## Assignement : "Detail analysis and/or modeling of a network dataset(s)"

## Project : Predict the virality of memes/hashtags in a social network (Twitter)

**Team members :**

* BOUCHNAIF Samy : [Samy.BOUCHNAIF@ensae-paristech.fr](mailto:Samy.BOUCHNAIF@ensae-paristech.fr)
* BROCHARD Maxence : [Maxence.BROCHARD@ensae-paristech.fr](mailto:Maxence.BROCHARD@ensae-paristech.fr)
* TRAN Duc-Vinh : [duc.vinh.TRAN@ensae-paristech.fr](mailto:duc.vinh.TRAN@ensae-paristech.fr)

**References:**

* Dataset description : Sampled public tweets from Twitter streaming API between March and April 2012
* Dataset source : <http://carl.cs.indiana.edu/data/#virality2013>
* References : <https://arxiv.org/abs/1403.6199>

**Context and motivations :**

Virality is, in social networks, an important issue for corporations, political campaigns and influencers as they spend enormous resources and efforts to make their products or messages go viral in order to catch attention and spread their influence/activities to a wider audience.

Thus, understanding the complex mechanism of virality may help one control its effects over the network:

* How does the network structure affect the diffusion?
* How to model the contagion, etc.

Proposal brought by the paper : the broad idea is that network communities allow predict virality by its early spreading pattern. A simple, popular approach in studying hashtags diffusion is to consider hashtags as diseases and apply epidemic models. However, recent studies demonstrate that diseases and behaviors spread differently.

We can see huge potentiality for applications in social media marketing : social networks could give better advice to their users as to which posts are likely to give best advertising Return on Investment.

**Project phases :**

1. Introduction : Modeling viral propagation through classic random graph structures & study different type of spreading. This will serve as a presentation for next week while being a good opportunity to get familiar with the igraph package.
2. Considering the amount of data and sizes of graphes, a sampling phase will be necessary to conduct before analyzing the graphs. How does this sampling phase will influence our studies and how to take it into account for the modeling.
3. Study the graph structures of the provided dataset and its propagation phenomenon :
   1. Apply community detection (Infomap)
   2. Complex contagion vs regular infection
4. Predict the virality, a machine learning approach:
   1. Define a metric for a hashtag popularity
   2. Extract graph features
   3. Predict.

**Friday’s presentation:**

* Introduction:
  + Motivation
  + References
  + Project phases
* Data Presentation:
  + Sources
  + Description
  + Simple analysis
* Sampling:
  + Why? How?
* Modeling of hashtags spreading
* Prediction of “infection”