SEVASTOPOL
STATE UNIVERSITY

Analysis and preliminary classification of social network communities by the degree of attitude to the conflict

O V Chengar, V I Shevchenko, E N Maschenko,
D M Obolensky and A S Soina

This research is based on the researches, those aim is to solve an interdisciplinary fundamental scientific problem – the study of big data analysis technologies for developing scenarios of communication mechanisms formation for working with Internet communities in order to reintegrate a post-conflict society, for example, Ukraine. The strategy is based on the method of classification of VK user communities relying on keywords.

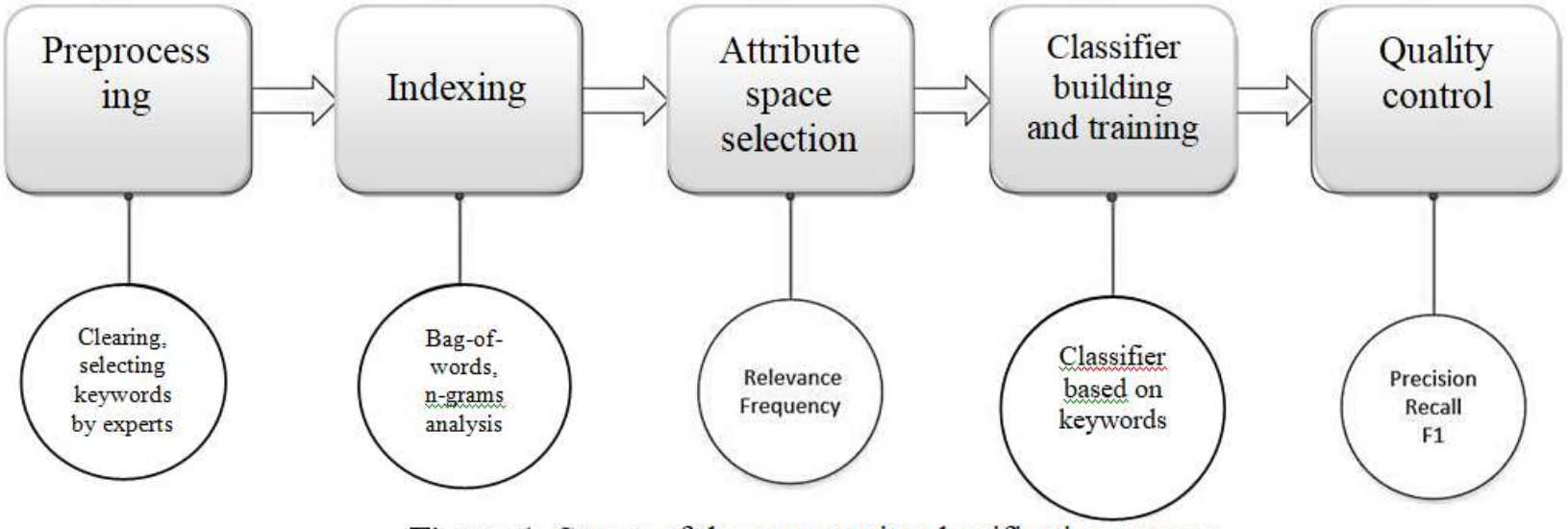

Figure 1. Stages of the community classification process

When analyzing the data, it was decided to exclude community profiles that have the following features:

- there is no description of the community (the reason – it is impossible to select keywords for creating an automatic classifier);

- paucity of a community (communities with fewer than 1000 members – such communities have little social influence);

- unacceptable content (content that violates the norms of morality – a list of keywords for identifying such communities is highlighted);

- the community type is “event” (not “group”): such communities have a limited time of existence.

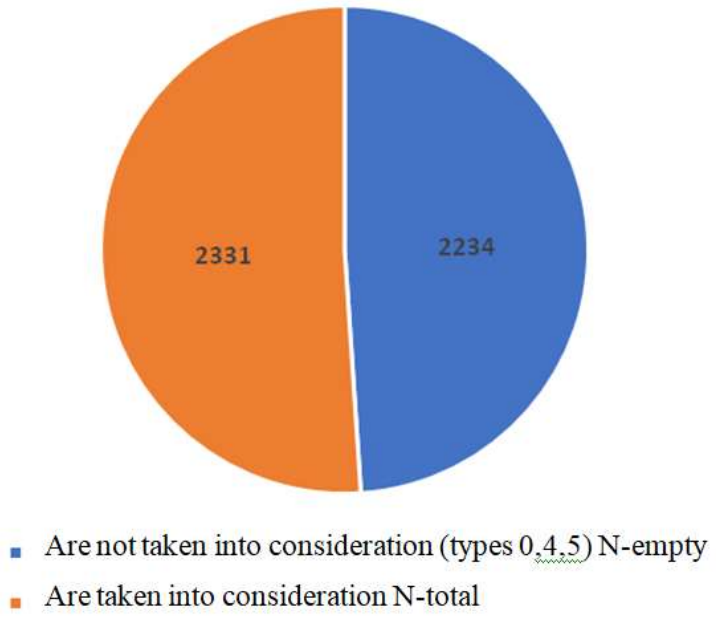

Figure 3. Data clearing results

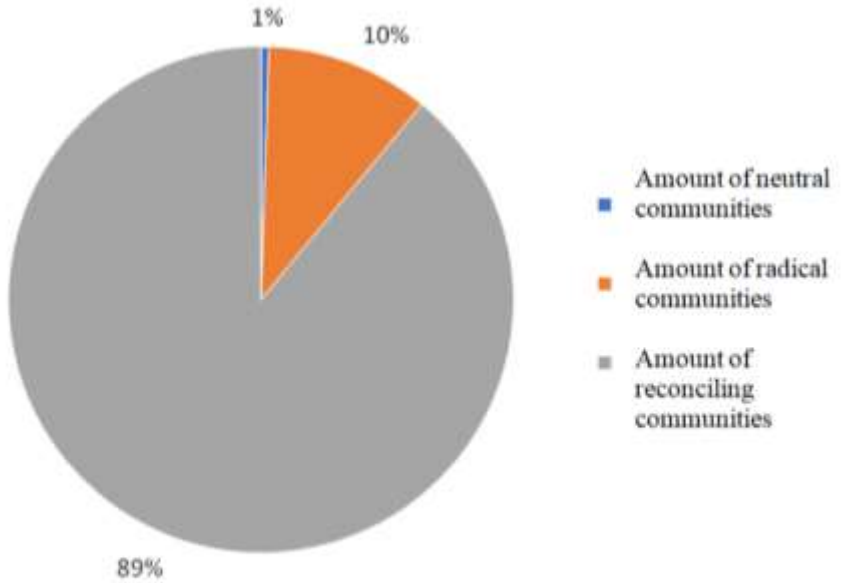

Figure 4. Distribution of users by community

Analysis and preliminary classification of social network communities by the degree of attitude to the conflict

SEVASTOPOL STATE UNIVERSITY

Table 5.Unigrams of keywords sets

Word	Weight of radical communities in the set of keywords	Word	Weight of reconciling communities in the set of keywords Word		Weight of neutral communities in the set of keywords
украін	5.644	групп	10.632	групп	11.967
українськ	5.426	наш	9.093	наш	11.605
наш	4.907	игр	8.647	работ	11.374
групп	4.700	год	8.340	заказ	11.232
груп	4.087	мир	8.224	товар	11.004
народ	3.807	сообществ	8.195	магазин	10.850
оруж	3.700	люд	8.022	доставк	10.690
люд	3.700	жизн	8.011	ваш	10.665
свобод	3.700	музык	7.889	цен	10.653
буд	3.585	правил	7.807	фотограф	10.586
росс	3.585	клуб	7.775	украин	10.338
файн	3.585	новост	7.721	одежд	10.223
українц	3.459	добр	7.665	люб	10.190
суспільств	3.459	которы	7.637	сайт	10.145
влад	3.459	друз	7.622	год	10.136
очищенн	3.459	вопрос	7.570	вопрос	10.120
комуністичн	3.459	всем	7.570	свадьб	10.063
вільн	3.459	фильм			
лиш	3.322	одн			
альбом	3.322	такж			
тільк	3.322	создан			
буд	3.322	виде			
сил	3.322	сериал			
організац	3.322	рок			

In this paper, data classification studies based on lexical and statistical analysis of keywords are carried out. Five experts analyzed training samples: class labels were placed and sets of keywords were generated for each class in order to automate the classification process. The experiments performed confirmed the representativeness of the obtained class labels.

The perspective of further research is the use of classification methods based on training with a teacher, which will improve the accuracy of classification.

Table 8. The results of classification by training set.

Community class	$Precision_{c_i}$	$Recall_{c_i}$	$F_{1\,c_i}$	Number of class labels
C_1 –Radical	0.6667	0.1250	0.2105	80
C_2 – Reconciling	0.4252	0.6176	0.5037	2210
C_3 – Neutral	0.9131	0.8379	0.8739	11350
Average metrics	0.6683	0.5268	0.5294	13640
Weighted avg	0.8326	0.7980	0.8100	13640

Thank you
for your attention