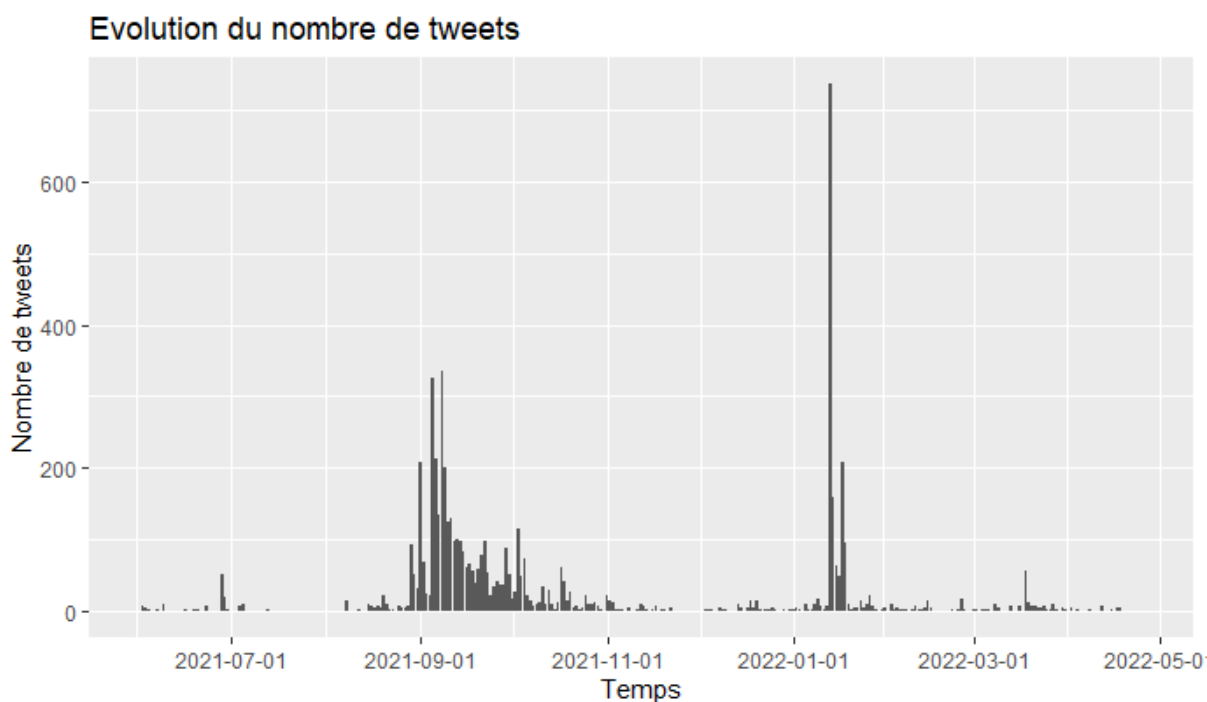


Figure 5.2. Evolution du nombre de tweets par jour du film Boite noire de juin 2021 à avril 2022

Un autre pic est observable début janvier 2022. Il est dû au fait qu'un tweet a généré beaucoup de retweets : « *RT @MehdiOmais: #Concours 3 blu-ray de #BoiteNoire à gagner ! > RT pour tenter ta chance jusqu'au 18 janvier à 20h. > Tirage au sort le 19 janvier* ». Il est normal que ce type de tweets génère un grand nombre d'interactions – ici, précisément, 1287 retweets : il s'agit d'un concours dont la participation se fait sur base des dites interactions. Le phénomène n'est pas isolé puisque le second tweet le plus retweeté au sujet du film est également un concours.

Les autres films suivent le même schéma ; hausse de réactions au moment de la sortie du film ou lors de concours. Cependant, il faut rajouter qu'un engouement est observable pendant la période de Canne (juillet 2021) ou des Césars (février 2022) pour les films nominés (Benedetta et Les Olympiades) ou les films récompensés (Illusions perdues, Titane). En effet, une hausse des tweets peut être observé en février 2022 en ce qui concerne Illusions perdues. Ce film a remporté 7 césars, ce qui a entraîné un certain enthousiasme sur les réseaux sociaux. Tous les graphiques et codes se trouvent en annexe 2 et en annexe 3.

Dans les personnes ayant le plus généré de tweets et de retweets, en ce qui concerne Boite noire, Pierre Niney arrive en première place. Il s'agit, en effet de l'acteur principale; il est donc logique de le retrouver à cette place. C'est un compte indépendant qui se retrouve en

deuxième place : boxofficefr. Ce compte commente les entrées des box office de l'industrie du cinéma français. Dans le reste du classement, nous pouvons apercevoir StudioCanal en 6ième position, qui est la société de distribution du film. Les comptes montrant le plus d'activités sont, dans la plupart des cas, des personnes soit directement impliquées dans la production, la réalisation ou la distribution du film, soit investies plus largement dans le monde du cinéma.

Ce schéma ne trouve pas de régularité dans le cas des autres films. En effet, peu d'acteurs ou de représentants du cinéma possèdent un compte Twitter. Dès lors, les comptes ayant généré le plus de tweets au sujet de ces films sont souvent des amateurs de cinéma. La même personne se retrouvent parfois dans plusieurs tops. Le boxofficefr se retrouve également souvent dans le classement.

5.1.2. Construction des event log

Comme expliqué dans le [chapitre III](#), les event log sont construits comme suit : le case_id est le compte unique de chaque personne ayant tweeté ou retweeté, identifiable avec son id et le timestamp est la date de création du tweet ou du retweet. En ce qui concerne l'activité, il s'agit de la catégorie qui est déterminée dans cette partie.

Avant de classifier les tweets, il est important d'établir clairement les différentes catégories. Le choix a été fait en fonction de différents tweets observables dans la base de données et confirmé par une spécialiste du domaine, une assistante de production. Grâce à cela, 13 catégories ont pu être identifiées.

Ensuite, la classification des tweets a été réalisée grâce à la méthode dictionnaire. En fonction des termes utilisés dans le texte du tweet, la catégorie a été attribuée. Cependant cet outil comporte certaines limites qui seront explicitées dans le [chapitre VI](#). Le tableau 5.1. ci-dessous montre les catégories et mots clefs utilisés pour les identifier.

Catégorie	Mot clefs 1	Mot clef 2	Mot clef 3	Mot clef 4	Mot clef 5
Tournage	Tournage				
Affiche	Affiche				
Bande-annonce	Bande-annonce	Bande-Annnonce			
Avant-première	Avant-première	Avant première	AV	Première mondiale	
Statistique	Entrée				
Promo	Interview	ITW	Emission	Invité	Recevoir

Critique	Critique	Podcast			
Concours	Concours	Gagner			
César	César				
Plateforme	Dispo sur	Streaming			
Support matériel	Dvd	Bluray			
Canne	Palme d'Or	Canne			
Oscar	Oscar				

Tableau 5.1. Mode de classification des tweets en fonction des mots clefs

Les catégories « Tournage », « Avant-première » et « Bande-annonce » regroupent, comme leur nom l'indique, tous les tweets réagissant à une étape précise de la vie d'un film. Ces catégories sont facilement identifiables grâce à l'hypothèse suivante : si une personne veut faire référence à l'un de ces trois sujets, elle utilisera précisément ces mots.

La catégorie « Promo » regroupe tous les tweets faisant référence aux différentes actions que les acteurs et/ou réalisateurs peuvent entreprendre pour promouvoir leur film.

La catégorie « Critique » regroupe tous les tweets faisant référence à une critique dans un article, blog, podcast, site d'opinion, ...

En ce qui concerne la catégorie « Statistiques », elle regroupe tous les tweets indiquant le nombre d'entrées pour le film et son score au Box Office.

La catégorie « Concours » a été conçue pour les concours organisés dans le but d'offrir des produits liés au film (place de cinéma, dvd, ...).

La cérémonie des césars et le festival de Canne constituent les événements les plus prestigieux dans le monde francophone. Les tweets traitant de ces sujet ont donc été regroupés dans une catégorie particulière. A l'international, il a été choisi de considérer uniquement les oscars.

Pour finir, les catégories « Plateforme » et « Support matériel » sont utilisées pour les tweets indiquant respectivement toute information lié au plateforme de streaming sur laquelle le film est disponible et la vente de Dvd ou Bluray du film respectivement.

La catégorie « Affiche » a subi une deuxième catégorisation afin d'éliminer certains tweets non pertinents. Par exemple, l'expression « à l'affiche » pour désigner que le film est toujours en salle ou encore simplement le verbe « afficher » doivent être prise en compte. Un deuxième filtrage a été réalisé et tous les tweets comportant les mots expressions : « à

l’affiche », « s’affiche », « en tête », ou « entrées » ont été retiré de la catégorie. Ces tweets n’ont malheureusement pas pu être recatégoriser.

Sur les 6471 tweets récoltés pour le film Boite noire, 2886 tweets ont pu être catégorisé selon ce mode de fonctionnement, soit 44,59%.

Le tableau 5.2. montre le pourcentage de tweets classés par film.

FILMS	TOTAL	CLASSES	POURCENTAGE
Illusions perdues	5451	2137	39,20%
Benedetta	5840	3228	55,27%
L’évènement	3144	1637	52,06%
Les Olympiades	2081	974	46,80%
Titane	13968	8342	59,72%

Tableau 5.2. Pourcentage de tweets classés par film

Le pourcentage de tweets oscille entre 39,20 et 59,72% en fonction des films comme on peut le voir dans le tableau II.2.

La figure 5.3. ci-dessous montre que c’est la catégorie « Concours » qui a généré le plus de tweets et de retweets, suivie de loin par la catégorie « Statistiques » pour le film Boite noire. Seuls les 6 catégories ayant la plus grande quantité de tweets sont représentées.

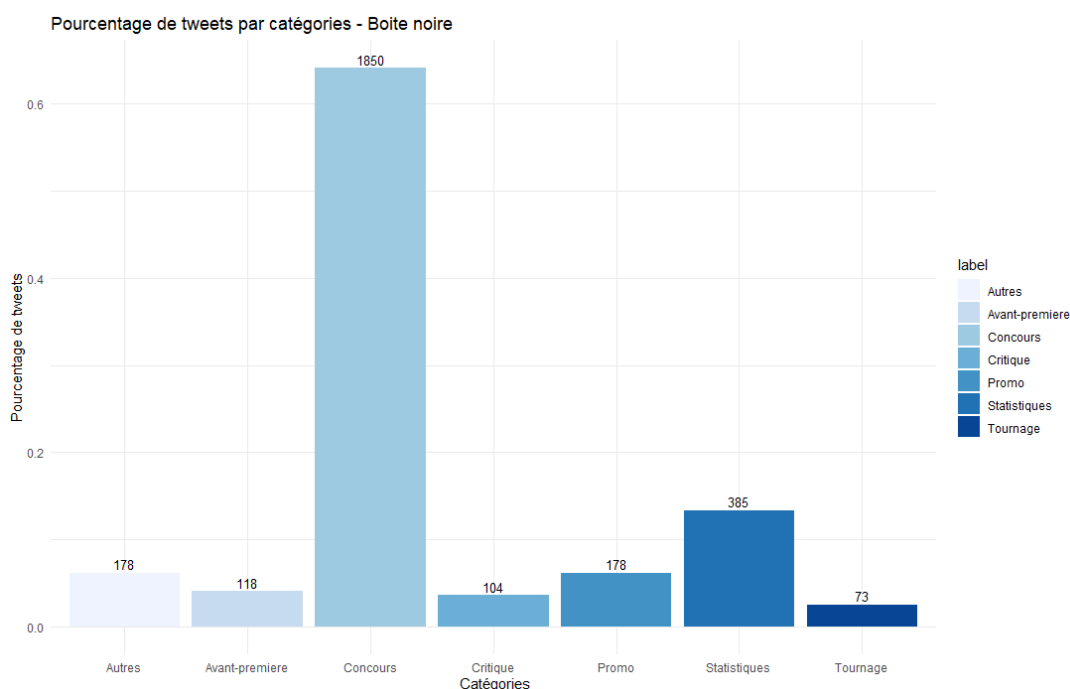


Figure 5.3. Pourcentage de tweets par catégories pour le film Boite noire

Les graphiques des autres films se trouvent en [annexe 3](#). De manière générale, la catégorie contenant la majorité des tweets est soit la catégorie « Concours » (Boite noire, Illusions perdues et L'évènement), soit le festival de Canne (Benedetta, Les olympiades et Titane). Si la catégorie « concours » ne se situe pas en première position, celle-ci est généralement en deuxième. Comme dit précédemment les concours génèrent beaucoup de réactions. Les festivals ou cérémonies sont également des catégories avec une grande quantité de tweets, notamment pour les films ayant remporté des prix. Le film Boite noire n'ayant reçu aucun prix dans ces 3 évènements, cette catégorie ne figure pas dans les plus citée.

Enfin, afin de ne pas avoir une surcharge de données dans la création des events log, seuls les renseignements essentiels ont été gardés : date de création du tweet, catégorie et id de l'auteur.

5.1.3. Processus par film

5.1.3.1. Processus général

Le code utilisé pour cette partie ainsi que les réponses de la console se trouvent en [annexe 5](#).

En considérant les 2886 tweets catégorisés du film Boite noire, 2886 traces ont été observées avec un nombre de case_id égale à 2137. Pour rappel, un case_id correspond à un utilisateur. Cela signifie qu'en moyenne une personne a tweeté, selon les catégories susmentionnées, 1,35 fois à propos du film.

Le tableau ci-dessous, tableau 5.3., montre le nombre de traces et de case_id ainsi qu'une moyenne des tweets par personne par film.

Films	Nombre Evènements	Nombre de cases	Moyenne de tweets par personne
Illusions Perdues	2137	1476	1,44
Benedetta	3166	2106	1,5
L'évènement	1637	1125	1,45
Les Olympiades	974	678	1,43
Titane	8342	4980	1,67

Tableau 5.3. Moyenne de tweets par personne par film

La moyenne générale est approximativement similaire pour tous les films.

Quand on observe le début et la fin d'activité (Cfr [annexe 5](#)) pour chaque event log pour le film Boite noire, il n'est pas incohérent d'obtenir l'activité « concours » dans les deux cas. En effet, étant donné que le nombre d'activités moyen s'élève à 1,35, beaucoup de personnes n'ont, en réalité, effectué qu'un seul tweet pour gagner un concours.

Ce schéma semble récurrent dans les autres exemples. Cependant, les films Benedetta et Titane ne s'y conforment pas. En effet, pour chacun de ces films, la première et dernière activités est le festival de Canne. Comme expliqué précédemment, cet évènement génèrent beaucoup de réaction. Etant donné que ces 2 films étaient en compétition pour la palme, ces nominations ont suscité la réaction d'un nombre important d'internautes.

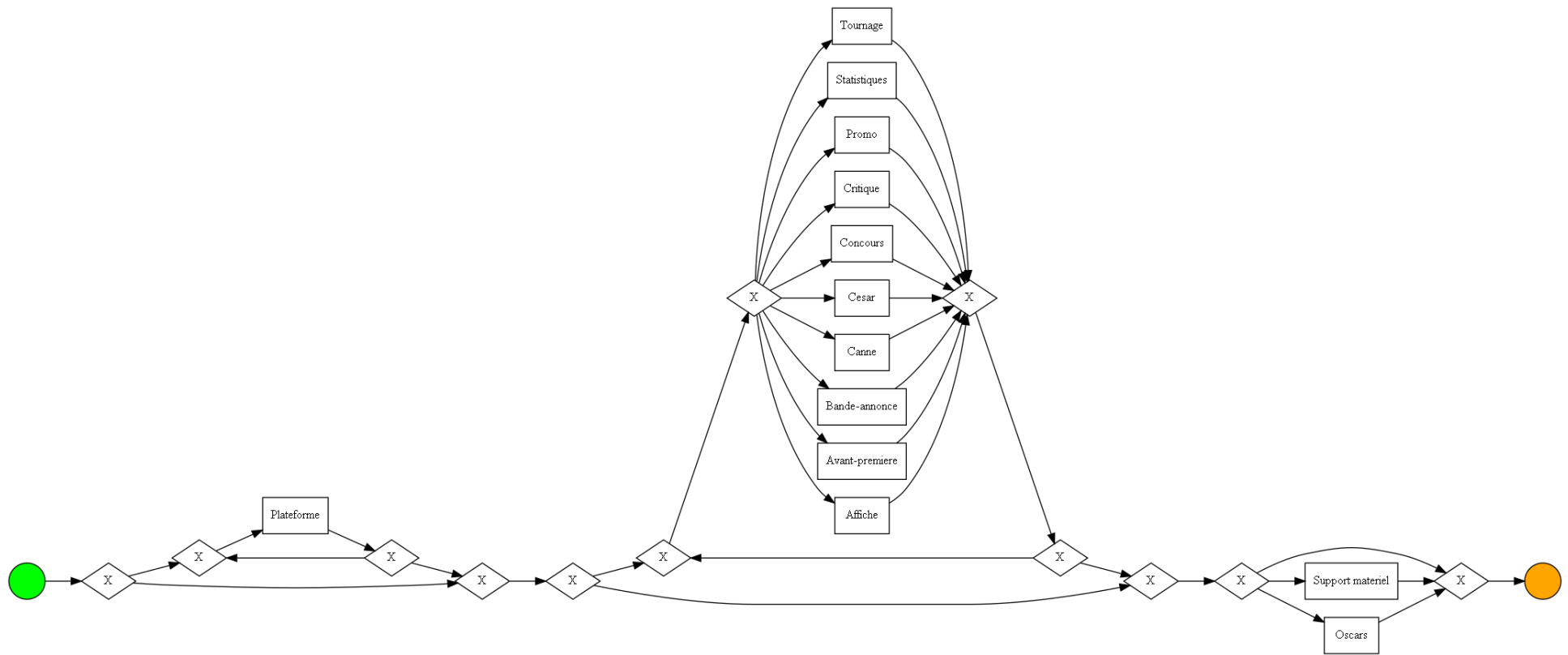


Figure 5.4. Processus de chaque utilisateur sur la plateforme Twitter s'exprimant sur le film « Boite noire

Dans ce processus, généré grâce à l'algorithme PM4PY, uniquement des « XOR » peuvent être identifiés, cela veut dire que chaque case_id, donc chaque utilisateur, prend un chemin ou l'autre. Ici, beaucoup de chemins sont possibles. Cela est dû au nombre de case_id différent. Pour remédier à cela, il est nécessaire d'effectuer du trace clustering. De plus, ce processus ne fait pas de différence entre ce qu'il est important ou non. Pour éviter cela, les prochaines analyses seront faites avec l'algorithme « fuzzy miner ».

5.1.3.2. Trace clustering

« La technique du trace clustering suppose que la complexité est due à la présence de traces hétérogènes des event log décrivant les différentes variantes du processus » (Diamantini et al., 2017). Afin d'avoir des processus plus clairs et compréhensibles, les case_id sont séparés en catégories également.

Dans le contexte du cinéma, le choix a été fait de séparer les comptes en trois catégories : les comptes des personnes impliquées dans le film (acteur, réalisateur, société de distribution, ...), ceux des médias (incluant les journaux, les émissions, les youtubeurs axés cinéma, ...) et pour finir, ceux des utilisateurs individuels. L'[annexe 4](#) indique les différentes catégories pour le cas du film Boîte noire. Au vu de la quantité d'utilisateurs individuels, ceux-ci ne sont pas repris dans le tableau. Ces choix ont été réalisés avec l'aide d'une spécialiste dans le domaine, une chargée de production.

Pour automatiser le processus de classification des comptes, la méthode par mots clefs en fonction de description de chacun des comptes a été mise en place. Tous les comptes contenant les mots : « journaliste cinéma », « rédaction », « attaché de presse », « rédacteur », « suivez-nous », « chargé de communication », « Podcast », « TF1 » et « RTL » ont automatiquement été classés dans la case Média.

5.1.3.3. Processus parties-prenantes

Dans le cas du film Boîte noire, très peu de personnes directement impliquées dans le film sont présentes sur Twitter. En effet, le nombre de cases s'élève à 7 pour 22 event en tout. 6 variantes peuvent être retrouvées ce qui signifie que la majorité des personnes ont pris un chemin différent. Le processus est disponible en [annexe 6.1.](#)

Ce phénomène est présent pour l'entièreté des films. En effet, peu d'acteurs ou réalisateurs se trouvent sur Twitter actuellement concernant le cinéma français.

5.1.3.4. Processus média

La figure II.3 montre le processus de média du film Boite noire, on peut recenser 81 events pour 50 case_id. Afin d'obtenir un graphique plus général, le processus prend en compte 33% des chemins.

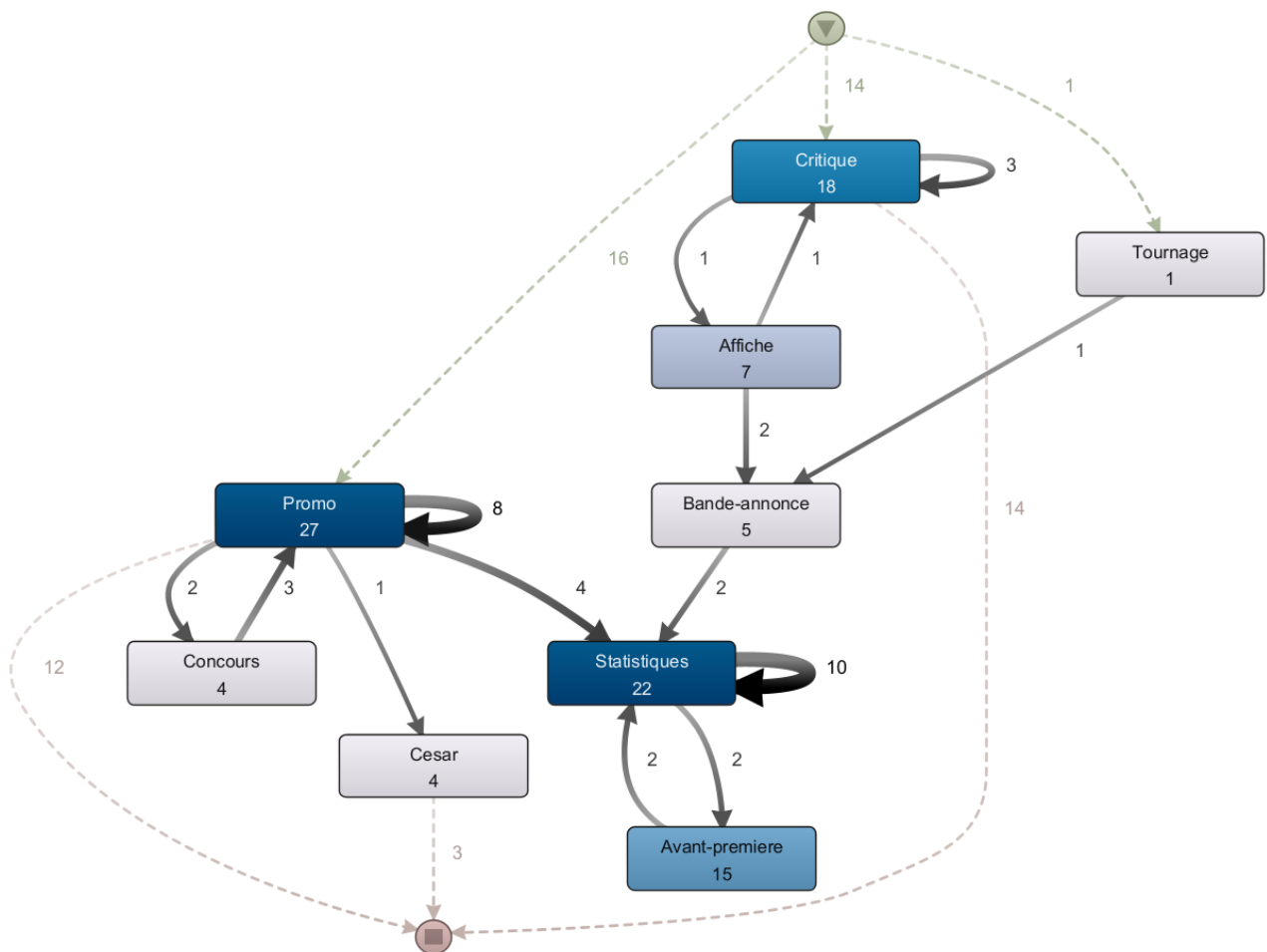


Figure 5.5 Processus des media des 50 cases_id du film Boite noire créé à partir du logiciel Disco

Parmi les 50 cases_id, 24 chemins différents ont été empruntés (Cfr [Annexe 6.2.](#)). Le plus fréquent est celui qui ne passe qu'une seule fois par l'activité « Critique ». En effet, 14 cases

commence et termine par celle-ci. Twitter permet à de nombreux médias indépendants de faire la promotion de leurs articles. A l'heure actuelle, il est facile, de créer un blog, un site internet, ... Cela explique que ce chemin est le plus fréquent.

La différence des chemins dépend des différents médias. Certains sont axés sur la critique, il s'agit des blogs indépendants, des sites d'avis, etc. D'autres commentent plus la vie du film en signalant l'affiche, la bande annonce, etc. Bien évidemment, une grande partie des médias sert à faire la promotion du film, c'est l'activité la plus effectuée. Pour rappel, cette activité représente tout tweet traitant des différentes actions par les membres du film pour le promouvoir : apparition de l'acteur dans une émission radio, tv, etc. Ce n'est pas insensé que ces médias suivent ce chemin plus particulier et avec moins de connexion avec les autres activités. En effet, ces émissions parlent généralement du film uniquement au moment de la présence de l'invité.

L'activité concours est beaucoup moins effectuée. Pour rappel, Cette catégorie était pourtant celle qui suscitait le plus d'activité. Ce changement est dû au fait que ce ne sont pas les médias qui participent aux concours mais en sont plutôt les organisateurs.

En ce qui concerne les processus des autres films, des similitudes peuvent être retrouvées. Les processus se trouvent en [annexe 7](#). Même si cela est parfois moins visible, l'activité « critique » reste une activité effectuée par une partie des médias (les blogs indépendants, sites d'avis, ...). Certains processus, comme celui de Benedetta par exemple, ont une activité concours avec beaucoup plus d'occurrences, cela peut s'expliquer par le fait que certains médias sont en réalité de réelles personnes, passionnés de cinéma, voulant participer aux concours. C'est notamment eux qui constituent les médias indépendants. Encore une fois, les films ayant gagné beaucoup de prix ou participé à l'une ou/et l'autre compétition (Césars et Canne) ont cette activité au cœur de leur processus. C'est l'activité la plus exécutée et beaucoup de chemins passent par elle. Pour les festivals de Canne, Benedetta, les Olympiades et Titane sont les exemples. Titane ayant remporté à la Palme d'or, cette activité est encore plus centrale. En ce qui concerne la cérémonie des Césars, Illusions perdues se différencie. En effet, ce film en a gagné 8.

De manière générale, l'activité « Affiche » se place toujours avant l'activité « Bande-annonce ». Cela semble logique étant donné qu'on divulgue l'affiche avant la bande-annonce.

La figure 5.6. ci dessous montre le nombre d'activités par case pour le cas de Boite noire.



Figure 5.6. Graphique représentant le nombre de par case_id du processus des médias du film Boite noire. Graphique créé avec le logiciel Disco

La première colonne représente le nombre de case n'ayant eu qu'une seule activité. Cela représente la majorité des cas. C'est le cas également pour tous les autres films. Attention, il ne faut pas oublier qu'il s'agit du nombre de tweet tweeté en fonction de catégorie et non du nombre de tweet réel.

5.1.3.5. Processus utilisateurs

Le processus des utilisateurs n'a pas donné de résultat probant. En effet, au vu du nombre de comportements différents des utilisateurs, le processus retiré est difficilement analysable. Le processus se trouve en annexe 5. Il est quand même possible d'observer un attrait pour les concours, en effet c'est l'activité la plus réalisée avec 1846 occurrences.

5.1.4. Généralisation

Afin de pouvoir identifier si des processus généraux peuvent être observés, une visualisation d'un processus comprenant tous les films est nécessaire.

Sur le processus ci-dessous, 5 films sont représentés ; Boite noire, Illusions perdues, L'évènement et les Olympiade et Benedetta. Le processus correspond au processus de toutes les personnes ayant tweeté qu'importe leur catégorie, c'est-à-dire qu'il n'y a pas de distinction entre les utilisateurs, les médias ou les potentiel parties prenantes. Le film Titane ayant des différences plus importante par rapport aux autres films, celui-ci ne figure pas dans le processus. En effet, étant donné que ce film a remporté la Palme d'or, l'activité « Canne » peut changer complètement le processus.

Etant donné la quantité de données et de chemin différent, ce processus ne montre que 75% des activités les plus exécuté et montre uniquement les chemins les plus utilisés.

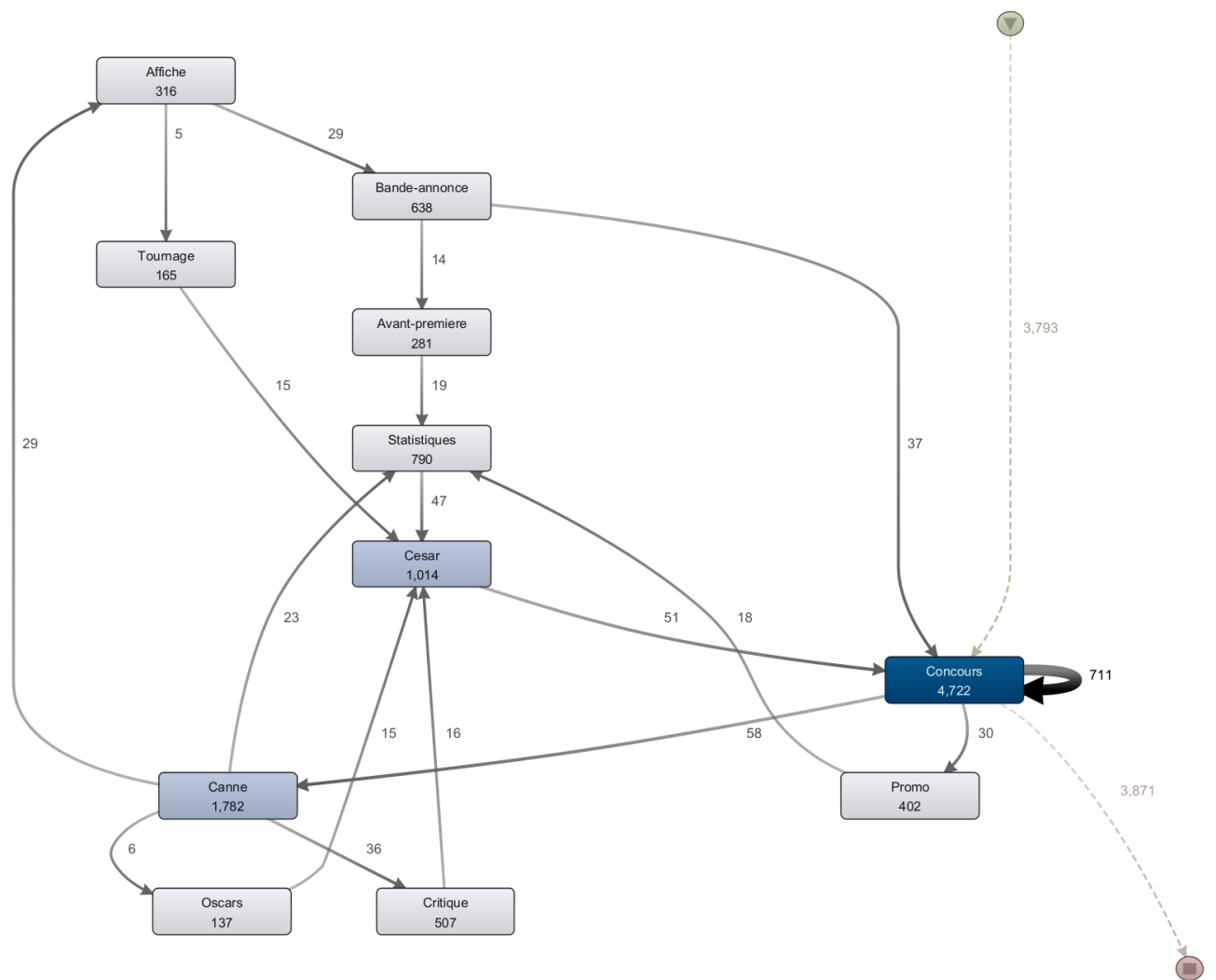



Figure 5.7. Processus générale des films Boite noire, Illusions perdues, L'évènement et les Olympiade et Benedetta.

Grâce au fait que nous n'avons pris en compte que les éléments et chemins le plus exécutés, le processus est facilement interprétable.

Tout d'abord, l'activité « concours » est la plus exécutés. Cela fait sens par rapport aux précédentes analyses. Ensuite, l'emplacement des activités les une par rapport aux autres est très facilement explicable. En regardant le chemin « Affiche », « Bande-annonce », « Avant-première », « Statistiques » puis « César », l'ordre correspond au cycle de vie d'un film. Entre l'activité « Avant-première » et l'activité « Statistiques » se trouve la sortie du film au cinéma. La cérémonie des césars se déroule après la sortie au cinéma. Quant à l'activité « Canne », son emplacement est également logique. En effet, ce festival a lieu généralement avant la sortie des films au cinéma.

L'activité « tournage » ne se trouve pas avant l'activité « Affiche », hors l'affiche est généralement dévoilé bien après le tournage. Voici deux exemple bien distinct de l'activité. Dans le cas de Boite noire, l'activité est la première d'un case_id. Il s'agit du tweet : « *Après #UnHommeIdéal, Pierre Niney retrouve le réalisateur Yann Gozlan pour le thriller #BoîteNoire, aux côtés de Lou de Laâge et André Dussollier. Le film, qui retrace l'investigation du crash d'un avion entre Dubaï et Paris, vient d'entrer en tournage en Île-de-France. [...]* ». Le tweet a été publié au moment du tournage. Cependant pour le même film, ce tweet a également été publié : « * #JeudiPhoto retour sur le tournage du film #boitenoire au sein du laboratoire du @BEA_Aero le samedi 26 octobre 2019. <https://t.co/Af4PUYpxB9>* ». Ce tweet date du 5 décembre 2021. ».

Chapitre VI. Conclusion

6.1. Apports méthodologique et analytiques

6.1.1. Apport méthodologique

L'objectif de ce mémoire est de mettre en lumière le potentiel des techniques du process mining dans le cadre de l'analyse des réseaux sociaux et plus particulièrement Twitter.

Dans un premier temps, la revue de la littérature a permis de comprendre les enjeux et les possibilités qu'offrait le process mining. En effet, cette technologie ne cesse de se développer, de nombreux algorithmes voient le jour pour permettre des analyses de plus en plus diverses.

Grâce aux différents outils employés, il a été possible de réaliser des processus afin de modéliser la façon dont les personnes parlent d'un sujet en particulier sur le réseau social Twitter. Tout d'abord, l'extraction de données est réalisée grâce à l'API de twitter. Ensuite, la réalisation d'un nettoyage des données et de la transformation de celles-ci en données exploitables, en events log est nécessaire pour l'analyse de process mining. Cette étape est cruciale pour le travail étant donné qu'elle contribue fortement au processus final. Il s'agit d'ailleurs d'une étape avec un fort potentiel d'amélioration.

6.1.2. Apport analytiques

Les analyses descriptives ainsi que les processus ont pu montrer certains attraits pour des tweets en particulier. Les concours sur Twitter génèrent beaucoup de réactions. Beaucoup d'utilisateurs ne s'expriment sur un film qu'à travers eux, en retweetant le concours pour gagner des places de cinéma ou encore des DVD.

De plus, les grands événements du cinéma comme Cannes ou les césars génèrent également un certain engouement. Cet engouement est d'autant plus flagrant si le film a remporté des prix lors de ces cérémonies.

Les processus ont pu mettre en avant les différents types de médias présents sur Twitter. En effet, cette plateforme permet à tout un chacun d'avoir un compte dans le but de faire la promotion de leurs blogs, podcasts, ... Il est donc possible de distinguer ce type de médias aux médias traditionnels comme les émissions ou journaux reconnus.

Enfin, de manière générale, le cycle de vie d'un film se retrouve dans le processus. Les utilisateurs vont d'abord mentionner les premières étapes d'un film comme l'affiche ou la bande-annonce.

6.2. Limites du travail

Quelques étapes de la méthodologie peuvent être améliorées.

Tout d'abord, la partie de la classification des tweets peut être réalisée d'une autre façon. En effet, la méthode actuelle ne permet pas de prendre en compte tous les tweets mais ne se concentre, en général ; que sur moins de 50% des tweets. Ce manque de classification se répercute par la suite sur les processus. Dans cet étude de cas, cela concerne énormément les tweets de utilisateurs parlant de leur ressenti sur le film. Un algorithme capable de classer en prenant en compte l'entièreté du contenu des tweets serait plus adéquat. De plus, certains tweets faisaient référence à deux catégories. Cependant, une seule activité pouvait lui être attribuée. Ce même problème se retrouve quant il s'agit de classer les comptes pour le trace clustering.

Pour finir, le nombre de films analysés peut être augmenté. Le choix du mémoire a été d'identifier une méthodologie pour analyser le réseau social à l'aide de techniques de process mining. Cependant, si le but est de se concentrer sur l'étude de cas, c'est-à-dire le cinéma français, des analyses supplémentaires sont nécessaires pour confirmer et infirmer les premiers résultats de ce travail.

6.3. Pistes travaux futures

Comme énoncé précédemment, des améliorations peuvent être apportées à ce travail.

Dans un premier temps, il pourrait être intéressant d'utiliser des techniques de clustering sur les tweets et observer si des groupes naturels émergent des données. Des techniques de text mining peuvent également être utilisées.

De plus, analyser une plus grande quantité de film en choisissant des films sortis à d'autres moments de l'année ou dans d'autres langues pourrait permettre de mieux déterminer si le processus est généralisable ou non.

Enfin, comprendre comment les différentes catégories (médias, parties-prenantes, utilisateurs) interagissent entre eux peut faire l'objet d'un travail futur.

Références

- Aalst, W. v., Adriansyah, A., Medeiros, A. K., Arcieri, F., Baier, T., Blickle, T., . . . others. (2011). Process mining manifesto. *International conference on business process management*, (pp. 169–194).
- Barbier, G., & Liu, H. (2011). Data mining in social media. In *Social network data analytics* (pp. 327–352). Springer.
- Benedetta*. (n.d.). Retrieved Juillet 1, 2022, from AlloCiné:
https://www.allocine.fr/film/fichefilm_gen_cfilm=255795.html
- Boite noire*. (2022, Mai 22). Retrieved from AlloCiné:
<https://www.allocine.fr/rechercher/?q=Boite+noire>
- Centre d'assistance*. (2022, 08 02). Retrieved from Twitter: <https://help.twitter.com/fr/rules-and-policies/twitter-api>
- Diamantini, C., Genga, L., Marozzo, F., Potena, D., & Trunfio, P. (2017). Discovering mobility patterns of instagram users through process mining techniques. *2017 IEEE International Conference on Information Reuse and Integration (IRI)*, (pp. 485–492).
- Douarin, R. L. (2022, Juin 9). Le monde du cinéma français. (M. Grailly, Interviewer)
- Ghahfarokhi, A. F., Park, G., Berti, A., & van der Aalst, W. M. (2021). OCEL: A standard for object-centric event logs. *European Conference on Advances in Databases and Information Systems*, (pp. 169–175).
- Ghawi, R. (2016). Process Discovery using Inductive Miner and Decomposition. *arXiv preprint arXiv:1610.07989*.
- Gupta, E. P. (2014). Process mining a comparative study. *International Journal of Advanced Research in Computer and Communications Engineering*, 3, 5.
- Illusions perdues*. (n.d.). Retrieved Juillet 5, 2022, from AlloCiné:
https://www.allocine.fr/film/fichefilm_gen_cfilm=273621.html
- Les olympiades*. (2022, Juillet 31). Retrieved from AlloCiné:
https://www.allocine.fr/film/fichefilm_gen_cfilm=285885.html
- L'évènement*. (n.d.). Retrieved Juillet 16, 2022, from AlloCiné:
https://www.allocine.fr/film/fichefilm_gen_cfilm=275122.html
- Li, G., & De Carvalho, R. M. (2019). Process mining in social media: applying object-centric behavioral constraint models. *IEEE Access*, 7, 84360–84373.
- Li, G., Carvalho, R. M., & van der Aalst, W. M. (2017). Automatic discovery of object-centric behavioral constraint models. *International conference on business information systems*, (pp. 43–58).
- Li, G., Murillas, E. G., Carvalho, R. M., & Van Der Aalst, W. M. (2018). Extracting object-centric event logs to support process mining on databases. *International Conference on Advanced Information Systems Engineering*, (pp. 182–199).

- Liao, L., & Huang, T. (2021). The effect of different social media marketing channels and events on movie box office: An elaboration likelihood model perspective. *Information & Management*, 58, 103481.
- Nasirifard, P., Peristeras, V., Hayes, C., & Decker, S. (2009). Extracting and utilizing social networks from log files of shared workspaces. *Working Conference on Virtual Enterprises*, (pp. 643–650).
- Oh, C., Roumani, Y., Nwankpa, J. K., & Hu, H.-F. (2017). Beyond likes and tweets: Consumer engagement behavior and movie box office in social media. *Information & Management*, 54, 25–37.
- Rui, H., Liu, Y., & Whinston, A. (2013). Whose and what chatter matters? The effect of tweets on movie sales. *Decision support systems*, 55, 863–870.
- Song, M., Günther, C. W., & Van der Aalst, W. M. (2008). Trace clustering in process mining. *International conference on business process management*, (pp. 109–120).
- Titane*. (2022, Juin 22). Retrieved from AlloCiné: <https://www.allocine.fr/rechercher/?q=Titane>
- Turner, C. J., Tiwari, A., Olaiya, R., & Xu, Y. (2012). Process mining: from theory to practice. *Business Process Management Journal*.
- Van Der Aalst, W. (2012). Process mining: Overview and opportunities. *ACM Transactions on Management Information Systems (TMIS)*, 3, 1–17.
- Van der Aalst, W. M., & Song, M. (2004). Mining social networks: Uncovering interaction patterns in business processes. *International conference on business process management*, (pp. 244–260).

Annexes

Annexe 1 – code R

1.1. Boite Noire

```
BoiteNoire<-
read.csv(file='C:/Users/grail/OneDrive/Documents/Mémoire/BoiteNoire/Excel/C
SV/BoiteNoireFinal.csv')

#Mettre la colonne created_at dans le format date

BoiteNoire$created_at = as.Date(BoiteNoire$created_at, "%Y-%m-
%dT%H:%M:%S.000Z")

#graphique nombre de tweet en fonction du temps

library("dplyr")
library("ggplot2")
library("hrbrthemes")
library("scales")

p= ggplot(BoiteNoire)+
  aes(x=BoiteNoire$date)+
  ylab("Nombre de tweets")+
  xlab("Temps")+
  ggtitle("Evolution du nombre de tweets")+
  geom_bar()

min <- as.Date("2019-09-01")
max <- NA
p + scale_x_date(limits = c(min, max), date_breaks = "6 month")

p= ggplot(BoiteNoire)+
  aes(x=BoiteNoire$date)+
  ylab("Nombre de tweets")+
  xlab("Temps")+
  ggtitle("Evolution du nombre de tweets")+
  geom_bar()

min <- as.Date("2021-06-01")
max <- NA
p + scale_x_date(limits = c(min, max), date_breaks = "2 month")

#vue globale de toutes les données
summary(BoiteNoire)
```

Réponses console :

created_at	retweet_count	reply_count
Min. :2019-10-08	Min. : 0	Min. : 0.0000
1st Qu.:2021-09-05	1st Qu.: 0	1st Qu.: 0.0000
Median :2021-09-15	Median : 6	Median : 0.0000
Mean :2021-09-27	Mean : 189	Mean : 0.2437
3rd Qu.:2021-10-27	3rd Qu.: 72	3rd Qu.: 0.0000
Max. :2022-04-25	Max. :1287	Max. :226.0000

like_count	quote_count	username
------------	-------------	----------

Min.	: 0.000	Min.	: 0.0000	pierreniney	: 96
1st Qu.:	0.000	1st Qu.:	0.0000	boxofficefr	: 54
Median	: 0.000	Median	: 0.0000	summergirly7180:	40
Mean	: 3.927	Mean	: 0.0919	club300allocine:	35
3rd Qu.:	0.000	3rd Qu.:	0.0000	gysmoanne	: 35
Max.	:1553.000	Max.	:30.0000	STUDIOCANAL	: 35
		NA's	:497	(Other)	:6176

1.2. Illusions perdues

Réponse console :

```
username
DasyBab      : 628
ZerroualN   : 190
summergirly7180: 63
Laura_Dambron : 59
Gaumont     : 51
boxofficefr : 31
(Other)     :4429
```

1.3. Benedetta

Réponse console :

```
username
ben80707145 : 69
summergirly7180: 64
sebyoda     : 58
MitraEtemad : 34
PatheFilms  : 34
ThierryBarnaud : 25
(Other)     :5556
```

1.4. L'évènement

```
username
wildBunch   : 62
summergirly7180: 53
RectangleProd : 31
MitraEtemad : 22
GubernCristina : 20
boxofficefr  : 15
(Other)     :2941
```

1.5. Les olympiades

Réponse console :

```
username      date
mementodistrib : 79
alexmento     : 54
panciutello   : 50
summergirly7180: 50
Infinemusic   : 45
BorisBastide  : 31
(Other)       :1772
```

1.6. Titane

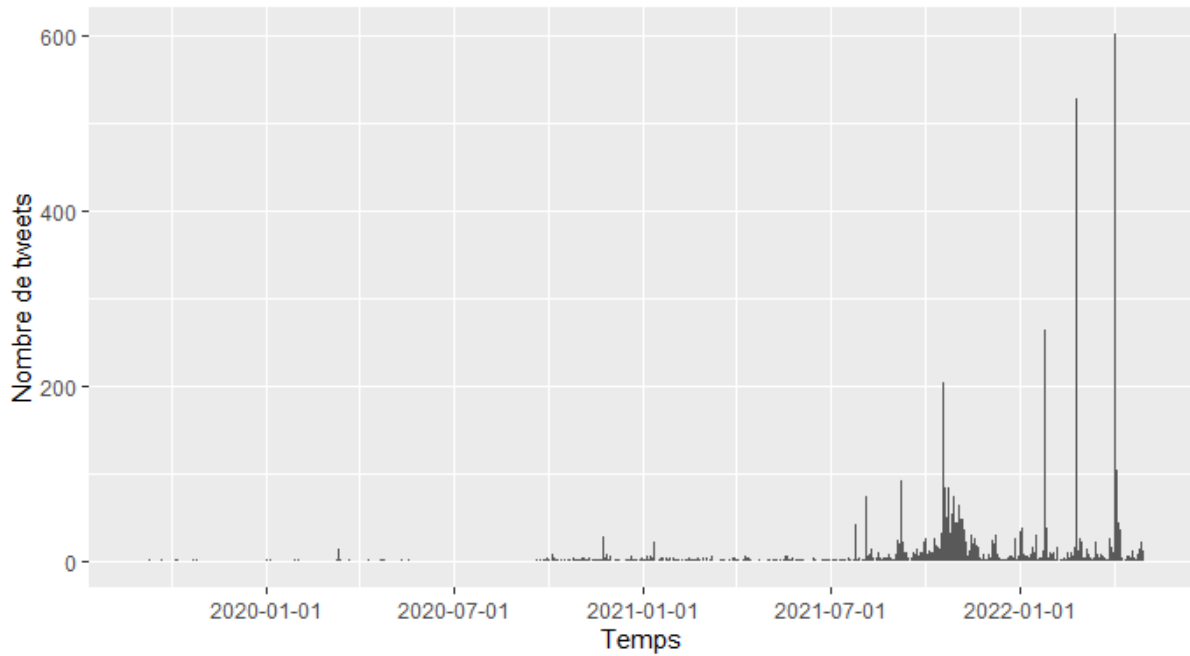
Réponse console :

```
username
summergirly7180: 65
cinecharlie : 51
FlorianDurand7 : 48
boxofficefr : 33
diaphana : 33
destinationcine: 31
(Other) :8081
```

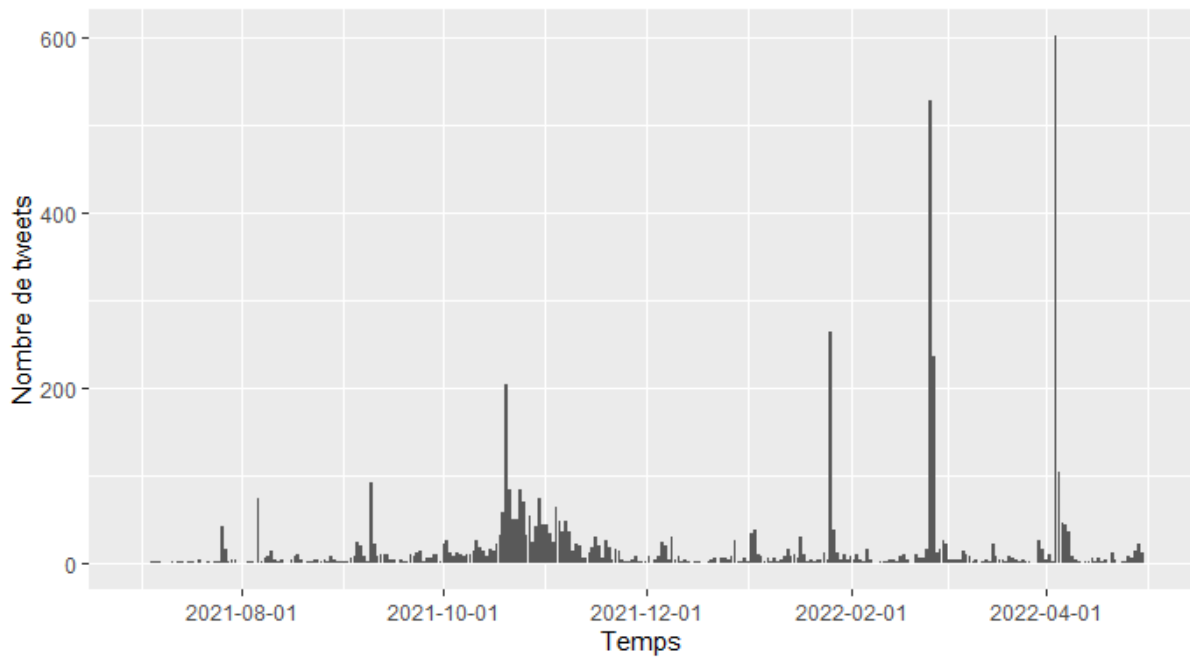
Annexe 2 – Graphiques - Autres films

2.1. Illusions Perdues

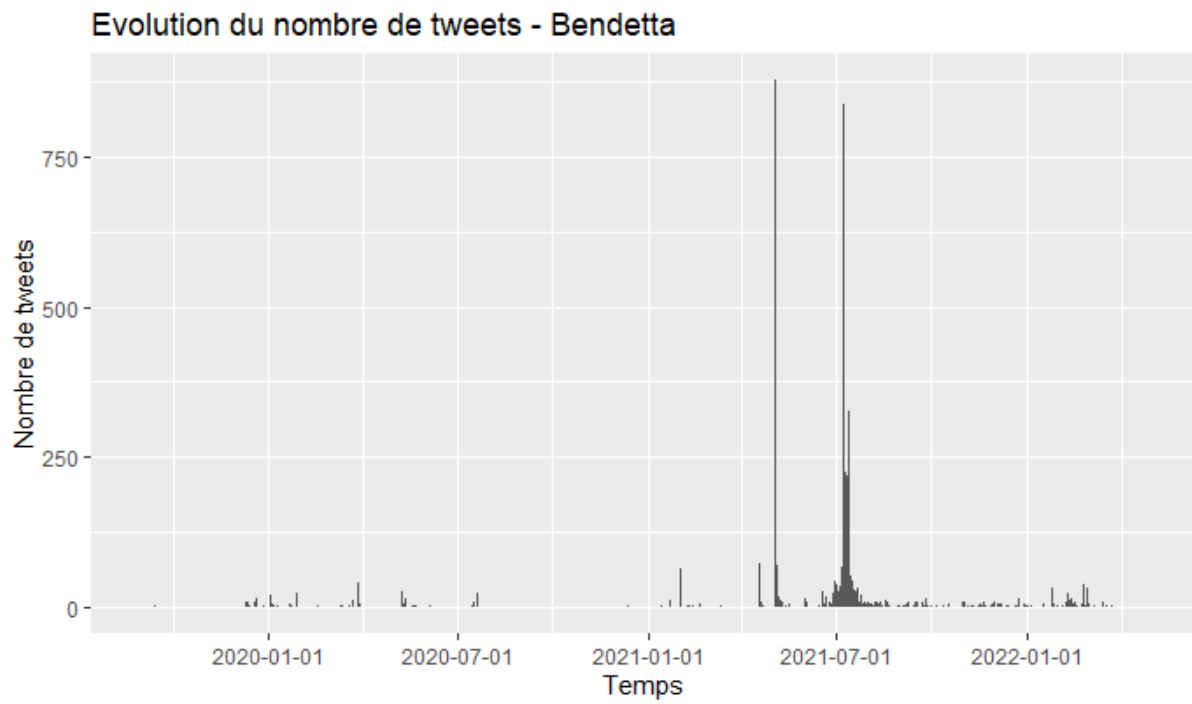
Evolution du nombre de tweets - Illusions perdues



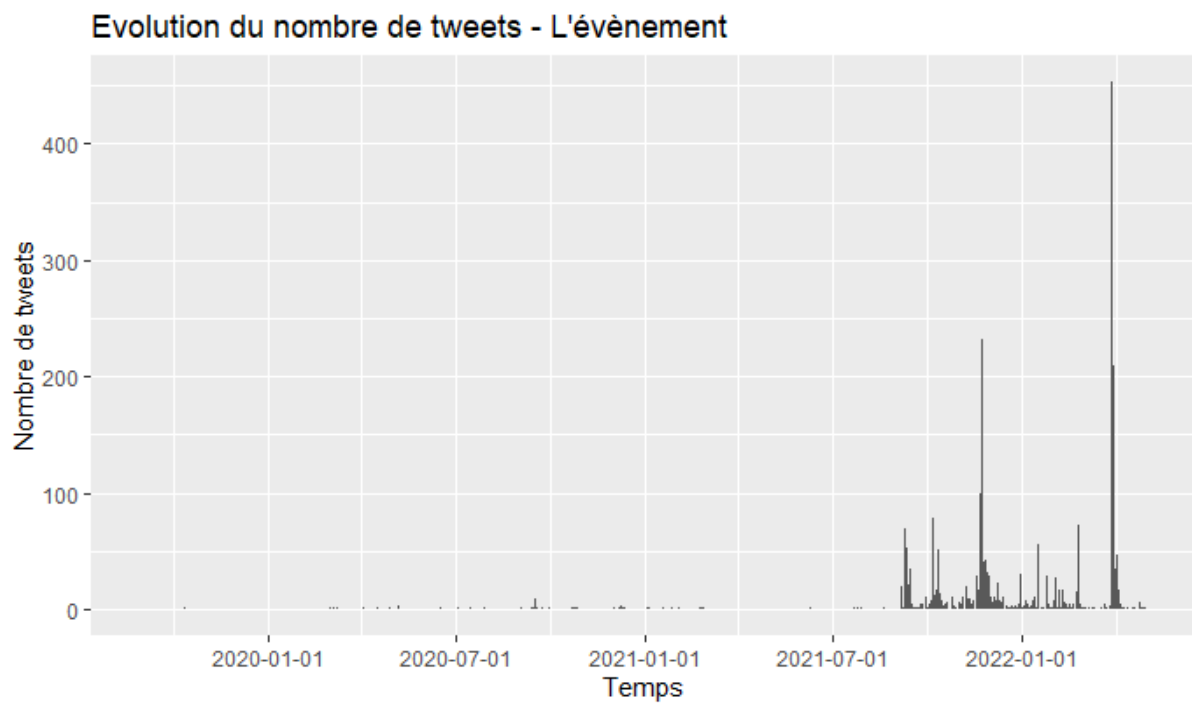
Evolution du nombre de tweets - Illusions perdues



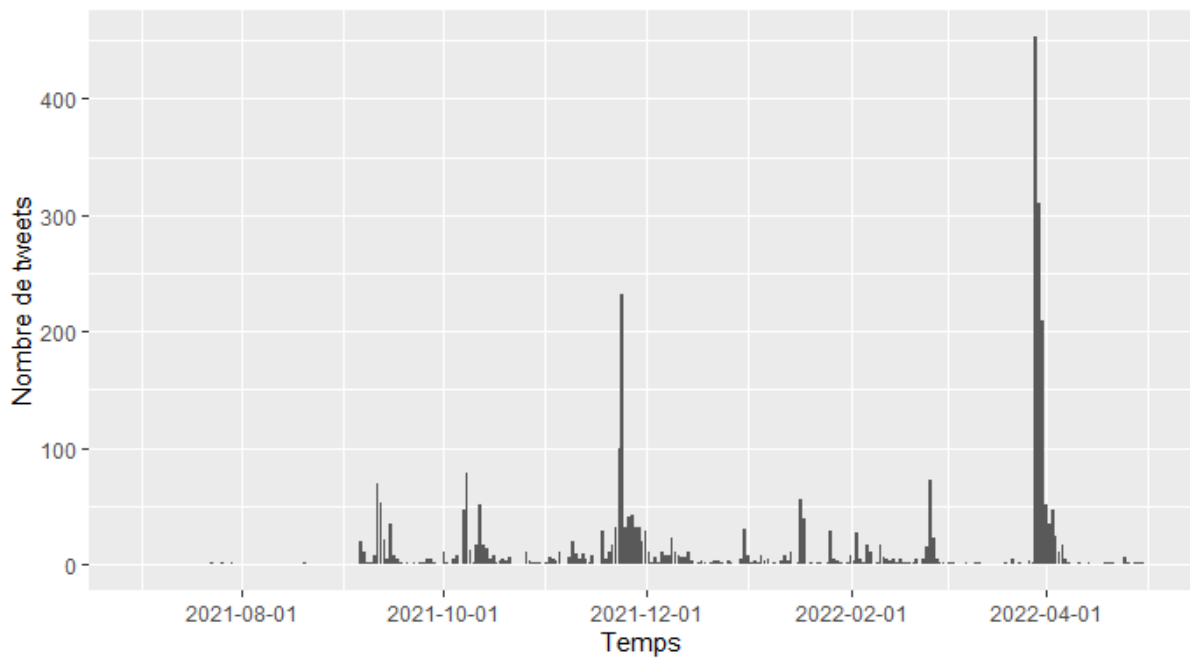
2.2. Benedetta



2.3. L'évènement

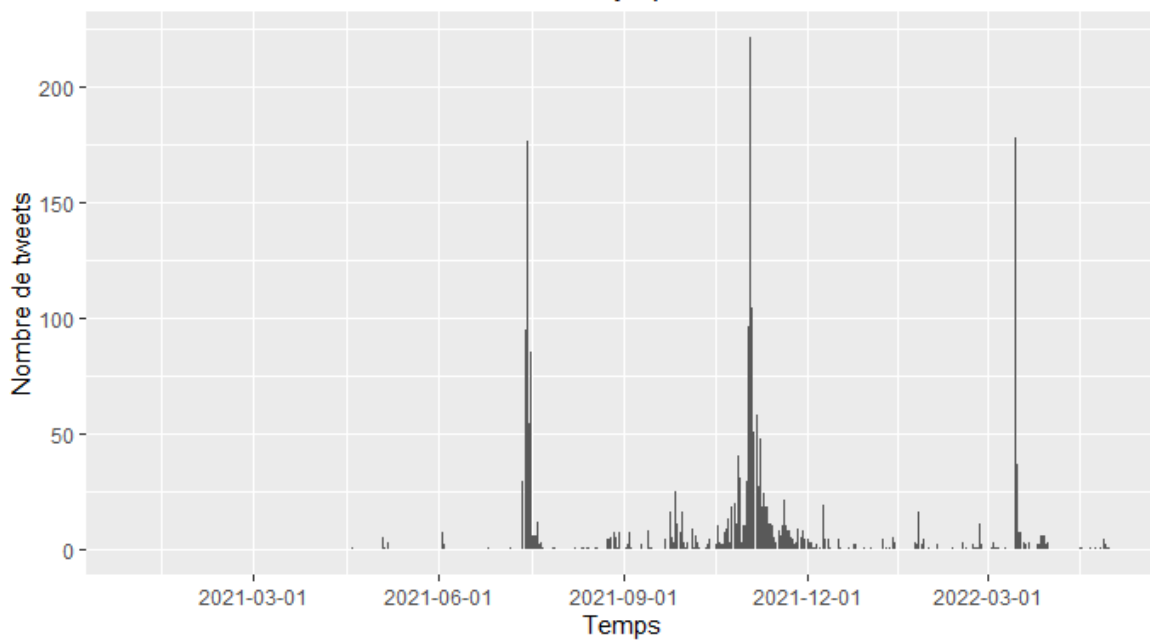


Evolution du nombre de tweets - L'évènement

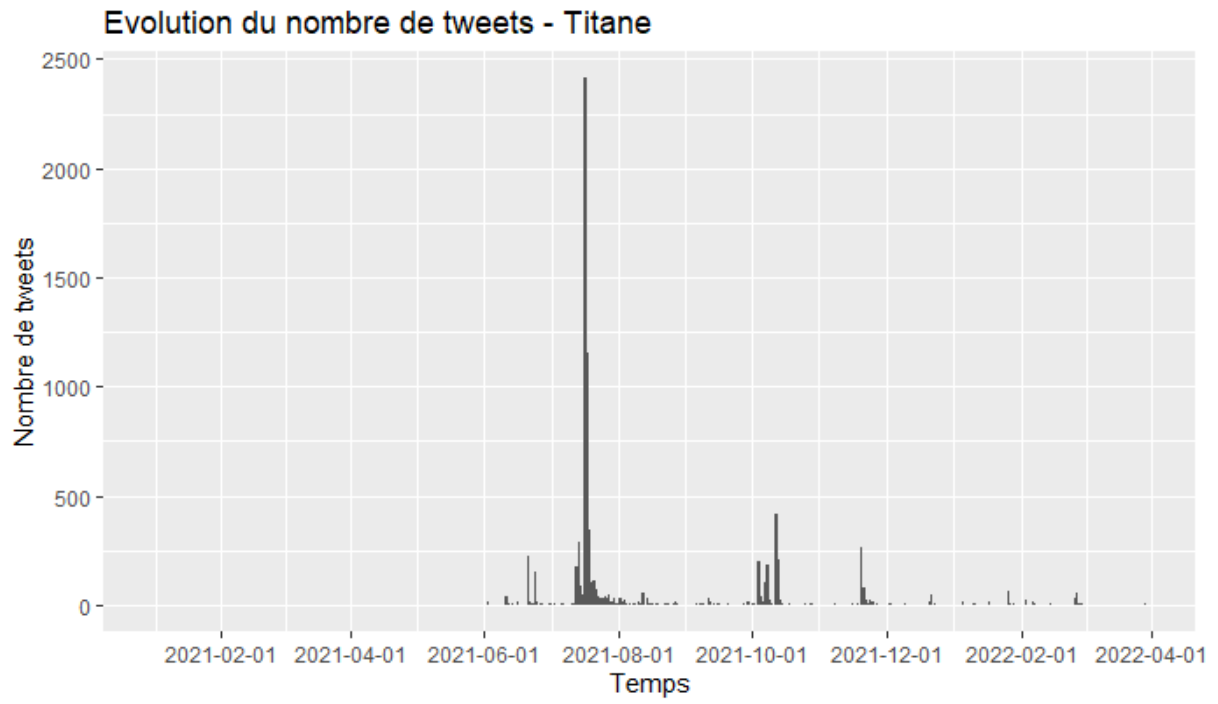


2.4. Les Olympiades

Evolution du nombre de tweets - Les olympiades

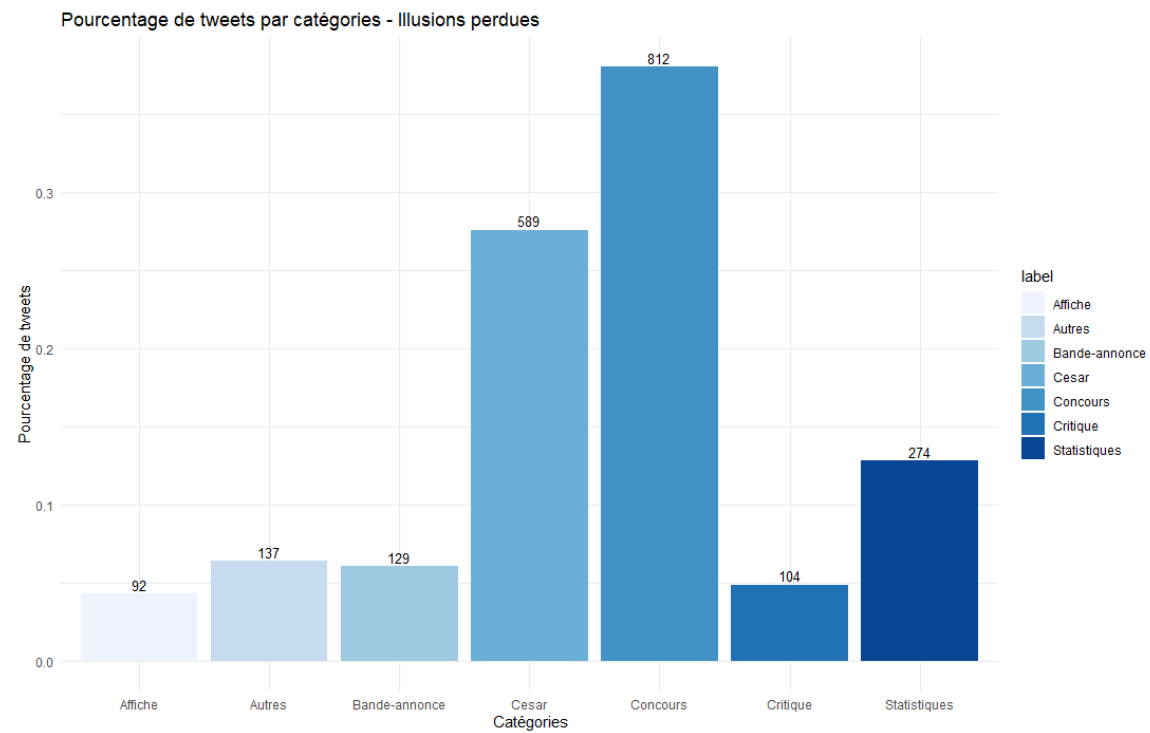


2.5.Titane

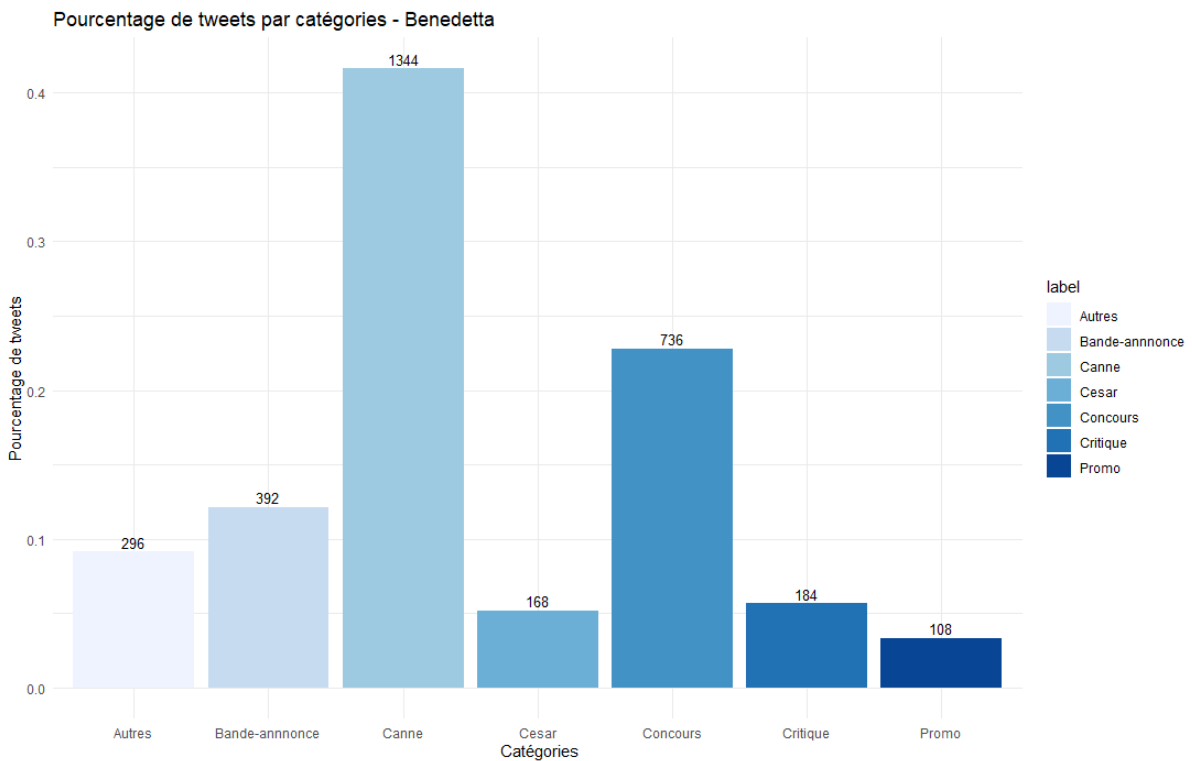


Annexe 3 – Graphiques tweets en fonction des catégories

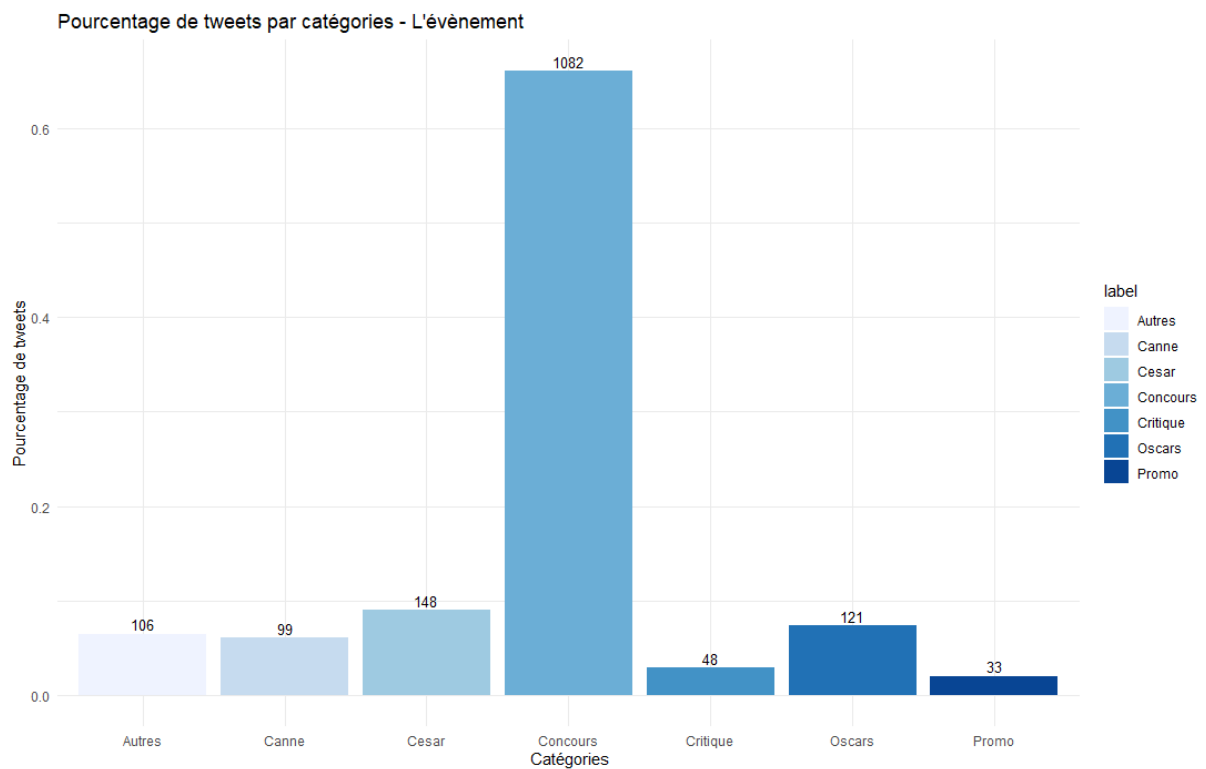
3.1. Illusions perdues



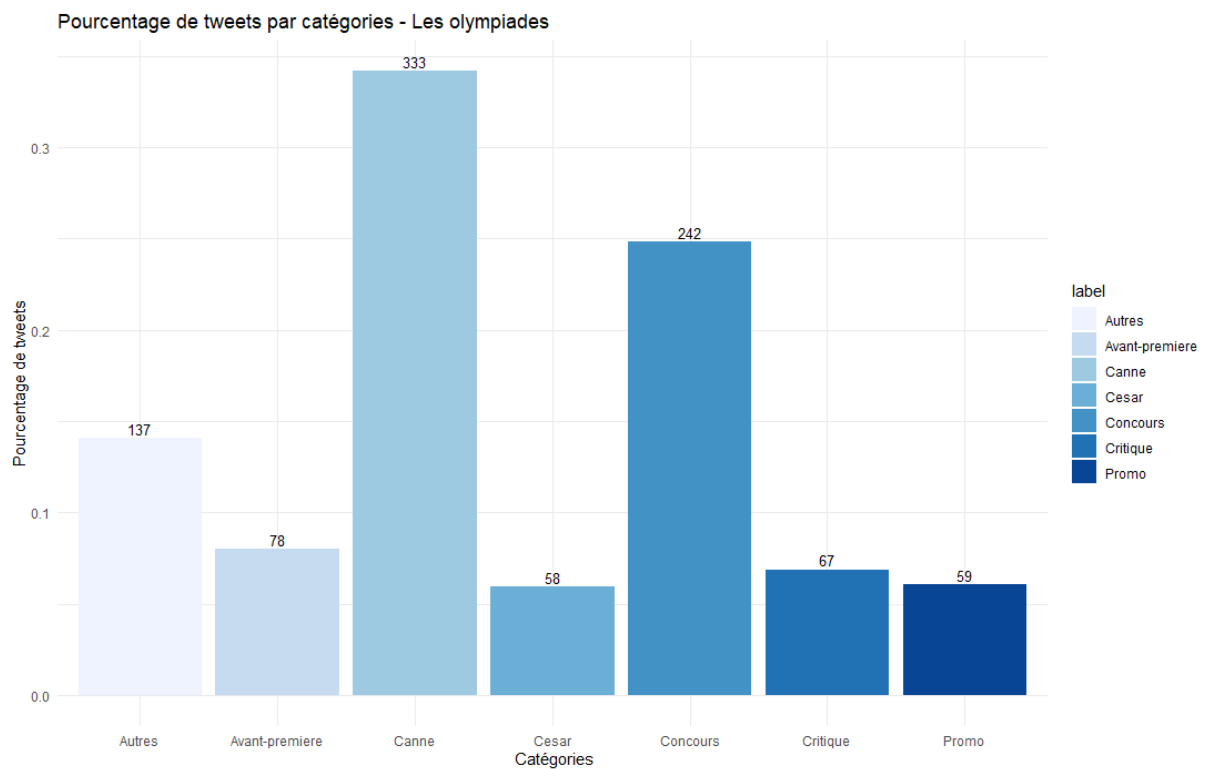
3.1. Benedetta



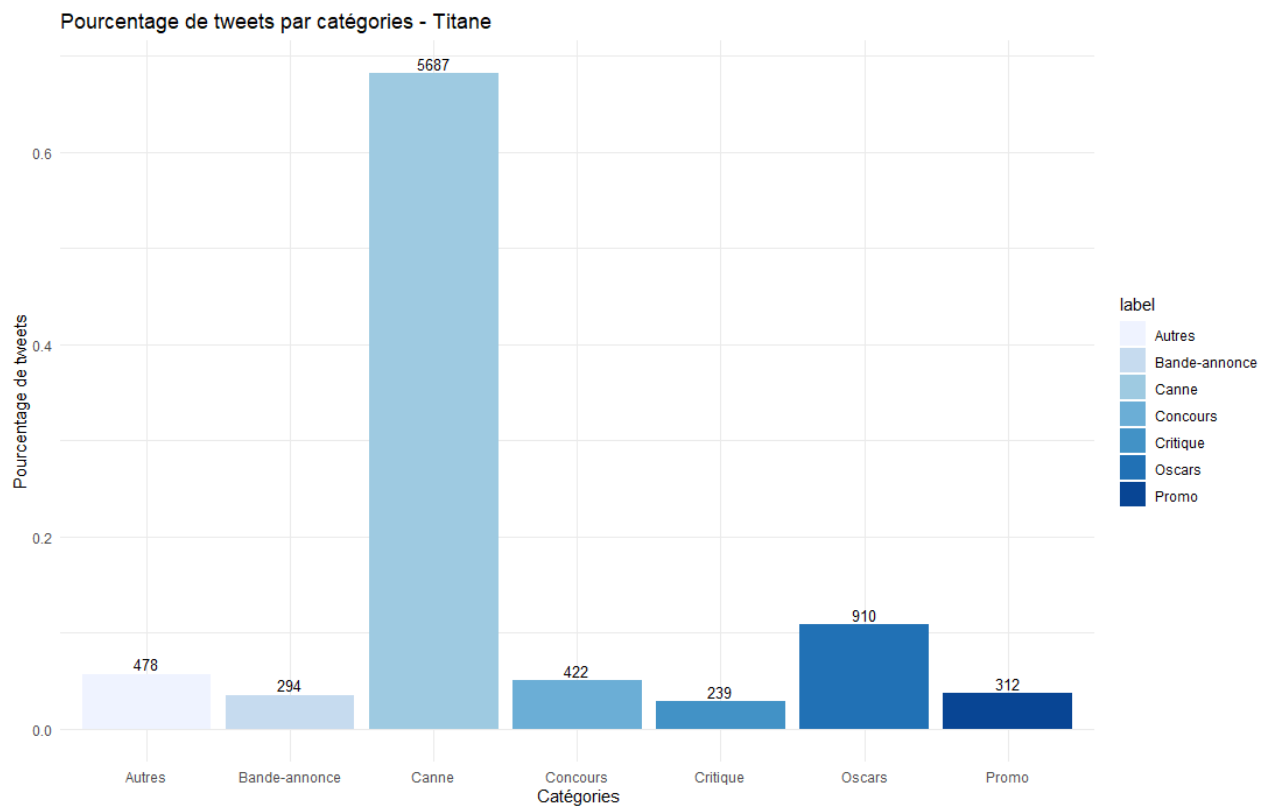
3.2. L'évènement



3.3. Les Olympiades



3.4. Titane



Annexe 4 – Les différentes catégories pour les cases_id dans le cadre du film Boite noire

Catégorie	Rôle	Nom	Pseudo Twitter
Personnes impliquées dans le film	Acteur	Pierre Niney	@pierreniney
	Société de distribution	Studio Canal	@STUDIOCANAL
	Chaîne de télévision du groupe Canal +	Canal +	@canalplus
		Canal + Group	@canalplusgroupe
	Canal + cinéma	@canalpluscinéma	
Salles de cinéma	Cinéma Pathe Gaumont	@PatheGaumont	
Medias	Site fournissant des information sur le cinéma	Allo Ciné	@allociné
	Communauté des membres les plus actifs d'allo ciné	Le club de 300 Allo ciné	@club300allocine
	Journaliste cinéma	Mehdi Omais	@MehdiOmais
	Journaliste cinéma	Stéphane Boudsocq	@Stephhbou08
	Rédactrice Podcast Cinéma	Yael Yermia	@yaelyermia
	Youtubeur axé cinéma	Clapman	@Clapman_YT
	Emission sur la culture de la France	Laissez-vous tenter RTL	@LVT-RTL
	Animateur RTL France	Eric Dussart	@E-Dussart
	Emission	On refait la télé	@OnRefaitLaTele
	Site de critiques cinématographique	Ciné chez moi	@cinechezmoi
	Site de critiques cinématographique	Pettri	@PETTRIllesite
	Podcast sur le cinéma	24FPS	@24FPSpodcast
	Blog et chaîne youtube sur le cinéma	Ciné-O-maX	@cineo_max
	Magazine ciné et podcast	L'infini Détail	@infiniDetail
	Site consacré au cinéma	Cinéphiles44	@cinephiles44
	Présentateur podcast sur le cinéma : La dernière séance	Antoine	@Antoine_mon
	Youtubeur cinéma	RodMovies	@MisteRRod
	Podcast cinéma	Cinecast	@cinecastbe
	Emission Radio Cinema	Tea Time Ciné	@teamtimecine
	Blog sur le cinéma	La tentation culturelle	@LaTentaCulture
	Média d'actualité cinésérie	Watch Buddy	@WatchBuddy
Blog sur le cinéma	CinéVerse	@CineverseOff	
Utilisateurs	Tous les autres comptes		

Annexe 5 – Code PM4Py

Importation des données + nombre d'événement log et nombre de case_id :

1. Boite noire

```
pip install pm4py

import pandas

def import_csv(file_path):
    event_log = pandas.read_csv(file_path, sep=';')
    num_events = len(event_log)
    num_cases = len(event_log.case_id.unique())
    print("Number of events: {}\nNumber of cases: {}".format(num_events,
num_cases))

import_csv("C:/Users/grail/OneDrive/Documents/Mémoire/BoiteNoire/Excel/CSV/
Event Log")
```

Réponse console :

```
Number of events: 2886
Number of cases: 2137
```

2. Illusions perdues

```
import_csv("C:/Users/grail/OneDrive/Documents/Mémoire/Données/Films/Illusio
ns Perdues/Excel/Eventlog_Illusionsperdues.csv")
```

Réponse console :

```
Number of events: 2137
Number of cases: 1476
```

3. Benedetta

```
import_csv("C:/Users/grail/OneDrive/Documents/Mémoire/Données/Films/Benedet
ta/Excel/Eventlog_Benedetta.csv")
```

Réponse console :

```
Number of events: 3228
Number of cases: 2120
```

4. L'évènement

```
import_csv("C:/Users/grail/OneDrive/Documents/Mémoire/Données/Films/L'évène
ment/Excel/Eventlog_evenement.csv")
```

Réponse console :

```
Number of events: 1615
Number of cases: 1119
```

5. Les Olympiades

```
import_csv("C:/Users/grail/OneDrive/Documents/Mémoire/Données/Films/Les  
olympiades/Excel/Eventlog_Olympiades.csv")
```

Réponse console :

```
Number of events: 927  
Number of cases: 678
```

6. Titane

```
import_csv("C:/Users/grail/OneDrive/Documents/Mémoire/Données/Films/Titane/Excel/Eventlog_Titane.csv")
```

Réponse console :

```
Number of events: 8342  
Number of cases: 4980
```

Début et fin d'activité pour tous les event log :

1. Boite Noire

```
def import_csv(file_path):  
    event_log = pandas.read_csv(file_path, sep=';')  
    event_log = pm4py.format_dataframe(event_log, case_id='case_id',  
activity_key='activity', timestamp_key='timestamp')  
    start_activities = pm4py.get_start_activities(event_log)  
    end_activities = pm4py.get_end_activities(event_log)  
    print("Start activities: {}\nEnd activities:  
{}".format(start_activities, end_activities))
```

Réponse console :

```
Start activities: {'Concours': 1500, 'Statistiques': 161, 'Promo': 133,  
'Avant-premiere': 79, 'Critique': 74, 'Tournage': 70, 'Bande-annonce':  
44, 'Cesar': 33, 'Affiche': 26, 'Support materiel': 8, 'Oscars': 4,  
'Canne': 4, 'Plateforme': 1}
```

```
End activities: {'Concours': 1514, 'Statistiques': 171, 'Promo': 129,  
'Critique': 75, 'Avant-premiere': 64, 'Tournage': 64, 'Cesar': 45,  
'Bande-annonce': 39, 'Affiche': 16, 'Support materiel': 11, 'Oscars':  
5, 'Canne': 4}
```

2. Illusions perdues

Réponse console :

```
Start activities: {'Concours': 755, 'Cesar': 282, 'Statistiques':  
142, 'Bande-annonce': 96, 'Critique': 72, 'Affiche': 66, 'Avant-  
premiere': 27, 'Tournage': 19, 'Promo': 11, 'Support materiel': 4,  
'Plateforme': 2}
```

```
End activities: {'Concours': 797, 'Cesar': 303, 'Statistiques': 143,  
'Bande-annonce': 82, 'Critique': 67, 'Affiche': 36, 'Avant-premiere':  
16, 'Tournage': 14, 'Promo': 9, 'Support materiel': 7, 'Plateforme': 2}
```

3. Benedetta

Réponse console :

```
Start activities: {'Canne': 809, 'Concours': 564, 'Bande-annonce': 305, 'Cesar': 114, 'Critique': 107, 'Affiche': 68, 'Promo': 50, 'Tournage': 34, 'Support materiel': 21, 'Affiche': 17, 'Statistique': 16, 'Avant premiere': 9, 'Oscars': 6}
```

```
End activities: {'Canne': 811, 'Concours': 558, 'Bande-annonce': 254, 'Cesar': 138, 'Critique': 129, 'Promo': 59, 'Affiche': 57, 'Support materiel': 27, 'Affiche': 26, 'Statistique': 26, 'Tournage': 25, 'Oscars': 6, 'Avant premiere': 4}
```

4. L'évènement

Réponse console :

```
Start activities: {'Concours': 743, 'Oscars': 96, 'Canne': 89, 'Cesar': 84, 'Critique': 36, 'Bande-annonce': 19, 'Avant-premiere': 17, 'Promo': 15, 'Support materiel': 13, 'Statistiques': 8, 'Affiche': 4, 'Tournage': 1}
```

```
End activities: {'Concours': 763, 'Cesar': 97, 'Canne': 82, 'Oscars': 78, 'Critique': 34, 'Promo': 20, 'Support materiel': 16, 'Avant-premiere': 14, 'Bande-annonce': 11, 'Statistiques': 8, 'Tournage': 1, 'Affiche': 1}
```

5. Les Olympiades

Réponse console :

```
Start activities: {'Concours': 231, 'Canne': 214, 'Promo': 46, 'Critique': 45, 'Avant-premiere': 43, 'Affiche': 34, 'Cesar': 33, 'Bande-annonce': 24, 'Statistiques': 24, 'Tournage': 5, 'Support materiel': 2}
```

```
End activities: {'Concours': 239, 'Canne': 199, 'Critique': 52, 'Avant-premiere': 45, 'Cesar': 41, 'Promo': 40, 'Affiche': 38, 'Statistiques': 30, 'Bande-annonce': 12, 'Tournage': 3, 'Support materiel': 2}
```

Premier processus :

```
#Storing a Pandas DataFrame as a .xes file
```

```
event_log =  
pm4py.format_dataframe(pandas.read_csv("C:/Users/grail/OneDrive/Documents/Mémoire/BoiteNoire/Excel/CSV/Event Log.csv", sep=';'), case_id='case_id',  
activity_key='activity', timestamp_key='timestamp')  
pm4py.write_xes(event_log,  
"C:/Users/grail/OneDrive/Documents/Mémoire/BoiteNoire/Excel/CSV/Event log.xes")
```

```
log =  
pm4py.read_xes("C:/Users/grail/OneDrive/Documents/Mémoire/BoiteNoire/Excel/CSV/Event log.xes")
```



```
process_tree = pm4py.discover_tree_inductive(log)
bpmn_model = pm4py.convert_to_bpmn(process_tree)
pm4py.view_bpmn(bpmn_model)
```

Process interne :

```
def import_csv(file_path):
    event_log = pandas.read_csv(file_path, sep=';')
    num_events = len(event_log)
    num_cases = len(event_log.case_id.unique())
    print("Number of events: {}\nNumber of cases: {}".format(num_events,
num_cases))
```

```
import_csv("C:/Users/grail/OneDrive/Documents/Mémoire/BoiteNoire/Excel/CSV/
Event Log_interne.csv")
```

Réponse console :

```
Number of events: 38
Number of cases: 4
```

Process media

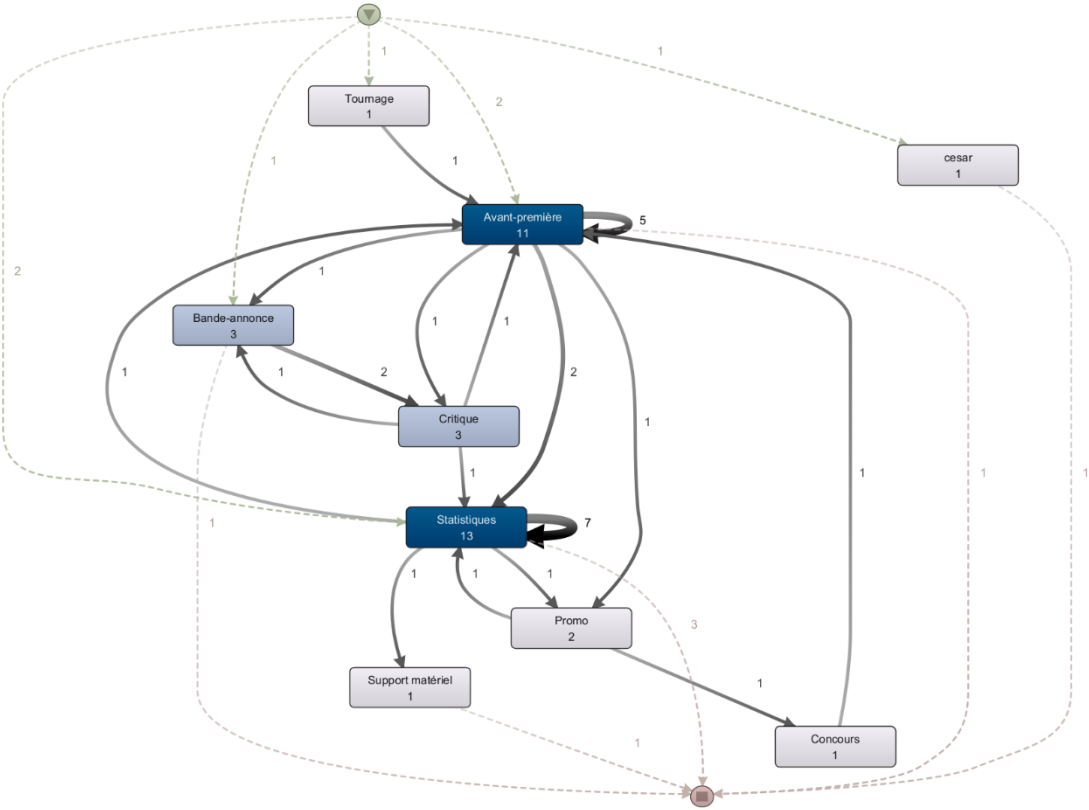
import pandas

```
def import_csv(file_path):
    event_log = pandas.read_csv(file_path, sep=';')
    num_events = len(event_log)
    num_cases = len(event_log.case_id.unique())
    print("Number of events: {}\nNumber of cases: {}".format(num_events,
num_cases))
```

```
import_csv("C:/Users/grail/OneDrive/Documents/Mémoire/BoiteNoire/Excel/CSV/
Event Log_media.csv")
```

Annexe 6 – Processus Boite Noire

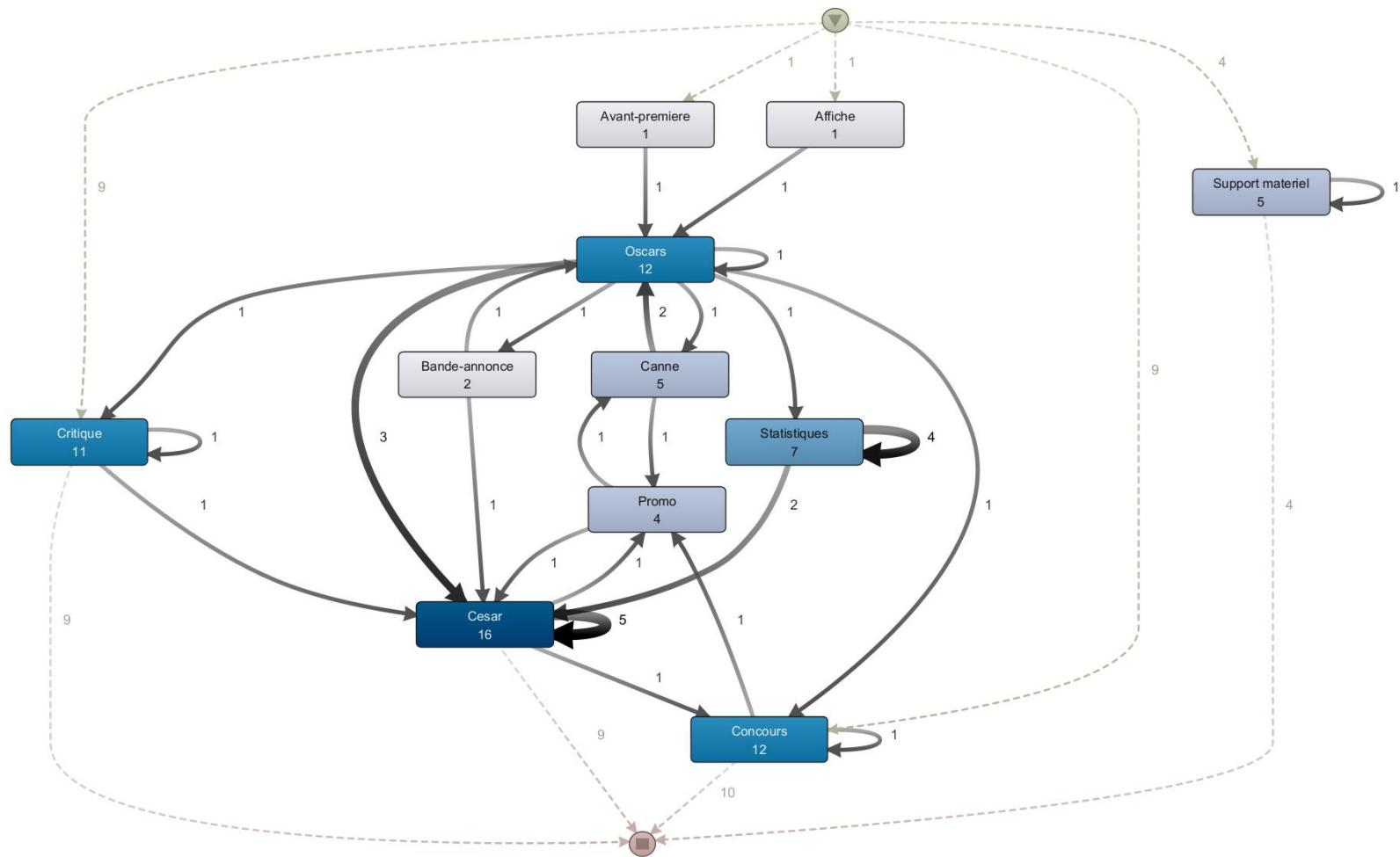
6.1.Processus parties prenantes Boite noire



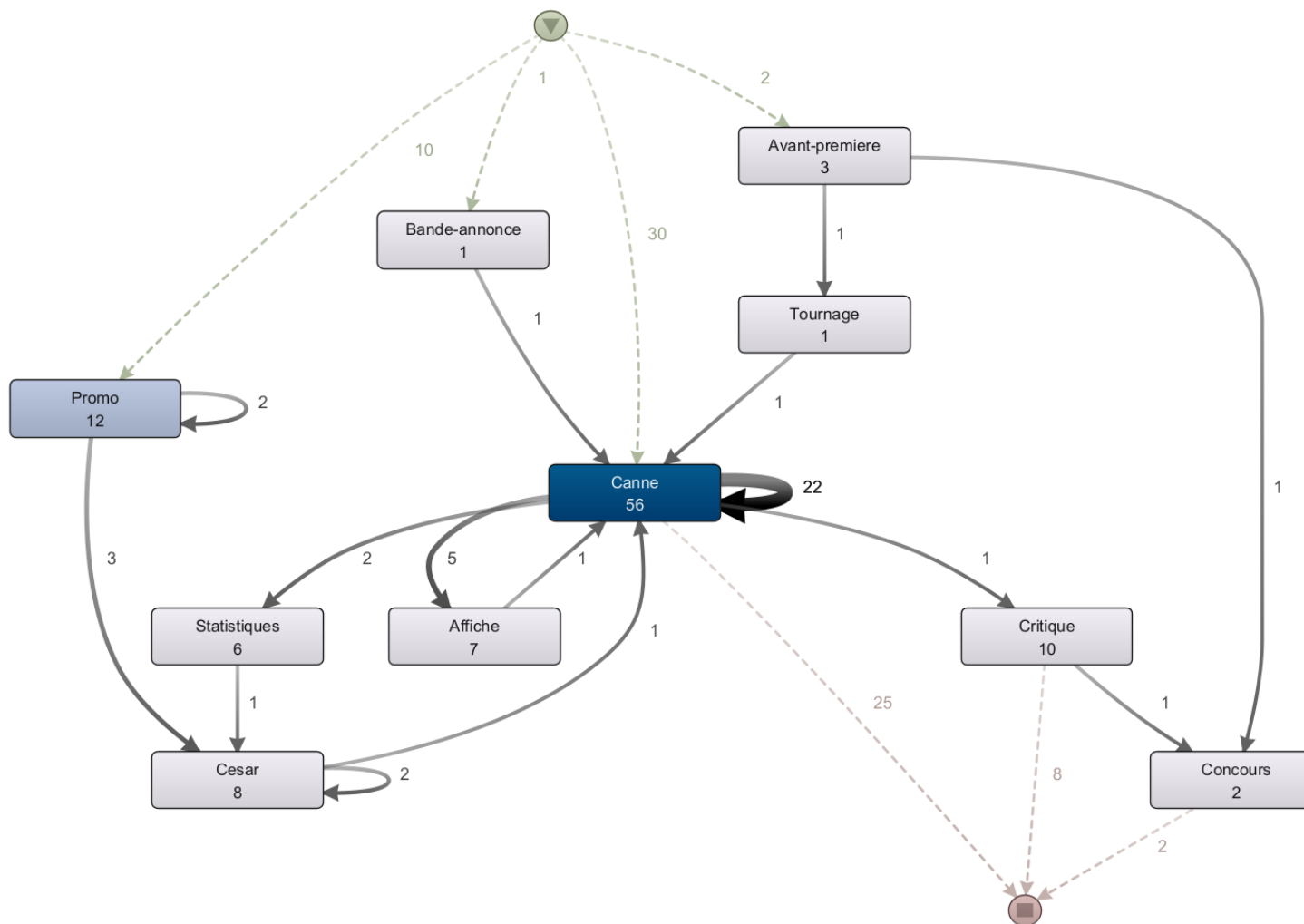
6.2. Tableau des variants

Variant	Cases	Events	Step 1	Step 2	Step 3	Step 4	Step 5	Step 6	Step 7	Step 8	Step 9
1	12	1	Critique								
2	7	1	Promo								
3	4	1	Avant-premiere								
4	3	2	Promo	Promo							
5	3	1	Statistiques								
6	2	2	Promo	Statistiques							
7	2	1	Affiche								
8	1	3	Promo	Cesar	Cesar						
9	1	5	Tournage	Bande-annon	Statistiques	Avant-premi	Cesar				
10	1	10	Promo	Promo	Statistiques	Statistiques	Statistiques	Statistiques	Statistiques	Statistiques	Statistiques
11	1	4	Statistiques	Statistiques	Statistiques	Statistiques					
12	1	3	Critique	Critique	Critique						
13	1	4	Concours	Promo	Statistiques	Statistiques					
14	1	2	Avant-premi	Statistiques							
15	1	4	Promo	Concours	Promo	Promo					
16	1	4	Concours	Promo	Promo	Concours					
17	1	2	Bande-anno	Statistiques							
18	1	9	Avant-premi	Statistiques	Avant-premi	Avant-premi	Avant-premi	Avant-premi	Avant-premi	Avant-premi	Avant-premi
19	1	2	Critique	Critique							
20	1	3	Promo	Promo	Promo						
21	1	3	Avant-premi	Affiche	Bande-annonce						
22	1	1	Bande-annonce								
23	1	3	Affiche	Affiche	Bande-annonce						
24	1	3	Affiche	Critique	Affiche						

7.3.L'évènement



7.4. Les olympiades



7.5.Titane

