

---

# Project Proposal

## Network tour of Data Science - Autumn 2017

Olivier Couque, Emma Lejal Glaude, Matthieu Sauvé

---

### Idea

We would like to explore the network of American Basketball Players. The NBA players are well known for their presence on Twitter which provides us a way to analyze the existing connections between players. Once we have a graph with the connections, we will be able to run clustering algorithms and discover if we can recognize the teams, the divisions or the conferences. There are 30 different teams, grouped in 6 different divisions (partitioned in roughly North-Center-South) again grouped in 2 different divisions (partitioned by East-West). At first glance, as we already know the number of clusters, k-means could be a good start.

Afterwards we could imagine that the graph also shows the « All Stars » (i.e. the teams composed of best players in both conferences playing one special game).

Finally if we still have time, we could try to extract the trades. For example, if a player A is highly connected to team T1 and team T2, can we verify that he changed from one team to the other ?

### Data Retrieval

The first dataset that we found is website with a list of all players with their Twitter account and with their team (Source : <http://hoopeduponline.com/2009/03/30/a-list-of-every-nba-player-on-twitter/>). This would have been the perfect dataset for us if it was up to date, unfortunately it was from 2013. As the NBA is changing a lot every year we cannot consider this dataset as it is.

In order to have the current composition of teams, we base ourselves on the Wikipedia page of the NBA teams (Source : [https://en.wikipedia.org/wiki/List\\_of\\_current\\_NBA\\_team\\_rosters](https://en.wikipedia.org/wiki/List_of_current_NBA_team_rosters)). We used the python library BeautifulSoup to retrieve this information in a Pandas Dataframe. On the personal pages of the players, some of them have their twitter account at the end and this is another strategy to get the Twitter account for the current players that are missing in our 2013 data.