Andrzej Jarynowski, *Department of Sociology, Stockholm University*

How people discuss on Stockholm riots 2013

We investigate Forum posts (Poloniainfo.se) and Twitter during and after riots. We choose those datasets, because, they are freely availed and can be legally crawled from Internet and both can be treated as a big data. Scientific community has already experienced the power of social network media since riots in Tottenham (London) in August 2011 [1]. Since then Twitter and any other social media were carefully investigated for almost every social movement with computational tools constructed for this problem [2]. We try some simple tools of NLP (natural language processing) and text-mining [3] to obtain some kind of hierarchical or network structure of concepts mentioned by Internet-users. The initial disturbances in Husby, north Stockholm, were triggered by the police shooting of old Maghreb-origin man. Most of discussion took placed during riots since 15.05 (incident with police) via period of actual riots 20-25.05 and shortly after that. Both Forum and Twitter data were collected from 15.05 till middle of July (15.07) so both data series are exactly 2 months long.

1. Twitter

1.1. Data description and objectives

We analyze ~14k twits in different languages (mostly Swedish and English) tagged with hash Husby. That's implying international perspective of people, who express their thought via Twitter. Because of Multilanguage perspective of such twits, we decide to analyze not whole content of those tweets, but only co-occurrence with other hash tags. We do not differentatiate whow was twititng sinmple users, mainstream media, non-mainstream media, bloggers, activists, or even and the police. Only ~ 8k twits were taken in our analyze (only those with more than one hash tag). We choose 20 most frequent tags, there Svpol is the most frequent with Sthlmriots and Migpol far behind and rest [Tab 1.1]. We decided to analyze hash tags, just as they are, but there could be possibility to categorized some hash tags in just one category (e.g. by combining Sthlmriots with Stkhlmriot).

No	Hash tag	Counts
1	Svpol	3897
2	Sthlmriots	1319
3	Migpol	436
4	Sthlmriot	236
5	Stockholm	200
6	Aftonbladet	142
7	Nymo	124
8	Rinkeby	109
9	Polisen	108
10	Sweden	100
11	Upplopp	92
12	Kista	89
13	Svtdebatt	82
14	Vpol	80
15	Debatt	76

16	08pol	75
17	Expressentv	72
18	Megafonen	71
19	Kravaller	70
20	Tensta	69

Table1.1) Most frequent tags

1.2. Longitude analyze

We found, that for example Svpol [Fig.1.1] and Migpol [Fig.1.2] are tags, which were in constant use for whole period. Shape of curve is almost linear. That's mean; they occurrence probabilities are equal in time. Moreover, every second twit since beginning of riots till middle of July is associated with Svpol hash tag. The rest of hash tags died out after riots have finished.

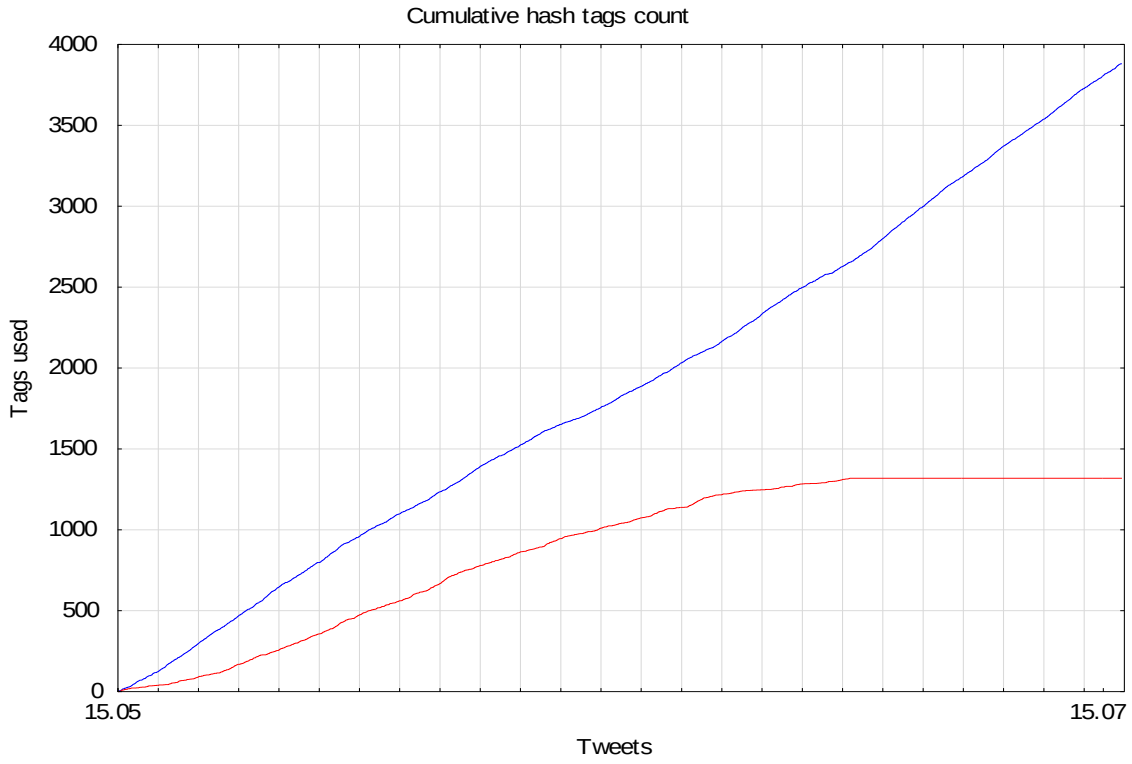

Figure 1.1) Longnitudal analyze of two most frequent tags usage. Blue- Svpol, red- Stkhlmriots

Moreover, for such as tags like Debatt or Svtdebatt, we observed, that people were use them only around the event [Fig. 1.2], which is very common phenomena in Twitter world. Some media names hash tags have stepwise shape like Nymo [Fig. 1.2], which also is characteristic for media. Those institutions provide some news, which are likely to be retwited. That explains big number of media hash tag use in short time surrounded by quite regions. The frequency of usage of given media hash tag could be also an indicator, how influential that media is.

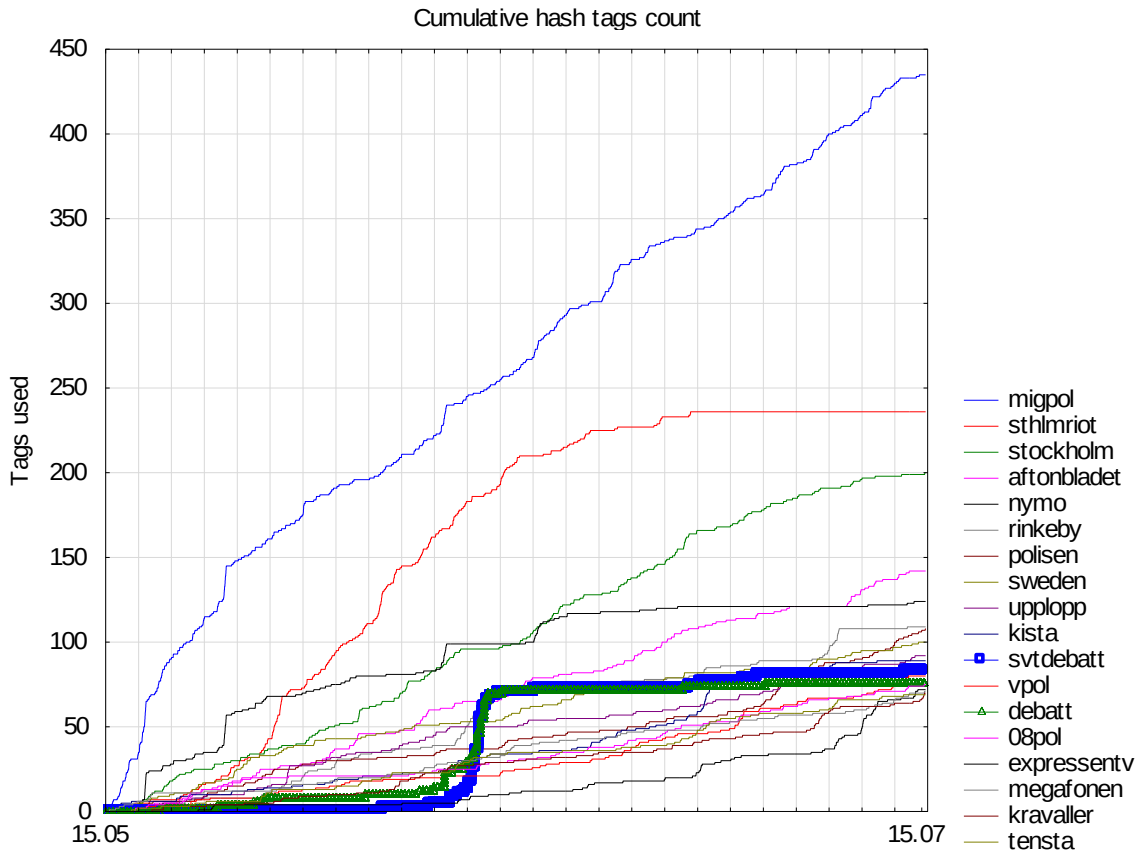

Figure 1.2) Longnitudal analyze of rest of most frequent tags.

1.3. Association analyze

We also try to find assotiation between tags. We define the link, when in the same tweet both tags coegsist. Hierarchical analyze shows leading role of dyad Svpol-Sthhlmriots [Fig.1.3, Tab 1.2]. To see hierarchical structure of rest of the system, we show once diagram with [Fig. 1.3] or without [Fig. 1.4] leading dyad. Morover, 2-gram elements (co-occurrence of 2 terms in one twit) of main dual dyad sthlmriots_svpol and svpol_migpol are few times more frequent than other elements [Tab. 1.2]. However, we cannot call triangle Migpol, Svpol, Sthlmriots as triad, because link between Migpol and Sthlmriots was observed only 37 times so it is on order of magnitude weaker than main double dyad (we do not see Migpol, Svpol, Sthlmriots cluster [Fig. 1.3]).

No	Hash tag	Counts
1	sthlmriots_svpol	533
2	svpol_migpol	353
3	svpol_sthlmriot	107
4	migpol_nymo	50
5	vpol_svpol	47

Table 1.2) The most frequent 2-grams. Evidence of importnace of dual dyad sthlmriots_svpol and svpol_migpol.

Let’s define construction of hierarchical diagrams [Fig. 1.3, 4]. We analyze synchronous time evolution of a pair of assets [4], where every series element is vector of occurrence of given hash tag. The distance is a Euclidean distance between every pair of hash tags.

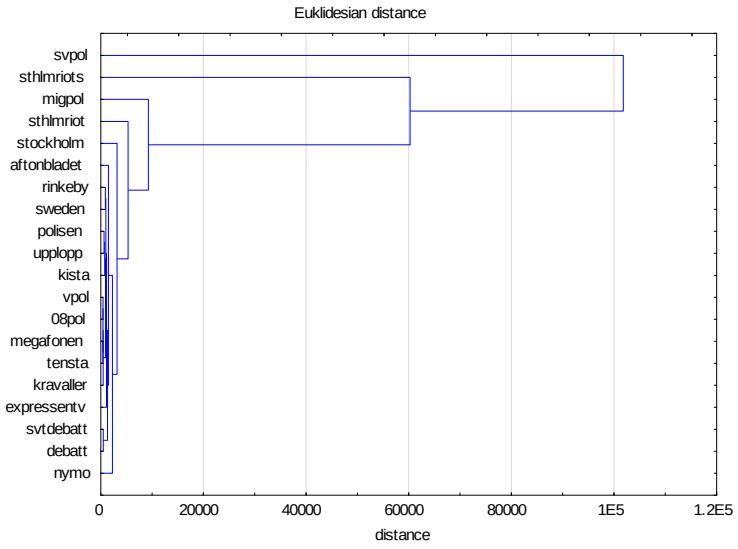

Figure 1.3) Hierarchical diagram for all tags

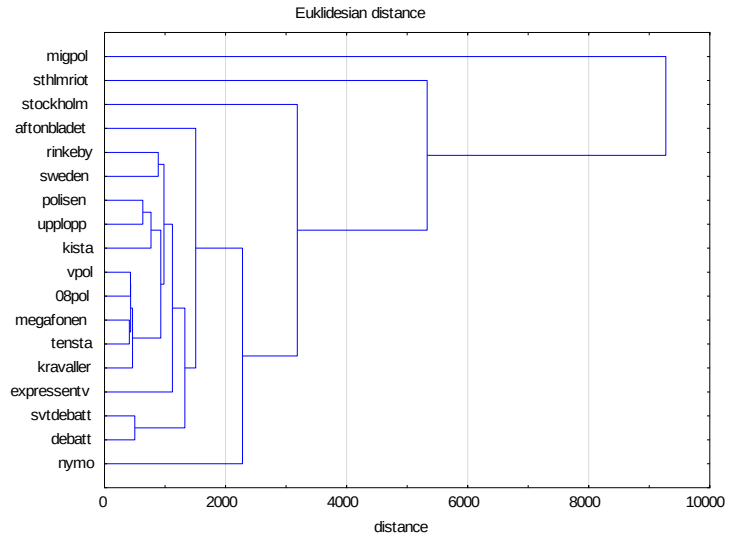

Figure 1.4) Diagram without two most frequent terms

Let’s define construction of the network [Fig. 1.5]. We decided to establish lower threshold on the level of 2 tweets, needed to create a link (everything below seems to be a noise, because link- association should be repeated at least once to avoid random effects). Thickness of the link corresponds to its weight (count of given 2-gram). On the other hand, dual dyad sthlmriots_svpol and svpol_migpol thicknesses were reduce not to cover whole figure.

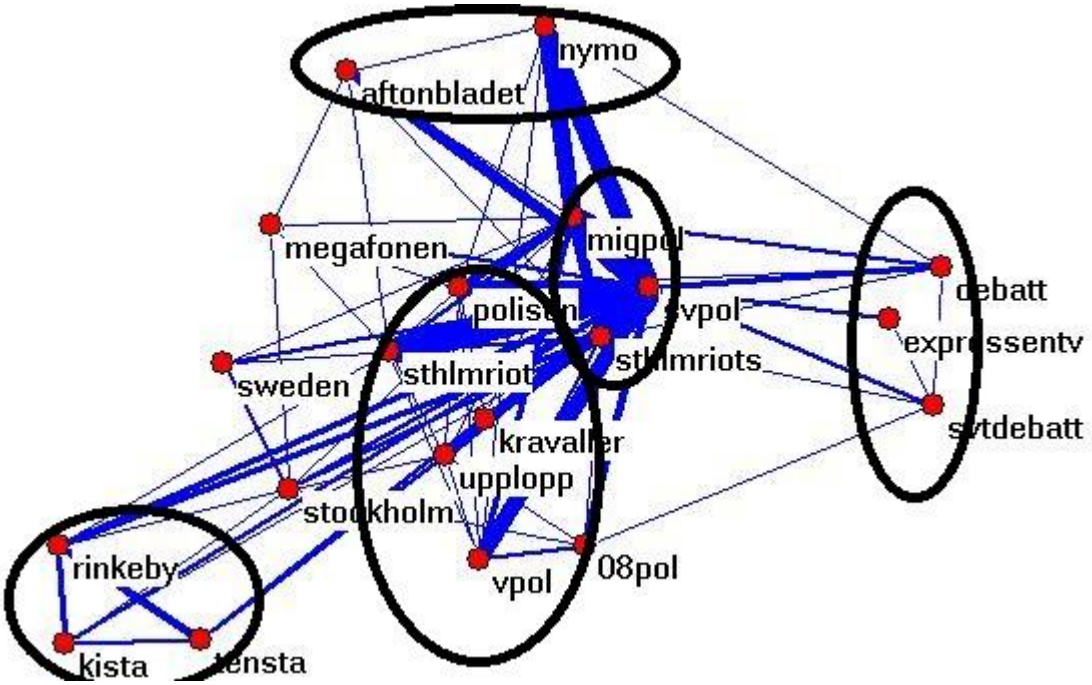

Figure 1.5) Network of connections (links thinness in dual dyad sthlmriots_svpol and svpol_migpol were decreased to see other links also)

Hierarchical and network analyze [5] show leading role of dyad Svpol-Sthhlmriots and quasi-triangle Migpol, Svpol, Sthlmriots. On the other hand we can find clusters of geographical districts, tags related to debate, Swedish words describing riots, and media institution.

1.4. Conclusion and future work

Provided analyze shows many features known already from other studies and general observation, but here they are presented in more systematized way. The most important topic is politics. Svpol as other hashes with the same meaning, are definitively the most frequent hash tags. Moreover only Svpol and Migpol seem to be used after riots with the same frequency as before. It would be interesting to see how hashes about politics co-occurrence with others change with time. Another question could be asked with sensitive analyze: how emotionally oriented are hashes about police.

2. Poloniainfo.se

2.1. Data description and objectives

Internet Forums like poloniainfo.pl are not to broadcasting media as Twitter, but relations between users are usually stronger and more personal. Quantitative researches could be deeper due to complex relations between users [6], but on the other hand amount of data is not as impressive like in Twitter. We look at frequencies of world used by Forum user in Topic about Riots in 525 Posts. We choose only Polish words in this analyze. Firstly we found extremely huge amount of personal and possessive pronouns of third person in plural form [Tab. 2.1]. Everything seems to be describing about "Them" more often called "others" in ethnological literature. "They" are native Swedes represented mostly by government and police and another "They" riots participants that indicates obrservative way of looking on the riots. Polish community did not take part of riots, and on the other hand have no influence on politics of Sweden. That makes this medium neutral, while both pro and against riots view were presented there. However, Polish people identify themself culturally with Swedish establishment and describe problems of Husby citizens non like their own perspective.

interesting pronoun	**Polish**	**frequency**	compared pronoun	Polish	frequency
them	**im**	**79**	us	nam	4
the	**tym**	**102**	us	nam	-
these	**tych**	**72**	ours	nasze/nasz/nasza	2
they	**oni**	**53**	we	my	1
these	**ci**	**48**	we	my	-
them	**nich**	**38**	us	nas	13

Table 2.1) Orientation of conversation on „them"

2.2. Methodology and data mining

We tried to categories all words used in discussion in few categories. To do so, we choose only those worlds, which have only one clear meaning somehow related to the topic. We found 386 different words which appear at least once in our sample and seem to have some important meaning. From them around 300 were attached to different categories 1-10 with subcategories described by some keywords [Tab. 2.2]. Every category allocate sum of number of unique words which belong to family of given keywords related to given category of subcategories.

1.1 Employment	6.0 Politics-general	7.1 They	8.2 Police-law
work/employees	government	immigrants	law
hardworking	debate	Arabian	cutthroat
rich	party	other nations	
money	politicians	origin	9.0 Riots-general
taxes	democracy		throw
		7.2 Swedes	riots
1.2 Unemployment	6.1 Politics-multikulti	Swedes	night
unemployment	invite	Swedish	street

social help	acclimatization	Sweden	violence
poor	multikulti	Europe	stones
	hope	nobility	car
2 Family	tolerance		
family	asylum	8.0 Police-general	9.1 Riots-pro
	get	police	rebellion
3 Religion	arrives	military	youth
Islam			protest
religion	6.2 Politics-segregation	8.1 Police-induce	vulnerable
	racist names	killed	
4 Education	eugenics	wounds	9.2 Vandalism
education	racism	induce	fires
schools	segregation	bullets	vandalism
learn	deportation	weapon	aggression
language	hate	shoot	
		Police in slang	10 External field
5 Living	7.0 Identity-general	knife	other riots
apartments/residents	nation	disarm	problems
district	Stockholm		wars
Husby	society		media

Table 2.2) Categories, subcategories and keywords describing them

2.3. Limitation of coding

Meaning of words used by people is very difficult to uniqueness classification. In our task we propose 10 main categories with 14 subcategories and attract presented words into given keywords related to descriptive category or subcategory. Classification based on ours subjective feeling. We tried to avoid words with many meanings. We had problems with words:

- Stockholm/Sweden (does not only relation with that city/country, but also geographical location);
- Swedish/Swedes (does not only describes citizenship of Sweden, but also the background of the riots);
- all words classified to categories pro riots or pro police (words connoted with law or rebellion have mostly positive meaning, but not always);
- all words classified to category external forces (e. g. none of media institution like radio, TV, press, or Internet companies names were include in investigation)

Also all of categories have very wide range of potential connotation and some of them could overlap (e. g. where should be border between employment and unemployment), but we tried to help ourselves with keyword list [Tab. 2.2].

2.4. Results of categorization

From results [Fig.1, 2] we can conclude, that main conflict is going on around identity, police operation and riots itself, work and living issues. Politics, education and family related issues plays secondary role, but still such topics were discussed by Forum users. The main subject of discussion seems to focus around identity (the biggest count of related words) which was

already observed from intensively of pronounce use [Tab. 1.1]. Moreover, motor of conflict could be define as Swedes-They. With Swedes outside of literary category 7.2 associate the second frequent category: Employment-which probably is the biggest single issue related to Stockholm's riots. The smallest subcategory was that one containing words related to indicated positive side of police operation. One order of magnitude often Police was described by negative or neutral connotation.

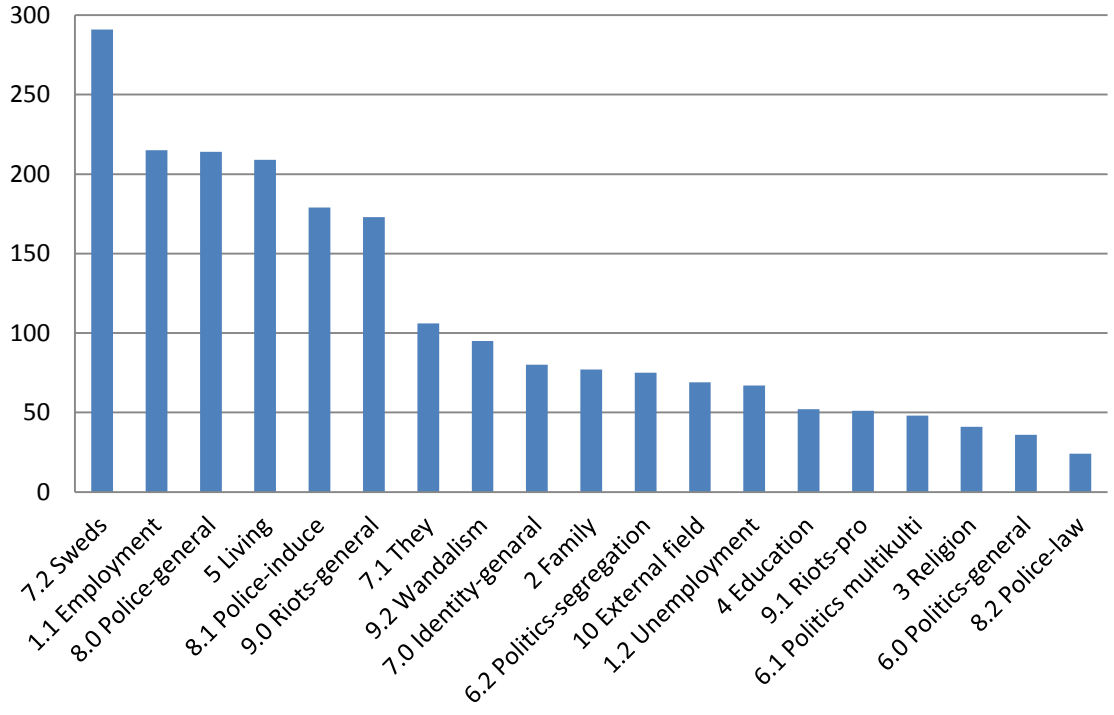

Figure 2.1) Categories and Subcategories from most to least popular

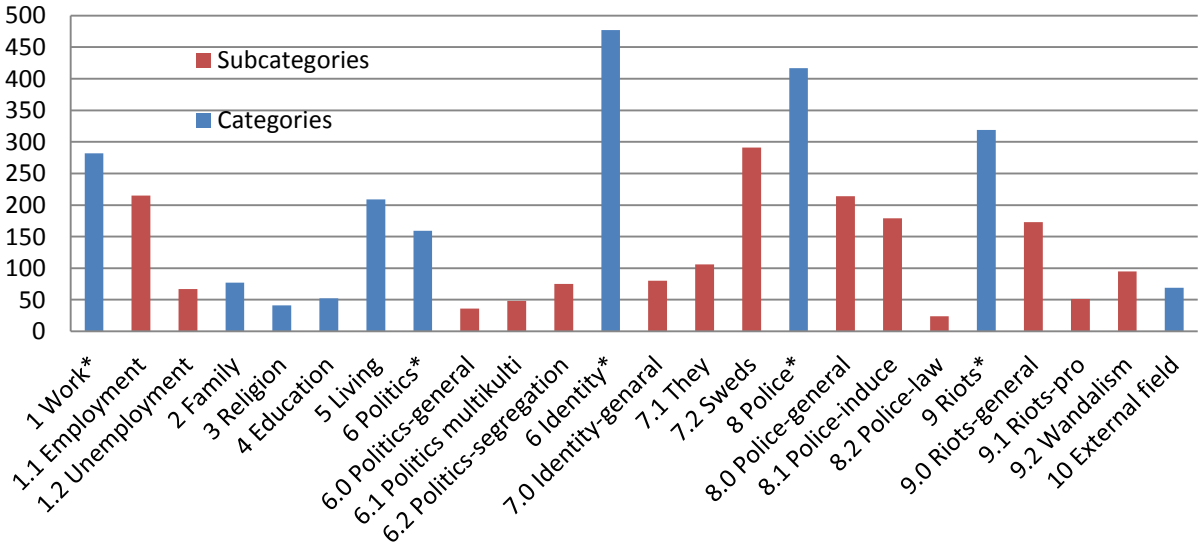

Figure 2.2) Categories and categories containing subcategories. * means sum of all counts of subcategories for given category

2.5. Sensitive analyze of 2-grams with Police

We extract all 2-grams where word police or police officer appears. We run sensitive analyze on each 2-gram. The sentence sensitive strength[1] could vary from -4 (very negative) to +4 (very positive). Most of them have been neutral and sensitive strength is 0. We additionally show all those 2-grams, which were found more than once and we show only emotionally oriented 2-grams below that threshold. Moreover Polish stop words (without meaning) were excluded also. To get effective power, we multiply frequency by sensitive strength. The overall score of power for all 2-grams containing word Police is slightly negative (-5).

Polish 2-gram	English 2 -gram	counts	sensitive strength	power
szwedzka policja	Swedish police	10	0	0
policja uzywa	police uses	3	0	0
granatniki policjanci	police launchers	2	-1	-2
jaka policja	what the police	2	0	0
mogla policja	police could	2	0	0
mordowac policja	murder, police	2	-3	-6
nie policja	not the police	2	0	0
policja bawiła	police amused	2	1	2
policja bedzie	police will	2	0	0
policja francji	French police	2	0	0
policja powinna	the police should	2	0	0
policja przyjezdza	police arrives	2	0	0
policja strzela	police shoot	2	1	2
policja szczula	police bully	2	0	0
policja szwedzka	Swedish police	2	0	0
policja zacznie	police will	2	0	0
policja zadecydowala	police decided	2	0	0
policja zrobila	the police had done	2	0	0
policjanci niech	let the police	2	0	0
policjanci pracujacy	police officers working	2	0	0
policjant brzuchem	policeman belly	2	0	0
policjanta poczuje	police feel	2	0	0
praworządność policjantów	the rule of law officers	2	0	0
reagowala policja	police respond	2	0	0
solna policjant	Solna policeman	2	0	0
bezpiecznie policja	Police safely	1	1	1
policja broni	Police defends	1	1	1
policja pogarsza	Police worsens	1	-1	-1
policja zabila	Police killed	1	-3	-3
policjantem kolega	cop friend	1	1	1
samotny policjant	lone policeman	1	-1	-1
policjant ratowal	policeman rescued	1	1	1
			Summary Power	**-5**

Table 2.3) 2-grams with Police, Policeman with their sensitive score and frequency.

[1] We use Sesnistrenth [2] for English translation of Polish sentences

2.6. Conclusions and future works

This work has only exploratory function, but even those preliminary results; we can propose few hypotheses, which should be checked by deeper investigation.

First one is related to Police operation. Please note, that we do not want evaluate professionalism of Police, but only public opinion feeling about their operation. Data-mining analyze proposed by us has many weak sides and it is usually very difficult to make clear conclusions from it, but with police case order of magnitude difference between positive and both negative and neutral cases seems to be something more than methodological bias or artifact. Why Forum users even, they identify themself with Swedish establishment did not say almost anything positive about Police? Did PR department of Police work properly? One of user wrote, that Swedish Police as a best paid organization in EU is one of the less effective the same time. We propose to do a survey about public evaluation about Police operation. Moreover sentiment and association analyze should also give some more insight, while preliminary result show little negative bias [Tab. 2.3]. This sample is unfortunately too small for any conclusions and bigger datasets should be used to estimate actual sensitive power.

Employment also seems to be relatively important topic [Fig. 2.1]. Work issue, with connotation with taxes and salaries should be more carefully investigate if that is really a leading factor of discussions about riots. It beats some aspects of identity (and even identity also if other coders would not include geographical word – e. g. Stockholm into identity category) and religion, living condition or education form frequentional analyze. It could come from bias, that Poles describing riots are mostly gastarbeiters and work, as a single theme, is the most popular within Polish community in Sweden.

3. General findings and limitation of both studies

Role of Police and politics should be investigated, because in Twitter study politics and in Forum Police is the main motor of opinion spread between people [Fig. 1.1, 2.2]. Both studies are only preliminary and just explore field to set up questions to be ask by pragmatic research. Both datasets are not representative for whole society and opinion shared in both mediums are very special to people who use them. Moreover in each study different methodology was used due to difference of data structure itself, and even in content. Some aspects revealed in one study were omitted in second one. For example media institution, which play important role in hash tag study, were not categorized anywhere in Forum case. However, similar issues come out from both mediums.

Bibliography

[1] theguardian&LSE, ”Reaading the Riots,” London, 2011.

[2] M. Thelwall, ”Sentiment strength detection for the social Web,” *Journal of the American Society for Information Science and Technology,* vol. 63, nr 1, pp. 163-173, 2012.

[3] B. Yuskiv, Content analysis. History development and world practice, Rowne, 2006.

[4] A. Buda och A. Jarynowski, Lifetime of Correlation and its applications, Wroclaw: Wydawnictwo Niezaleze, 2011.

[5] A. Buda och A. Jarynowski, ”Network Structure of Phonographic Market with Characteristic Similarities between Artists,” *Acta Physica Polonica A* , vol. 123, nr 3, 2013.

[6] A. Zbieg, B. Żak, J. Jankowski, R. Michalski och S. Ciuberek, ” Studying Diffusion of Viral Content at Dyadic Level,” i *ASONAM 2012*, Istambul, 2012.