## #Project Gender detection in VK

Students: Le Thanh Van Anh, Karataeva Azhar

**##Downloading and data processing**

* Downloading data via VK API - VK\_scraph.ipynb(Link github).

1. id members
2. id friends of members
3. data about members of group and their friends: user\_data\_all.pickle (Link github), user\_connections\_all.pickle (Link github)

* Find and save the largest connected component: connected\_components-all.ipynb(Link github).

1. All male user in the largest connected component: male\_all.txt (Link github)
2. All female user in the largest connected component: female\_all.txt(Link github)
3. Visualization of gender distribution
4. All information of users in the largest connected component: more\_infomation\_user\_data\_all.pickle (Link github)
5. Save graph "G\_subgraph\_all.gz" (Link github)

## ## Classification based on static features

* Implemention: classification\_on\_static\_features.ipynb Link github)
* Best model ROC AUC - 0.65
* Save best model in file: best\_static\_model\_all.pickle Link github)

## ## Classification based on Node2vec

* Implemention: classification\_on\_node2vec\_all.ipynb Link github)
* Best model ROC AUC - 0.52
* Save best model in file: best\_node2vec\_model\_all.pickle Link github)

## ## Classification based on Node2vec embeddings and static features

* Implemention: classification\_on\_node2vec\_and\_static\_features\_all.ipynb Link github)
* Best model ROC AUC - 0.55

**###Summary**

We trained the models with best results:

* Classifiers on static features – 0.65
* Classifiers on the concatenation of static features and Node2Vec embeddings – 0.55
* Classifiers on embeddings Node2Vec – 0.52

**### Future work**

* Get more users
* Collect other information such as followers, subscriptions, public pages, …
* Collect and add image recognition