# A Network Tour of Data Science

# Project proposal

In this project, Graph Based solutions will be incorporated into Recommendation Systems. The idea is to use graph models to represent the available data (crawled or downloaded, possibly enriched) and try different representations by restructuring the graph, using different data, and changing similarity function between the nodes etc. and finally comparing the performance using the different models.

*ex. "Constructing a network where the nodes are users, and the edges are weighted by different types of similarities between them (possibly thresholded)" or "Constructing a network where the nodes are items, and the edges represent different similarities between them" or "Constructing a network where the nodes correspond to both users and items, while the edges are user's rating for a given item". Finally trying different techniques of generating recommendations and evaluating the effects of the different models on the performance of the framework.*

Our work in finding a dataset, suitable for our application had lead us to a few out of the dozen that we reviewed. Namely we are left with:

1. The Movies Dataset - hosted by Kaggle
2. HetRec 2011

We are yet to analyse in a more detailed fashion the data available in each dataset and subsequently decide concretely which models would be used and what questions would be answered. (To be discussed in the Friday project discussion session.)

## Team members:

Martin Josifoski
Mazen Fouad A-wali Mahdi
Lorenzo Lazzara
Andrea Manzini