



Projet Data Mining

Pierre TURPIN, Jean-Marie COMETS

25 mars 2014

Table des matières

1	Caractérisation du flux vidéo	3
1.1	Popularité	3
1.2	Catégorisation	3
1.3	Les jeux vidéo populaires	3
2	Prédiction de l'audience d'un flux	5
3	Classement des "meilleurs" joueurs	5

Au vu des nombreux problèmes de "scaling" que nous pouvions rencontrer avec le jeu de données prévu, l'intégralité de ce rapport repose sur l'analyse d'un échantillon de données. Bien entendu, avant d'étudier les données, nous avons mélangé le jeu de données et pris un échantillon fixe pour l'ensemble de l'étude.

1 Caractérisation du flux vidéo

1.1 Popularité

En étudiant les différents indicateurs quantitatifs (attribus *_count), représentant le nombre de visionnages d'un flux, nous avons pu remarquer que quasiment toutes celles-ci sont indépendantes, mis à part l'attribut *stream_count*, qui ne représente que la somme du *embedded_count* et du *site_count*.

Nous avons donc retiré cet attribut, pour pouvoir définir la notion de **popularité** d'un flux, correspondant à une somme normalisée des différents indicateurs.

Cet indicateurs nous sert d'heuristique pour établir les différentes catégorisations suivantes :

1.2 Catégorisation

Un seul attribut permettant de catégoriser les différents flux est disponible, et ce uniquement à partir du jeu de données XML : *subcategory*. En remarquant que cet attribut concerne à la fois la catégorie du jeu et sa plateforme, nous avons séparé ces deux informations.

Ainsi, dans la figure 2, nous n'observons pas la plateforme PC, qui devrait cependant regrouper beaucoup de flux (plateforme non définie là où le type de jeu est bien défini). Cependant, nous pouvons remarquer qu'entre les différentes consoles de jeu, c'est la plateforme **XBOX** qui a le plus de succès.

Nous remarquons à partir de la figure 1 donc que la catégorie **strategy** se détache des autres, prenant plus de 60 % de la part du nombre de vues des flux. Ceci correspond bien à nos attentes, vu que Twitch est principalement connu pour des jeux développés pour PC, avec une préférence pour les jeux de stratégie/roleplay (Starcraft, League of Legends, etc...).

1.3 Les jeux vidéo populaires

Dans le jeu de données utilisé, la plupart des jeux étaient en doublons car ils n'étaient pas tous orthographiés de la même façon (majuscules/minuscules, espaces, ...). Nous n'avons pas pu, à cause de la taille des données, corriger toutes les entrées afin d'unifier

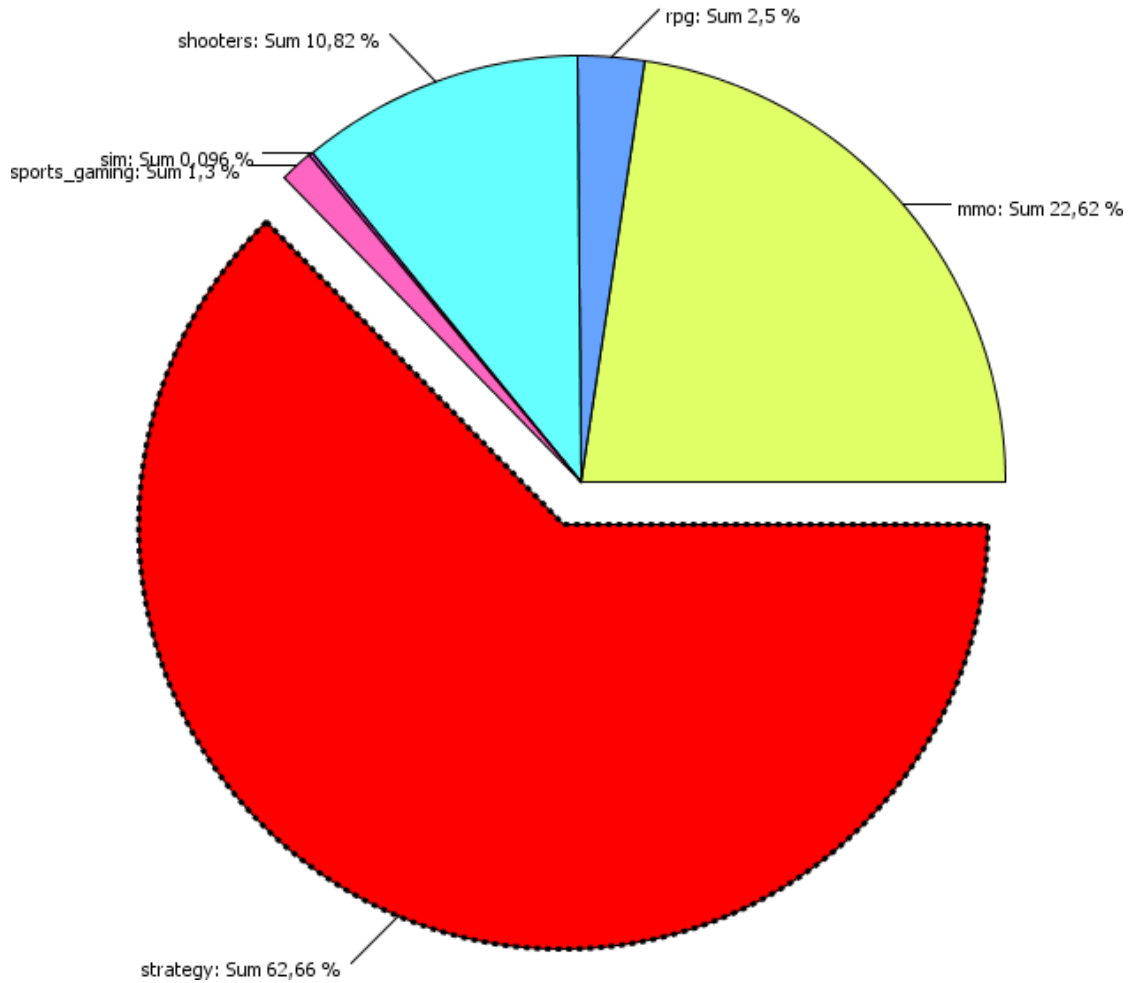


FIGURE 1 – Part des vues selon la catégorie du jeu

l'écriture des jeux. Les résultats ne sont donc pas complètement exacts mais donnent tout de même une bonne approximation de la réalité.

En tout il y a 160 jeux différents en comptant les doublons. Nous avons établi la part de vue des différents jeux en groupant sur le champ *meta-name* et en sommant l'heuristique de popularité. Comme une grande quantité de jeu était très minoritaire selon notre heuristique, nous avons regroupé ces derniers (en seuillant la popularité) dans une seule catégorie *misc*.

La figure 3 représente la popularité de chaque jeux. L'ensemble des jeux *misc* forment 19% de popularité tandis que 10 autres jeux prennent les 80% restant. Il y a donc une très grande disparité dans les jeux vidéo et une petite minorité de 10 jeux écrasent totalement 150 autres jeux.

Le tableau 1 montre alors un classement des jeux les plus populaires sur la plateforme

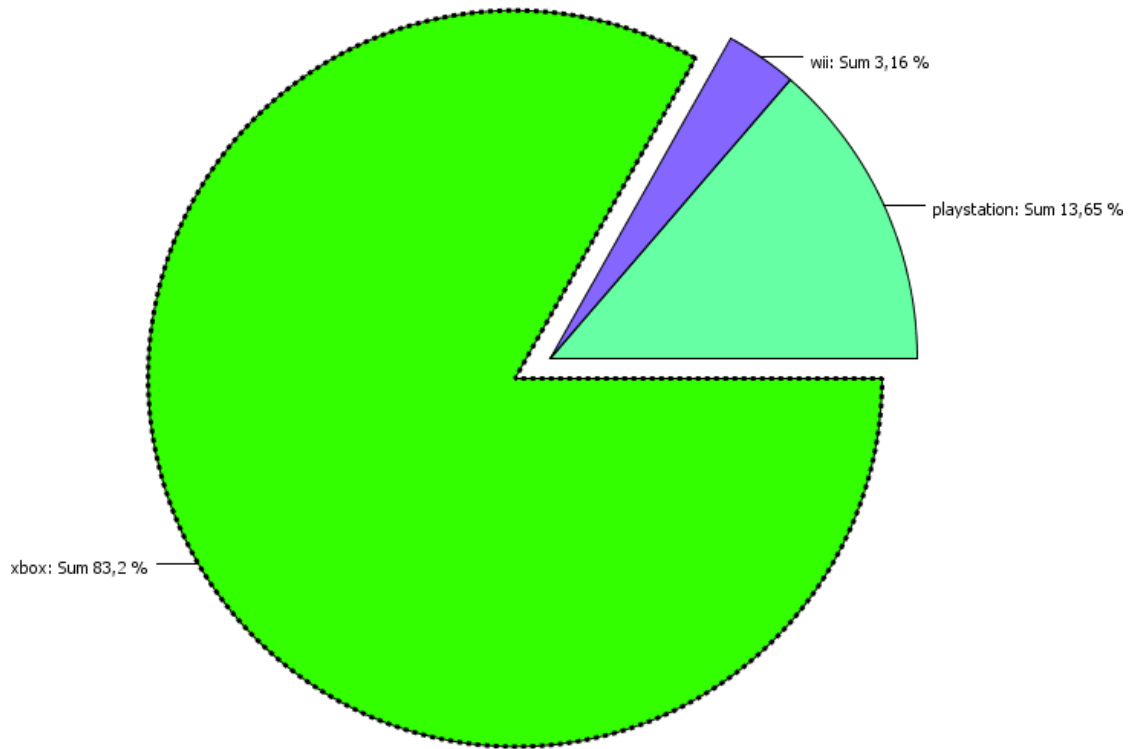


FIGURE 2 – Part des vues selon la plateforme (console) du jeu

Twitch.

2 Prédiction de l'audience d'un flux

3 Classement des "meilleurs" joueurs

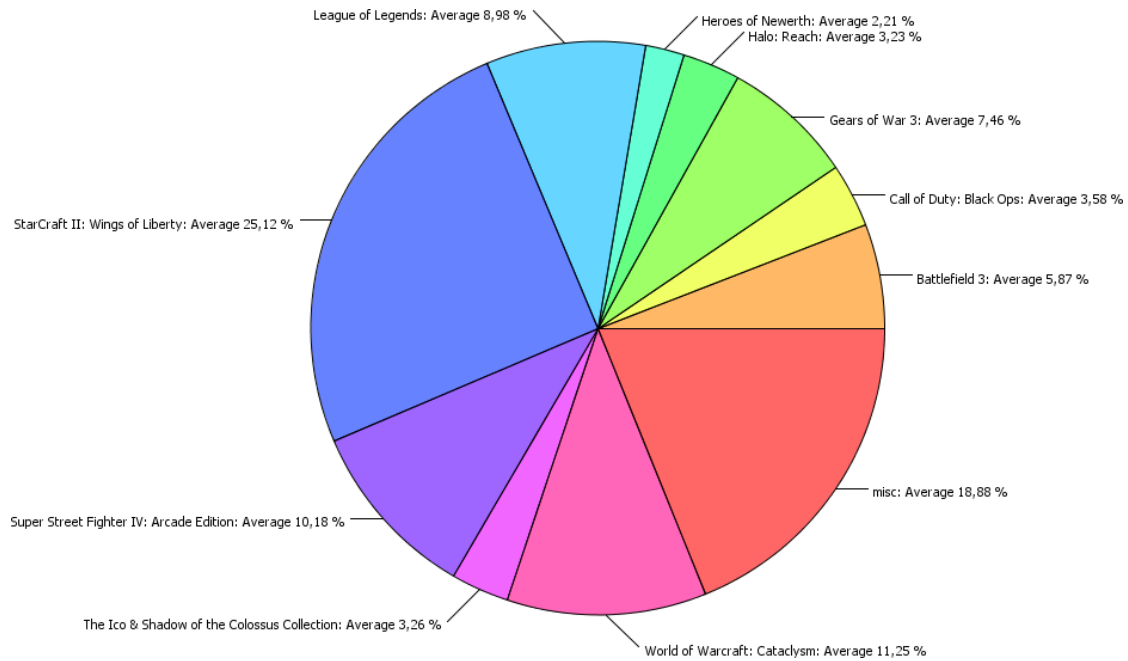


FIGURE 3 – Part de vue en moyenne des jeux vidéo

TABLE 1 – Classement des 10 jeux les plus populaires sur la plateforme Twitch.

Position	Jeux vidéo	Part de popularité (en %)
1	StarCraft II	25.12
2	World of Warcraft : Cataclysm	11.25
3	Super Street Fighter IV	10.18
4	League of Legends	8.98
5	Gears of War	7.46
6	Battlefield 3	5.87
7	Call of Duty : Black Ops	3.58
8	The Ico & Shadow of Colossus Collection	3.26
9	Halo : Reach	3.23
10	Heroes of Newerth	2.21

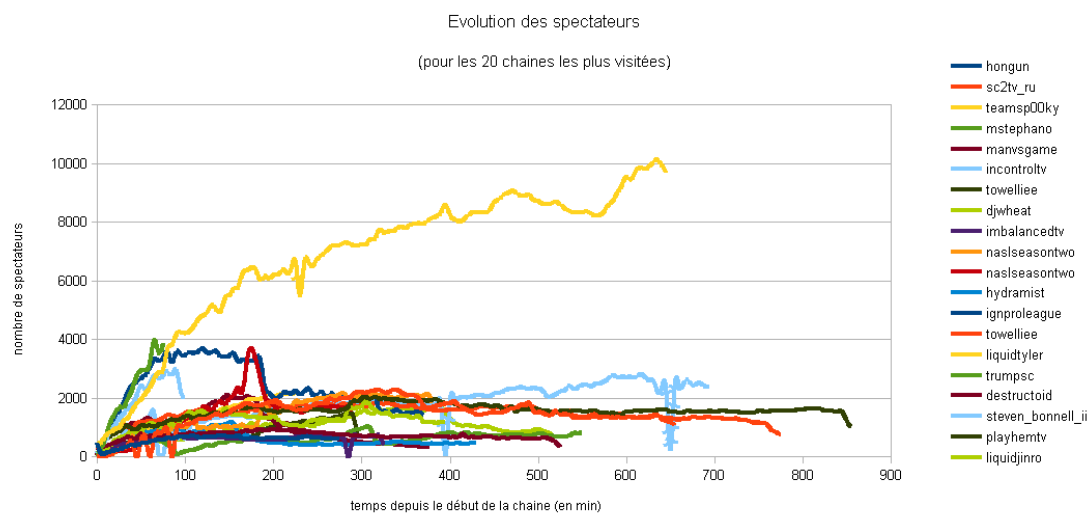


FIGURE 4 –

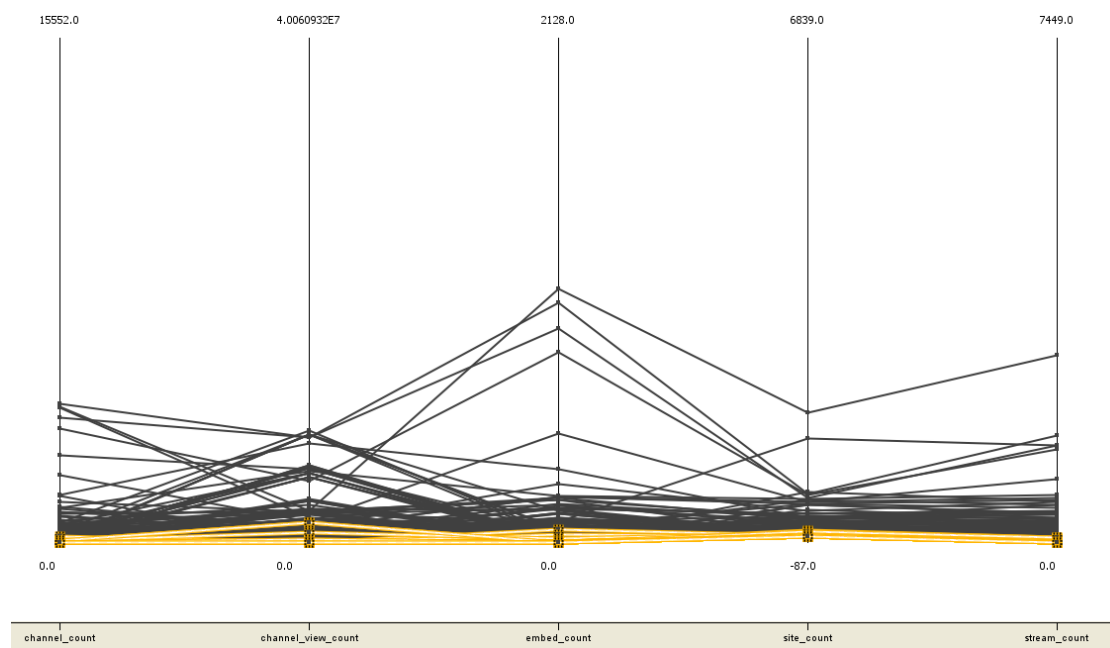


FIGURE 5 –

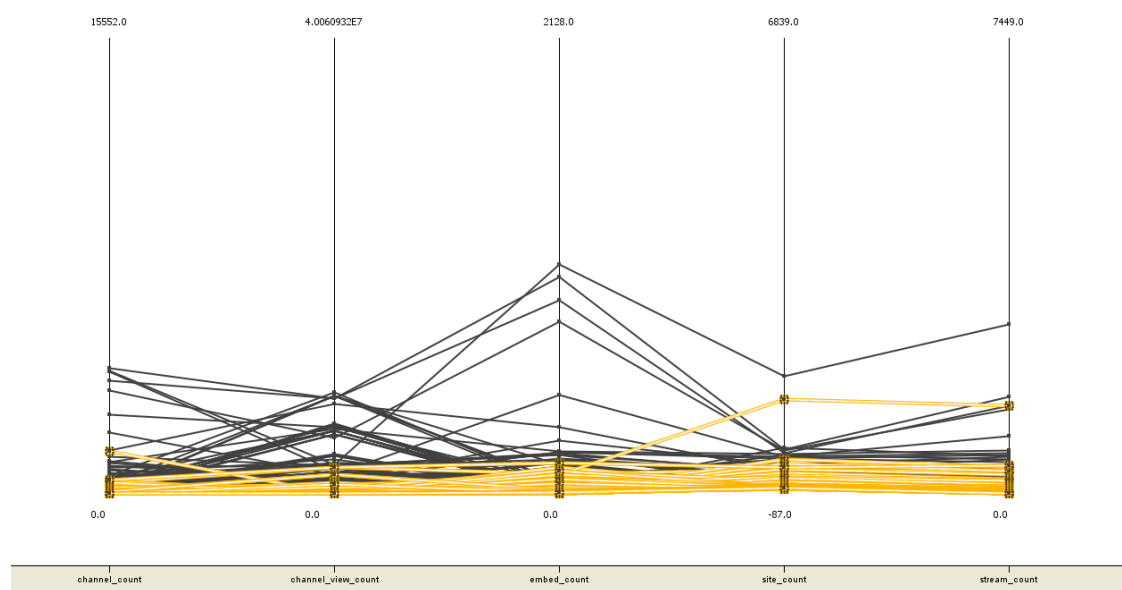


FIGURE 6 –