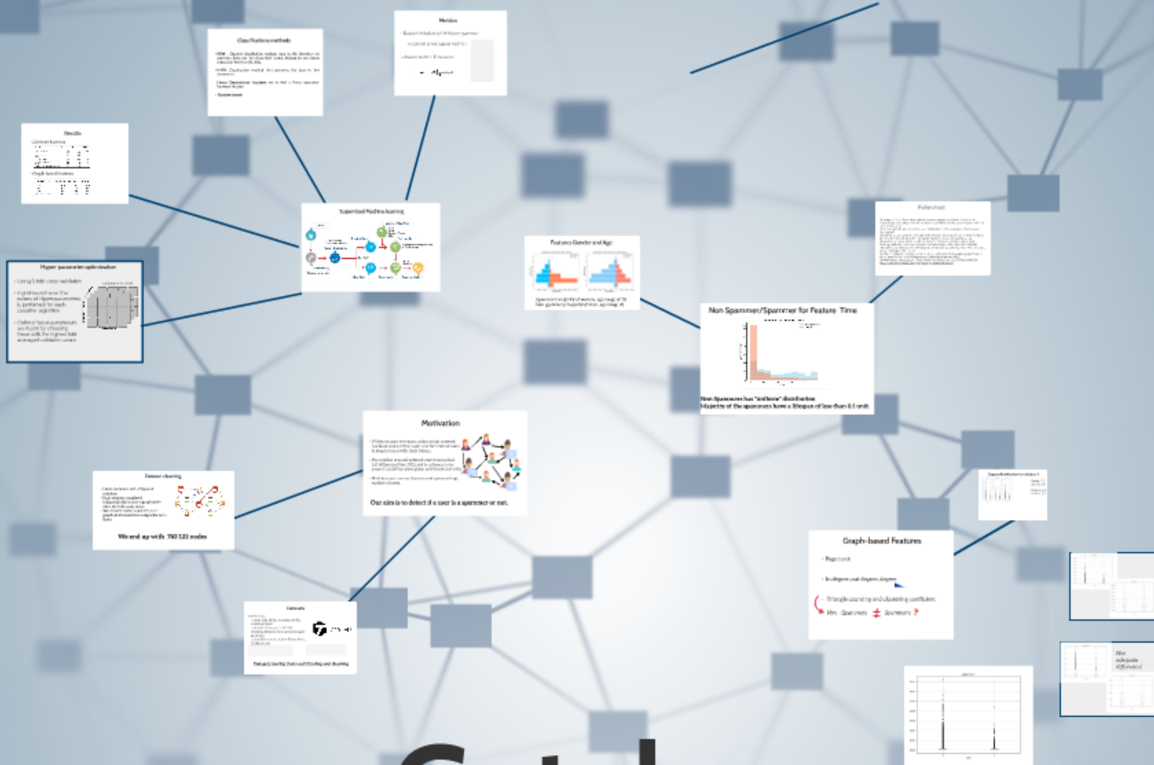
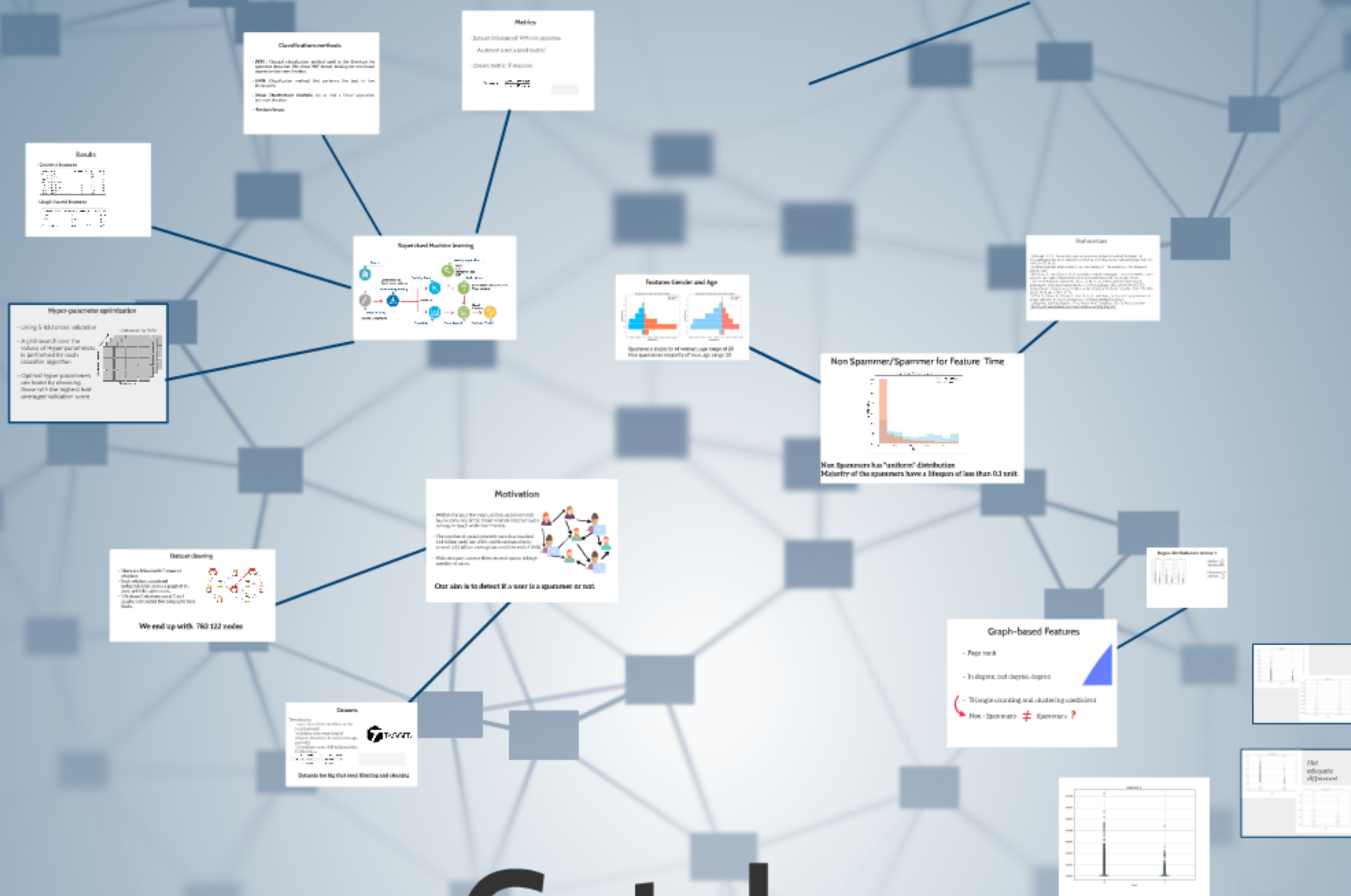


Spammer... Catch me if you can

TEAM 20
Hedi Fendri - Paul Jeha -
Christina Mantonanaki - Nguyet Minh Nguyet

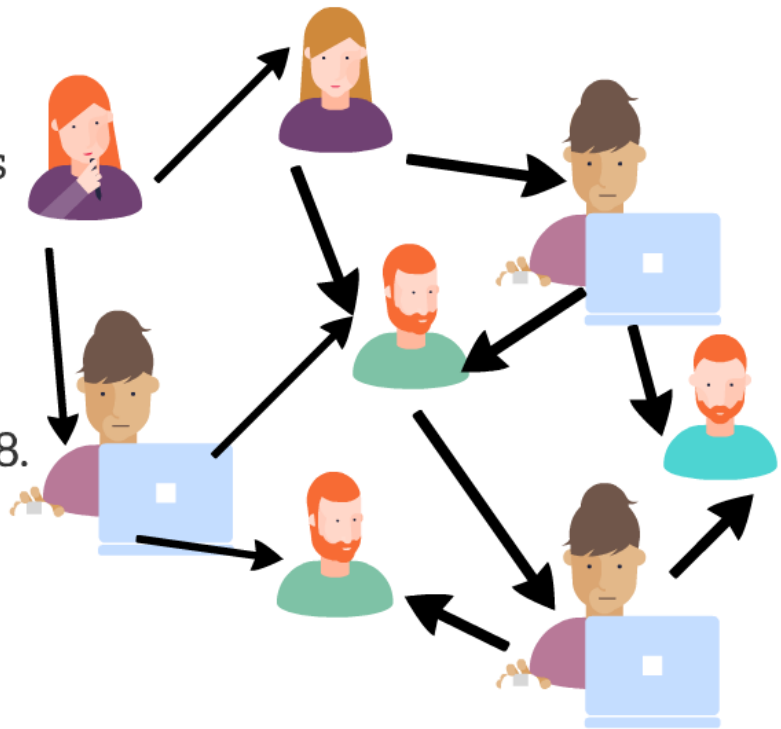


Spammer... Catch me if you can

TEAM 20
Hedi Fendri - Paul Jeha -
Christina Mantonanaki - Nguyet Minh Nguyet

Motivation

- Within the past few years, online social network has become one of the major way for internet users to keep in touch with their friends.
- The number of social network users has reached 1.61 billion until late 2013, and is estimated to be around 2.33 billion users globe, until the end of 2018.
- Malicious part can use them to send spams to huge number of users.



Our aim is to detect if a user is a spammer or not.

Datasets

Two datasets:

- users: lists all the members of the social network
- relations: lists every kind of relations between the users (message, pock etc.)
- 5.6 millions users, 858 millions links, 25 GB of data

TAGGED

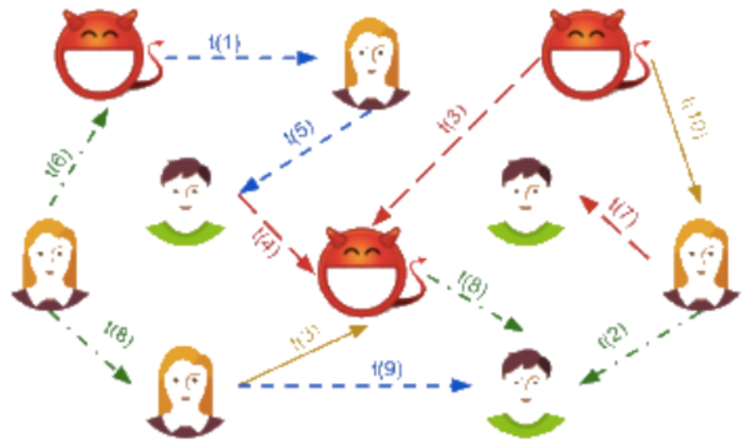
	day	time_ms	src	dst	relation
0	0	7857852	1	3993630	5
1	0	7860977	1	3181660	5

	User Id	Gender	Time	Age Range	Spammer Label
0	1	M	0.9000	30	0
1	190	M	0.2000	50	0

Datasets too big that need filtering and cleaning

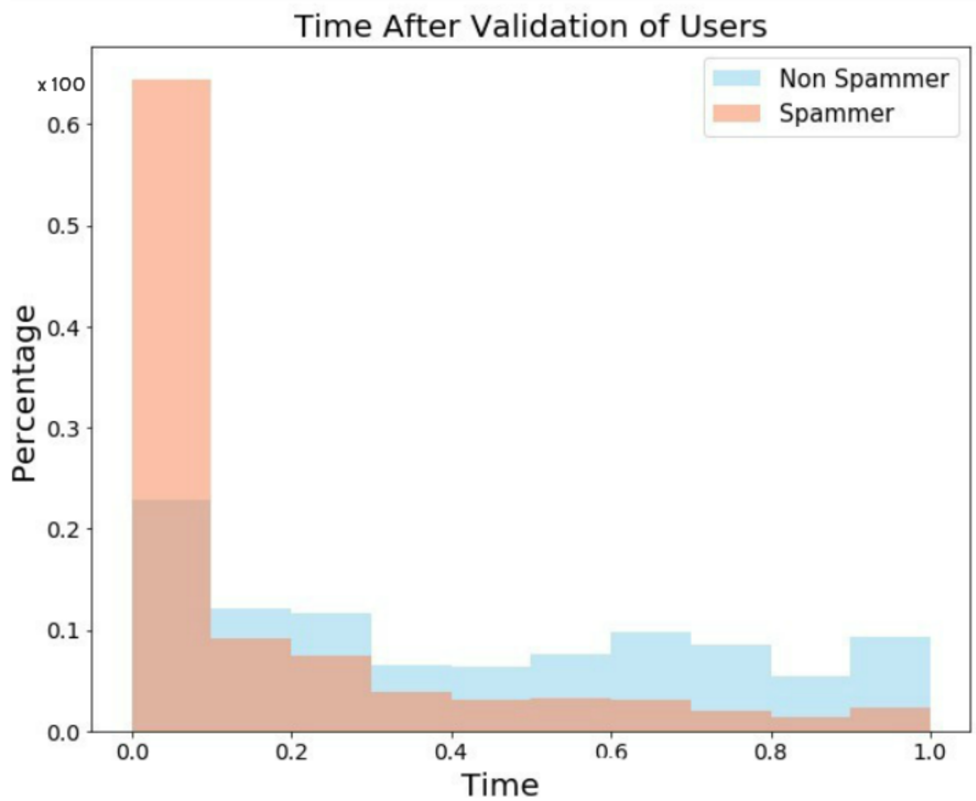
Dataset cleaning

- Users are linked with 7 types of relations
- Each relation considered independently creates a graph of it's own, with the same users
- We chose 5 relations out of 7, so 5 graphs and created five subgraphs from them.



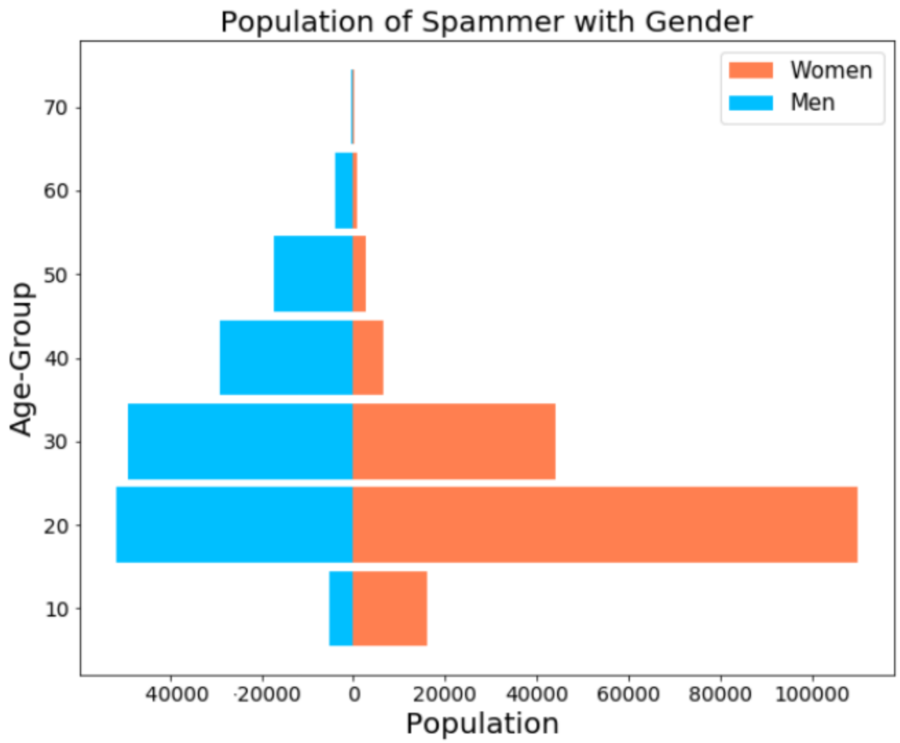
We end up with 783 122 nodes

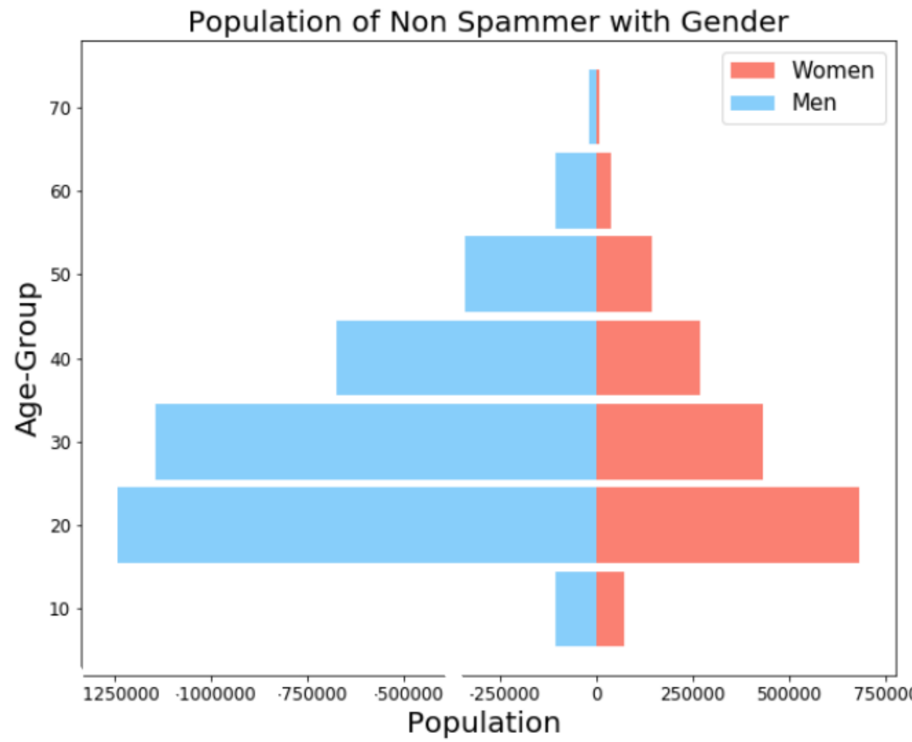
Non Spammer/Spammer for Feature Time



Non Spammers has "uniform" distribution
Majority of the spammers have a lifespan of less than 0.1 unit.

Features Gender and Age

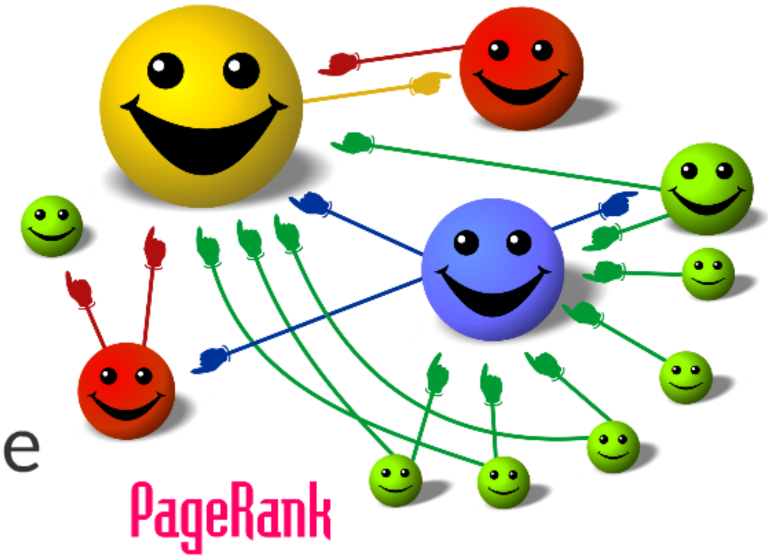




Spammers: majority of women, age range of 20
Non spammers: majority of men, age range 20

Graph-based Features

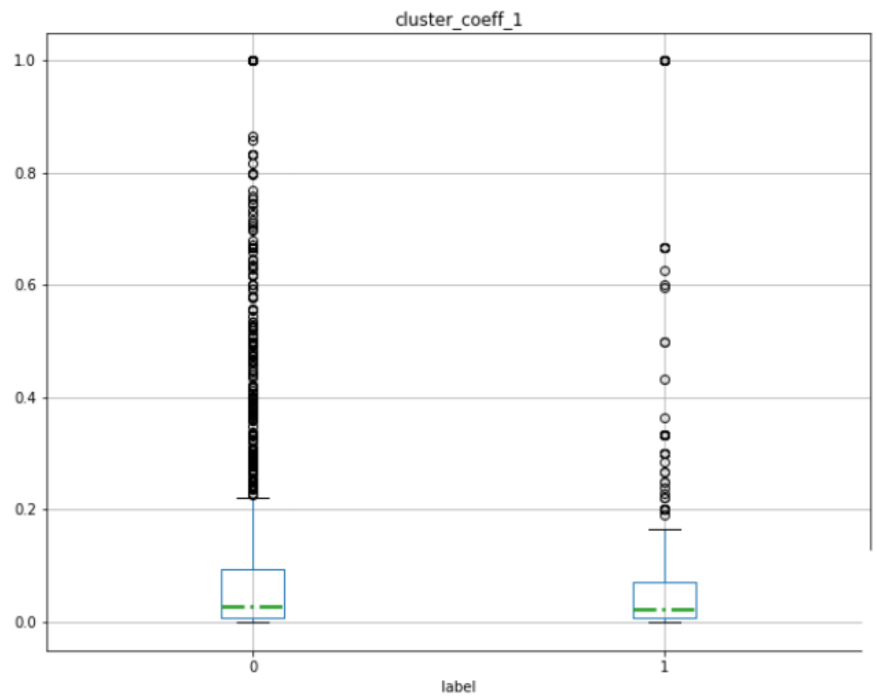
- Page rank



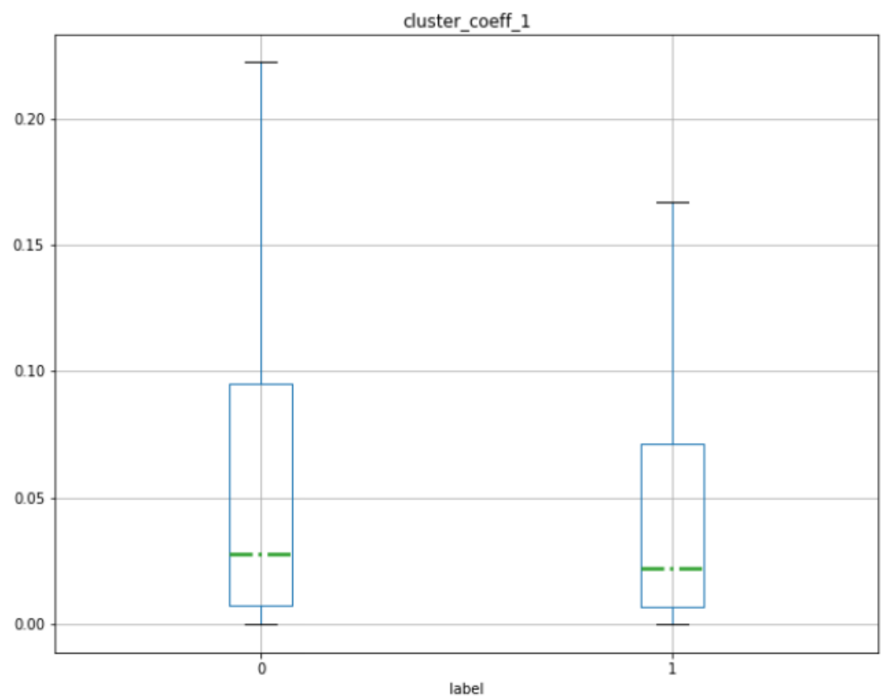
- In degree, out degree, degree

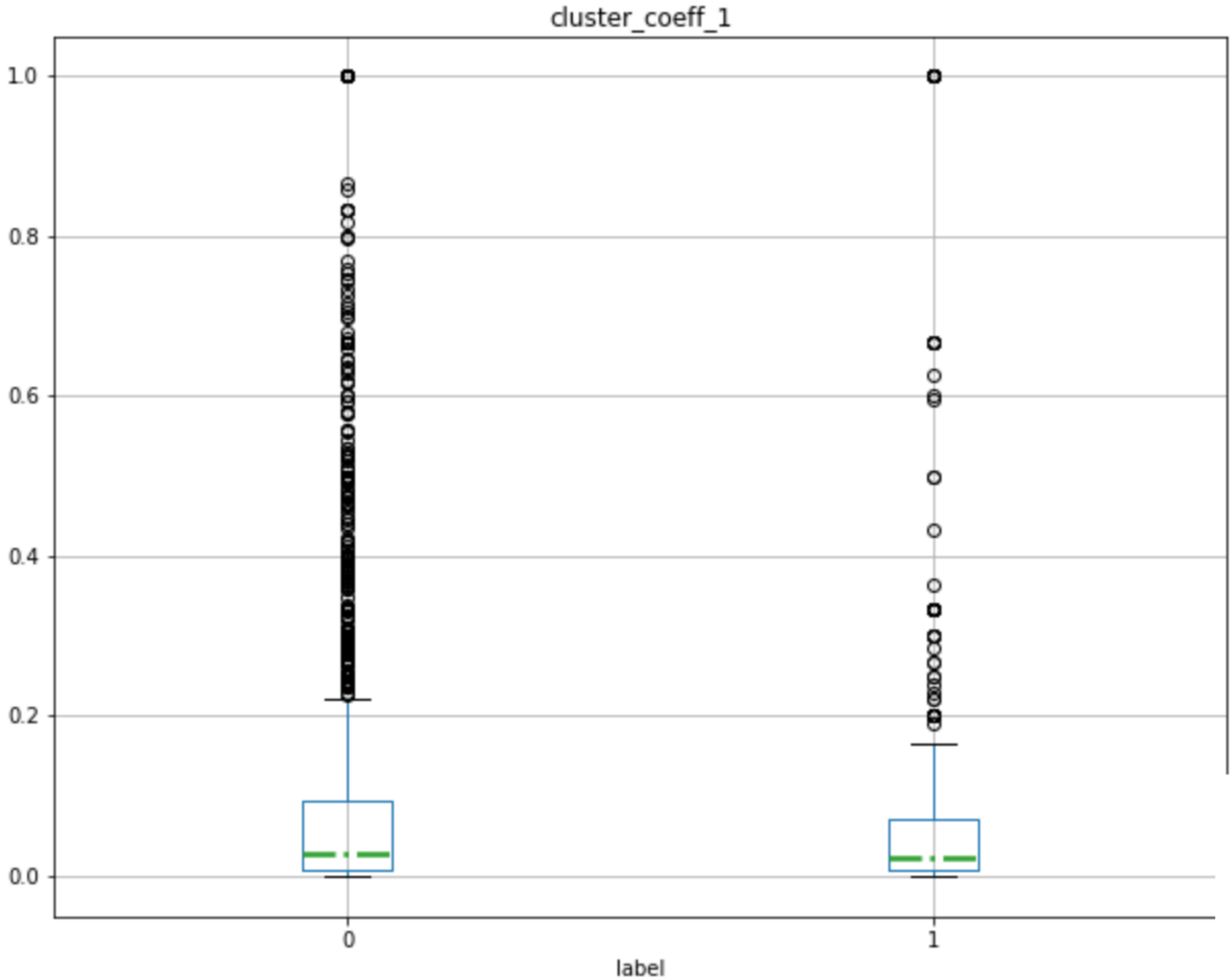
- Triangle counting and clustering coefficient

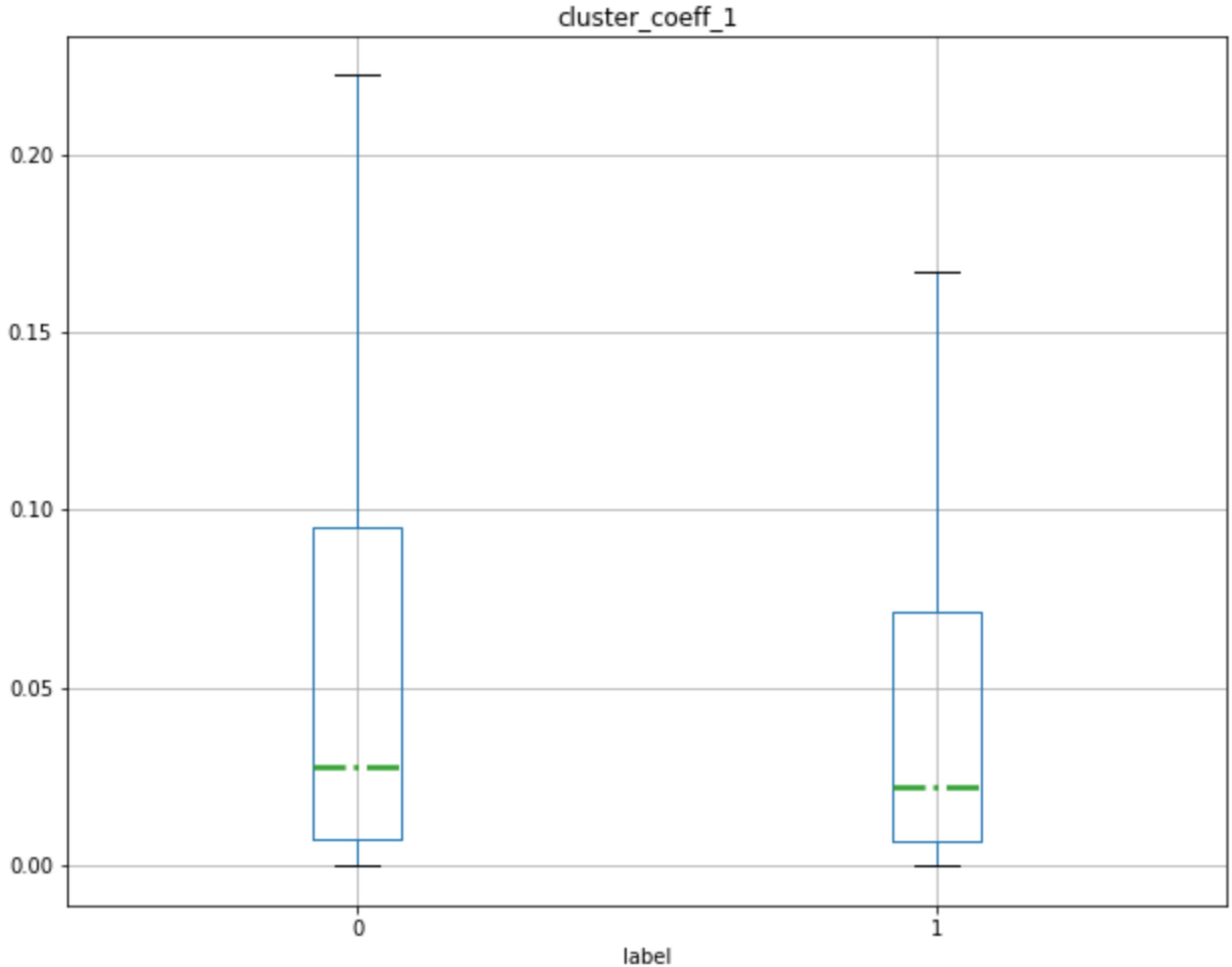
Non - Spammers $\neq$ *Spammers* ?

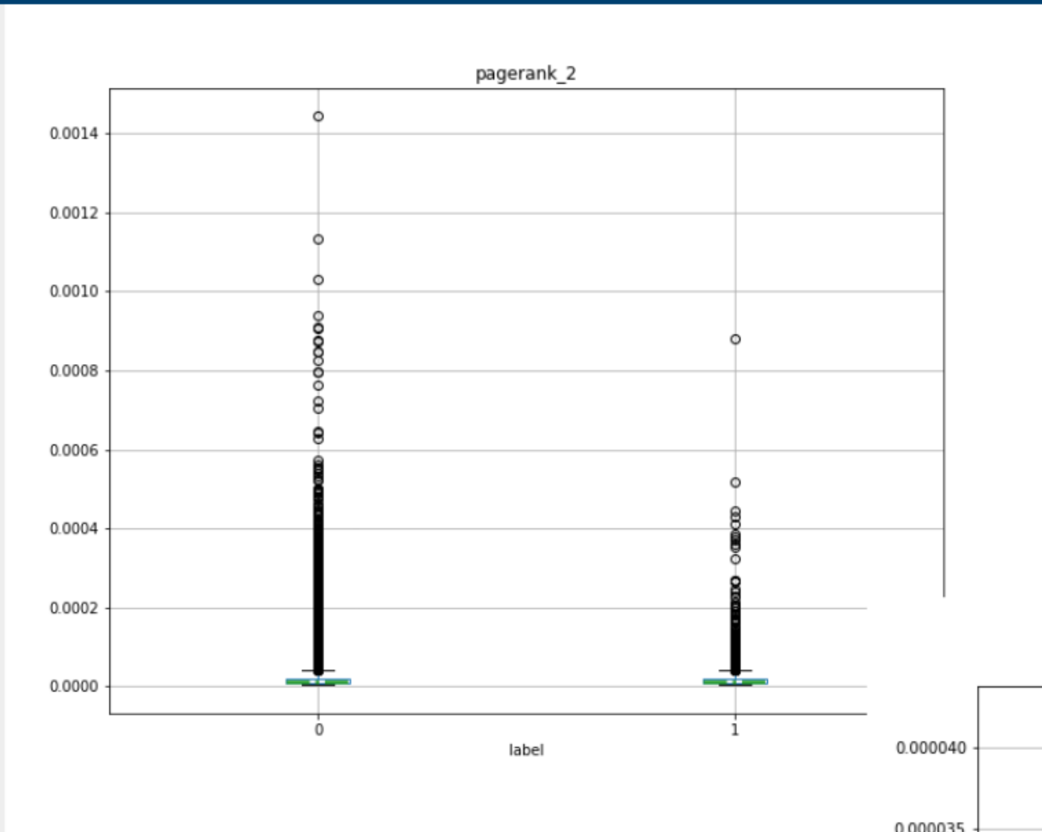


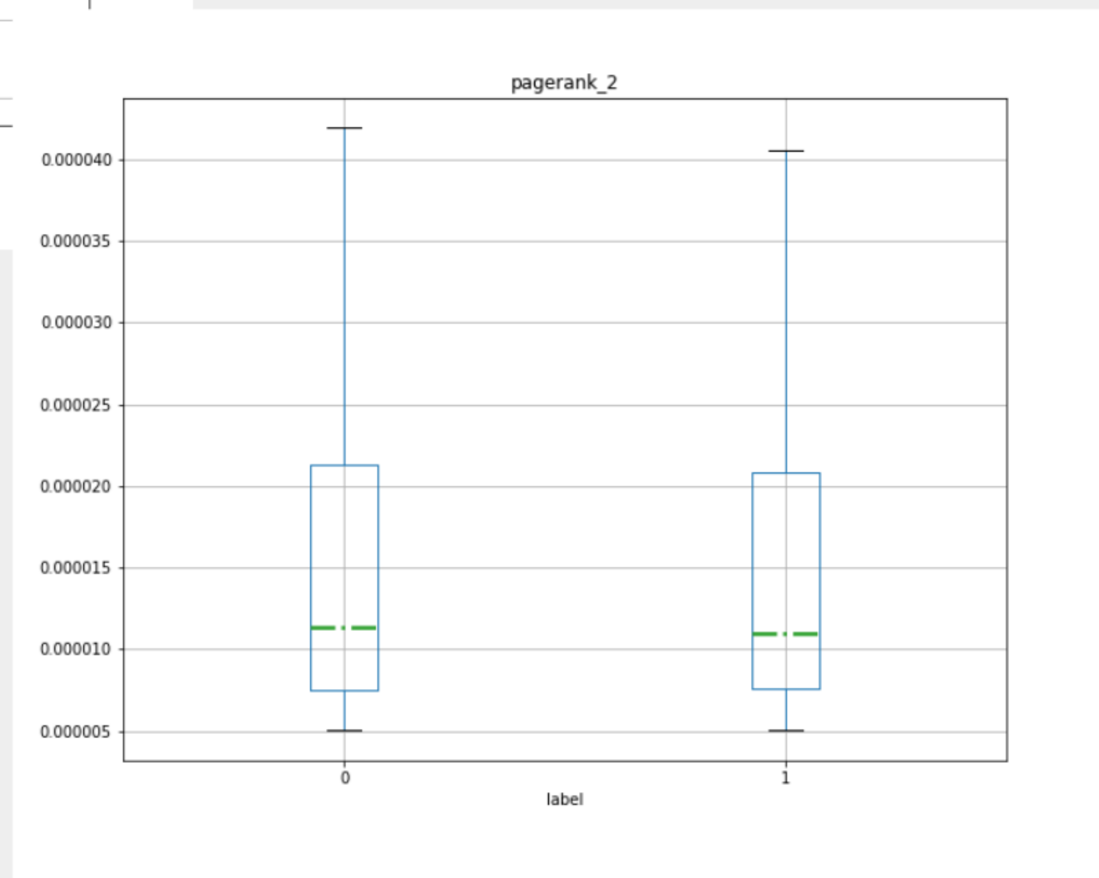
Not adequate difference!



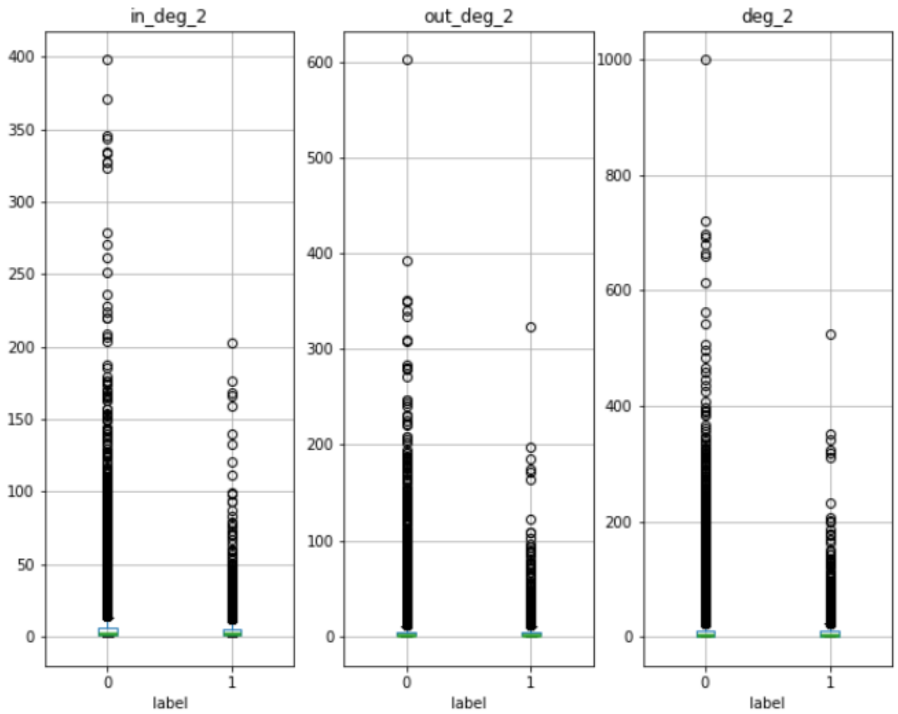

cluster_coeff_1
1.0
0.8
0.6
0.4
0.2
0.0
0
1
label


cluster_coeff_1
0.20
0.15
0.10
0.05
0.00
0
1
label

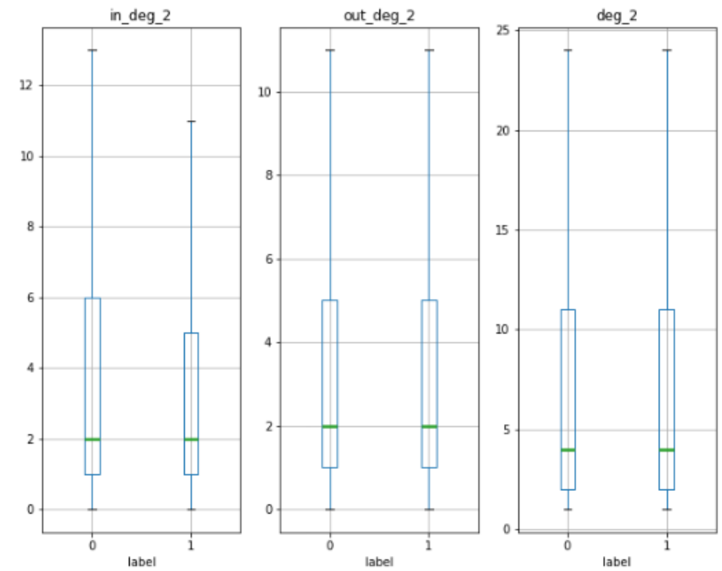

pagerank_2
0.0014
0.0012
0.0010
0.0008
0.0006
0.0004
0.0002
0.0000
0
1
label


pagerank_2
0.000040
0.000035
0.000030
0.000025
0.000020
0.000015
0.000010
0.000005
0
1
label

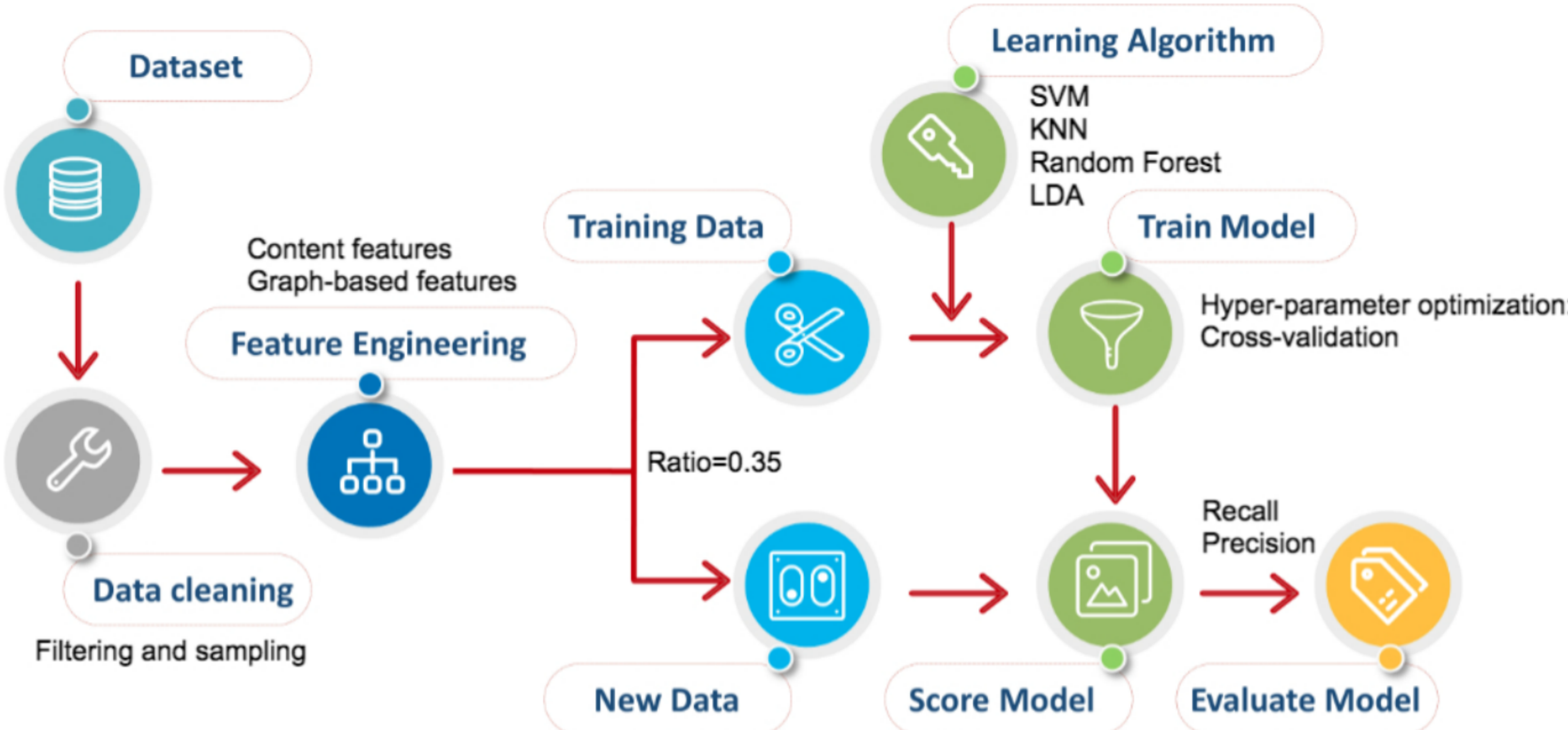
Degree distribution for relation 1



- Similar distribution
- Difference in outliers



Supervised Machine learning



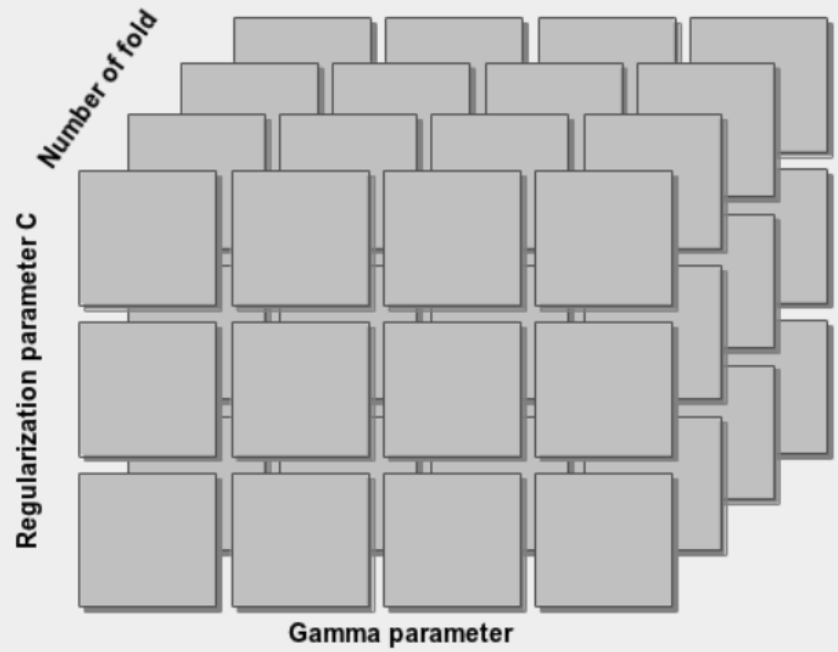
Classifications methods

- **SVM** : Classical classification method used in the literature for spammer detection. We chose RBF kernel, looking for non linear separation between the data.
- **k-NN:** Classification method that performs the best in low dimensions
- **Linear Discriminant Analysis:** try to find a linear separation between the data
- **Random forests**

Hyper-parameter optimization

- Using 5-fold cross validation
- A grid-search over the values of Hyper-parameters is performed for each classifier algorithm
- Optimal hyper-parameters are found by choosing those with the highest fold-averaged validation score

Gridsearch for SVM



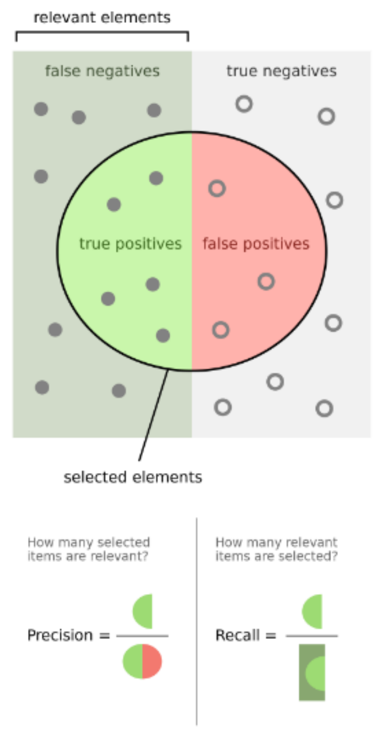
Metrics

- Dataset imbalanced: 90% non spammer

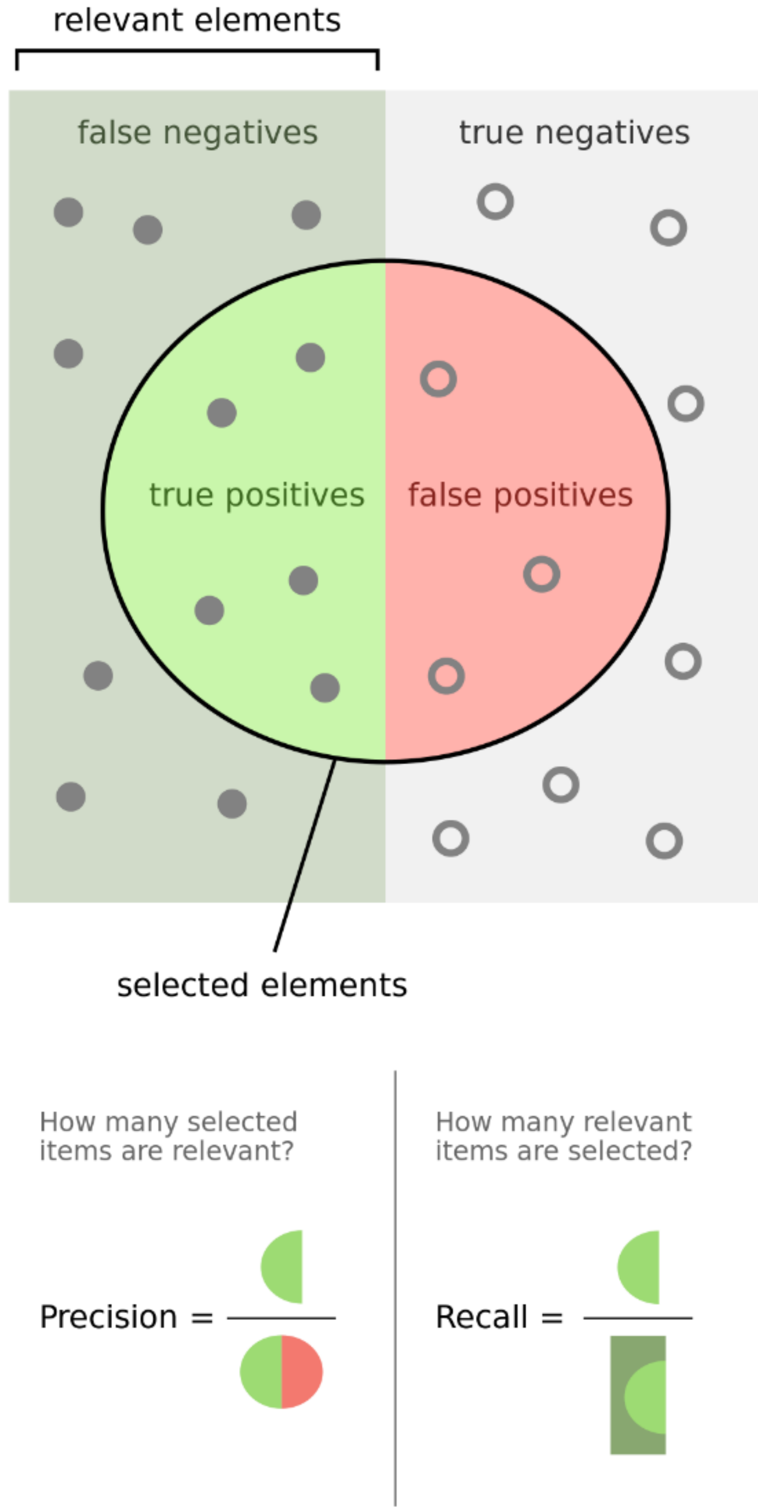
  Accuracy is not a good metric !

- Chosen metric : F-measure

$$\text{F-measure} = 2\frac{\text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}}$$



etric !


relevant elements
false negatives
true negatives
true positives
false positives
selected elements
How many selected items are relevant?
How many relevant items are selected?
Precision =
Recall =

Results

- Content features

classifier type	F-measure	Accuracy	AUROC
SVM age+gender	0.56	0.65	0.63
SVM age+gender+time	0.64	0.72	0.7
KNN age+gender	0.56	0.65	0.63
KNN age+gender+time	0.64	0.73	0.72
LDA age+gender	0.56	0.65	0.63
LDA age+gender+time	0.66	0.68	0.69
Random forest age+gender	0.56	0.66	0.64
Random forest age+gender+time	0.65	0.74	0.72

- Graph based features

classifier type	F-measure	Accuracy	AUROC
SVM	0.65	0.69	0.69
KNN	0.64	0.73	0.71
LDA	0.65	0.68	0.68
Random forest	0.66	0.74	0.71

Content features

classifier type	F-measure	Accuracy	AUROC
SVM age+gender	0.56	0.65	0.63
SVM age+gender+time	0.64	0.72	0.7
KNN age+gender	0.56	0.65	0.63
KNN age+gender+time	0.64	0.73	0.72
LDA age+gender	0.56	0.65	0.63
LDA age+gender+time	0.66	0.68	0.69
Random forest age+gender	0.56	0.66	0.64
Random forest age+gender+time	0.65	0.74	0.72

Graph based features

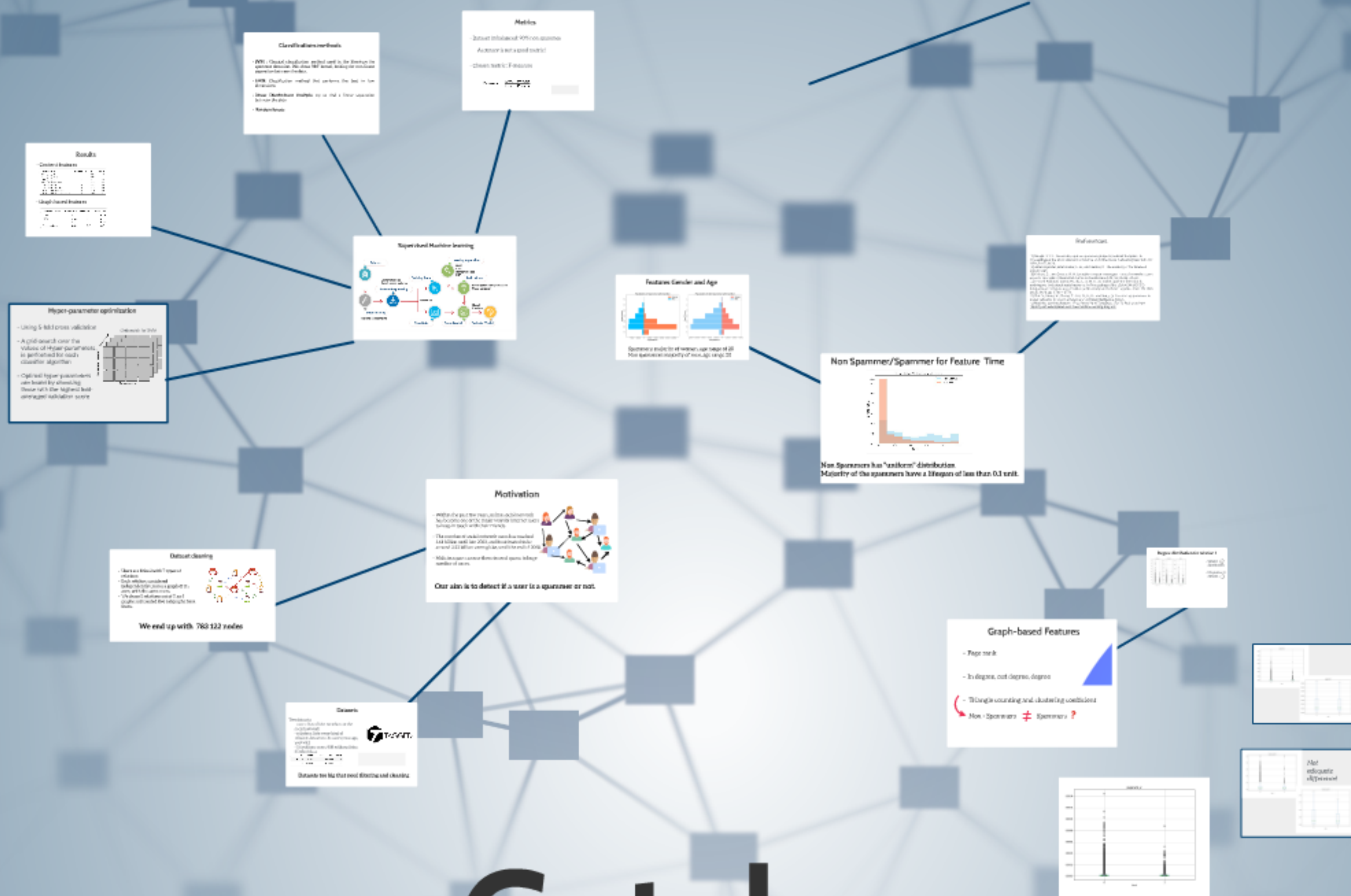
Random forest age + gender + time	0.69	0.74	0.72

Graph based features

classifier type	F-measure	Accuracy	AUROC
SVM	0.65	0.69	0.69
KNN	0.64	0.73	0.71
LDA	0.65	0.68	0.68
Random forest	0.66	0.74	0.71

References

[1] Akoglu, J. Y. L. Discovering opinion spammer groups by network footprints. In Proceedings of the 2015 ACM on Conference on Online Social Networks (New York, NY, USA, 2015), ACM.
[2] Johan Ugander, Brian Karrer, L. B., and Marlow, C. The anatomy of the facebook social graph.
[3] Pathan, S., and Goudar, R. H. Detection of spam messages in social networks based on svm. International Journal of Computer Applications 145, 10 (2016), 34–38.
[4] Shobeir Fakhraei, James Foulds, L. G. M. S. Collective spammer detection in evolving multi-relational social networks. In Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (New York, NY, USA, 2015), ACM, pp. 1769 – 1778.
[5] Zhu, Y., Wang, X., Zhong, E., Liu, N., Li, H., and Yang, Q. Discovering spammers in social networks. In AAAI Conference on Artificial Intelligence (2012).
[6] Machine learning diagram - Free PowerPoint Templates. (2017). Retrieved from https://yourfreetemplates.com/free-machine-learning-diagram/



Spammer... Catch me if you can

TEAM 20
Hedi Fendri - Paul Jeha -
Christina Mantonanaki - Nguyet Minh Nguyet