



## Projet Data Mining

Pierre TURPIN, Jean-Marie COMETS

25 mars 2014

## Table des matières

## 1 Sujet (à enlever)

Your goal is to give a longitudinal characterization of this new community by analyzing Twitch audiences. Previously, you crawled the list of active live video streams along with their respective number of viewers every five minutes from September 29th, 2011 to January 09th, 2012. Your data analysis should enable :

1. to characterize video streams qualitatively (identifying the games and the player location) and quantitatively through their viewer counts, durations, and audience
2. to early predict the audience of a stream
3. to rank the most popular players. These results are of major interest for all actors of this community. For example, popularity is key in a pro-gamer career, strongly influencing his revenues (sponsors, invitations to tournament with prizes and advertisement revenues while streaming).

## 2 Rapport

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 fixé des données, soit 5 000 streams.

### 2.1 Caractérisation du flux vidéo

*qualitatif : localisations des joueurs*

*qualitative : mettre graphique evolution (nombre de vues, durées)*

### 2.2 Prédiction de l'audience d'un flux

TODO

### 2.3 Classement des "meilleurs" joueurs

*classement par tableau/graphe*