The Belarus protests (2020) – a Twitter data analysis

This is a description of the technical steps taken for analyzing Twitter data related to the crisis of 2020 in Belarus, as part of the MOBILISE research project.

Data collection

From the day of the election (July 9, 2020), we set up a weekly collection of all tweets and retweets containing hashtags related to the events, with a particular focus on discontent and protester's slogans. This list of hashtags was curated by several researchers based on their observation of online and offline media. Data was collected through the legacy Twitter API (V1) with the help of open-source library Rtweet. Mining those hashtags led to approx. 2M tweets from August 9 to December 31, 2020.

In March 2022 we performed an additional collection, taking advantage of the Archive Search from Twitter API V2 and using the open-source library Twarc. First, we extended the search going as far back as June 1st, 2020, covering missing periods in the initial data collection. We then did a simple, manual keyword expansion, adding to our query list (1) the top 70 hashtags based on frequency in the original dataset, that we weren't already tracking, excluding those that were irrelevant to the election and protests ; (2) the top 70 hashtags in cyrillic alphabet only[1] ; and (3), keywords which were « trending in Belarus » pre-election and post-election according to two websites that archive Twitter trends. This last batch included non-hashtag keywords such as "psiphon" (name of a popular VPN used during Internet service interruptions), names of cities ("Жодино", "Брест" - Zhodino, Brest) and popular figures of the opposition ("Латушко", Latushko).

Data that we mined during this second stage was then merged with the original dataset. Figure below illustrates the number of tweets yielded at each step. Even for the same time period and hashtags, data collected using Rtweet and Archive Search do not entirely overlap : our legacy collection retained tweets and users that were since deleted, while the second batch includes retweets that appeared later, or were missed due to changes in Twitter indexing or API errors. When possible, the original data was retained, in order to match as closely as possible the Twitter activity at the time of the events (user description, number of retweets).

Dataset cleaning and subset selection

Removal of tweets from bots

The first step of data cleaning aimed to remove bot activity, which is reported to have increased after the election (Rice et al., 2021). We experimented with several measures to remove tweets and retweets from bots from the dataset, namely training a classifier on three of the features reported as most significant in the literature on automated bot detection : retweet to tweet ratio, friend to follower ratio, and mean frequency of tweet (Derhab et al. 2021). We also used a deep classifier (Yang, Ferrara, et Menczer 2022) and an ensemble classifier (Sayyadiharikandeh et al. 2020), both accessible through an open-source API. However, these methods seemed to have a high false positive rate. We thus simply

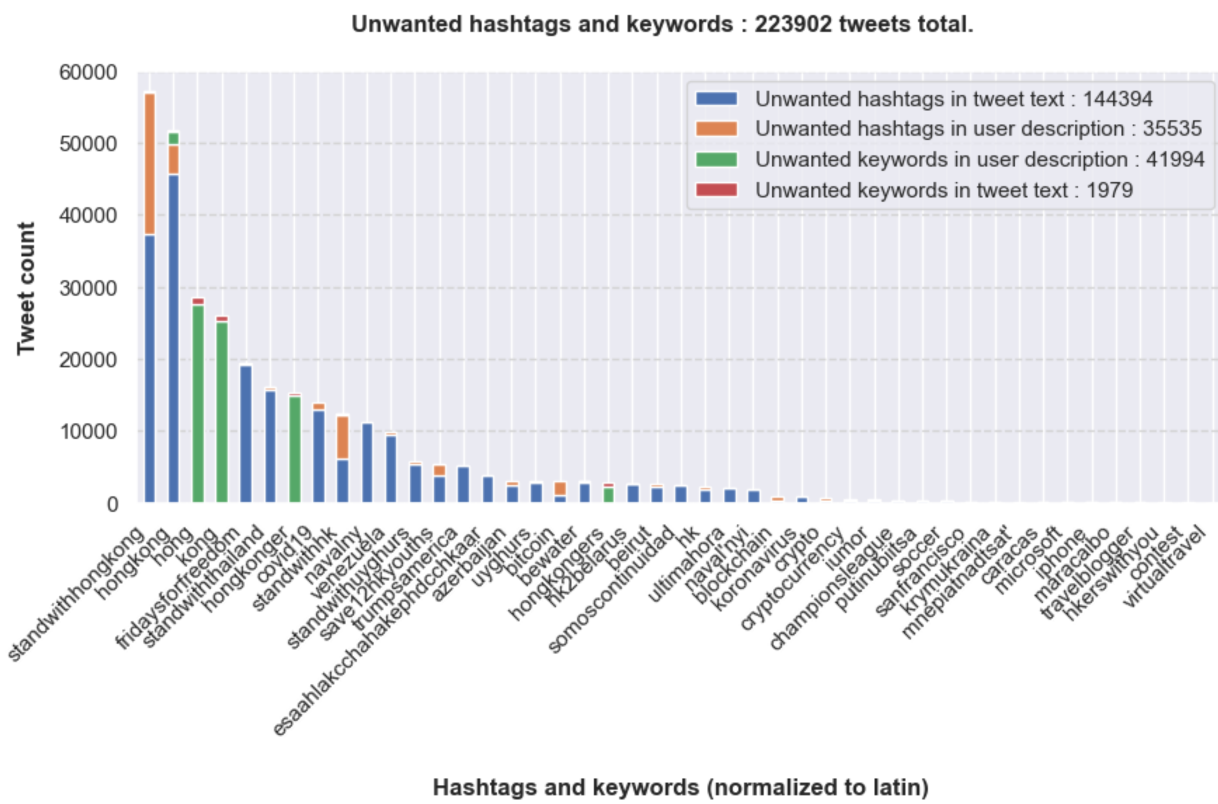
[1] This was done as a way to help balance the dataset and increase representation of Russian, Belarusian and other Slavic languages.

removed all tweets (approx.. 25 K) from users whose name contain various forms of the word « Bot » (Latin and Cyrillic).

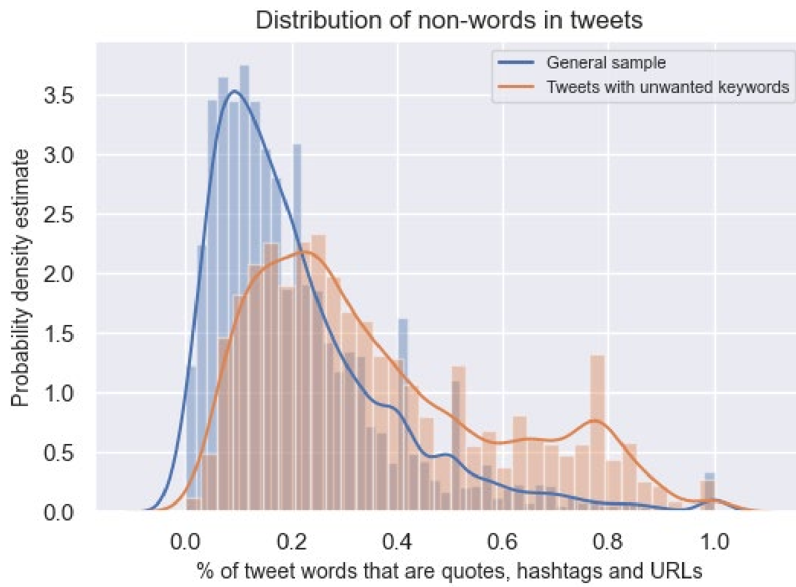
Removal of tweets based on irrelevant keywords

We then removed approx. 250K tweets based on certain unwanted hashtags and keywords, not only in the tweet text, but also in the user's descriptions; some were linked to completely irrelevant topics trending in the general Twitter sample (sports, travel, COVID-19), mainly by-products of collecting generic hashtags such as #Belarus. Other keywords were being used in conjunction to the Belarus protests keywords in order to raise awareness for other protests, or in some cases, marketing purposes or other forms of « newsjacking ». Notably, we chose to target keywords related to the Hong-Kong and Thai protests, whose promoters ("Milk Tea Alliance") had engaged in a very active support campaign for Belarus especially during the first days of the protests, creating a significant amount of noise for the purpose of textual analysis.

Figure below illustrates the keywords and data removed at this stage.



These tweets fit a pattern of retweeting, quoting, replying, or otherwise mimicking the content of tweets, without adding meaningful commentary. For instance, mean retweet count among tweets we deleted was around twice what was observed in the general sample (~600 vs ~1200, excluding outliers above 10k retweets). Moreover, when comparing original tweets, the tweets we deleted were far more likely to contain more "non-words" (that is, hashtags, URLS or quotes) than in the general sample (see fig. below).


Distribution of non-words in tweets
General sample
Tweets with unwanted keywords
Probability density estimate
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
0.0
0.2
0.4
0.6
0.8
1.0
% of tweet words that are quotes, hashtags and URLs


小龍貓
@edwardso924
#truth #justice #HumanRights #democracy
#FightForFreedom #StandWithHongKong
#MilkTeaAlliance #Belarus #Uyghurs #Mongolia #Tibet
5:26 PM · Sep 21, 2020

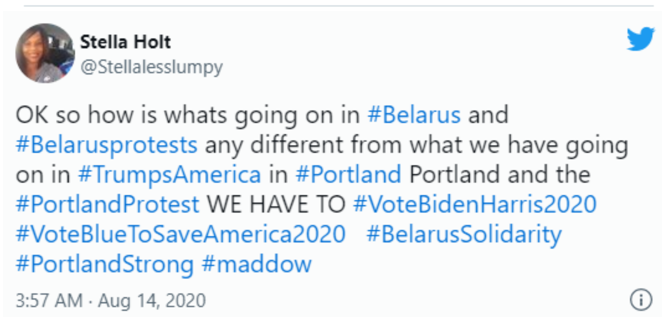
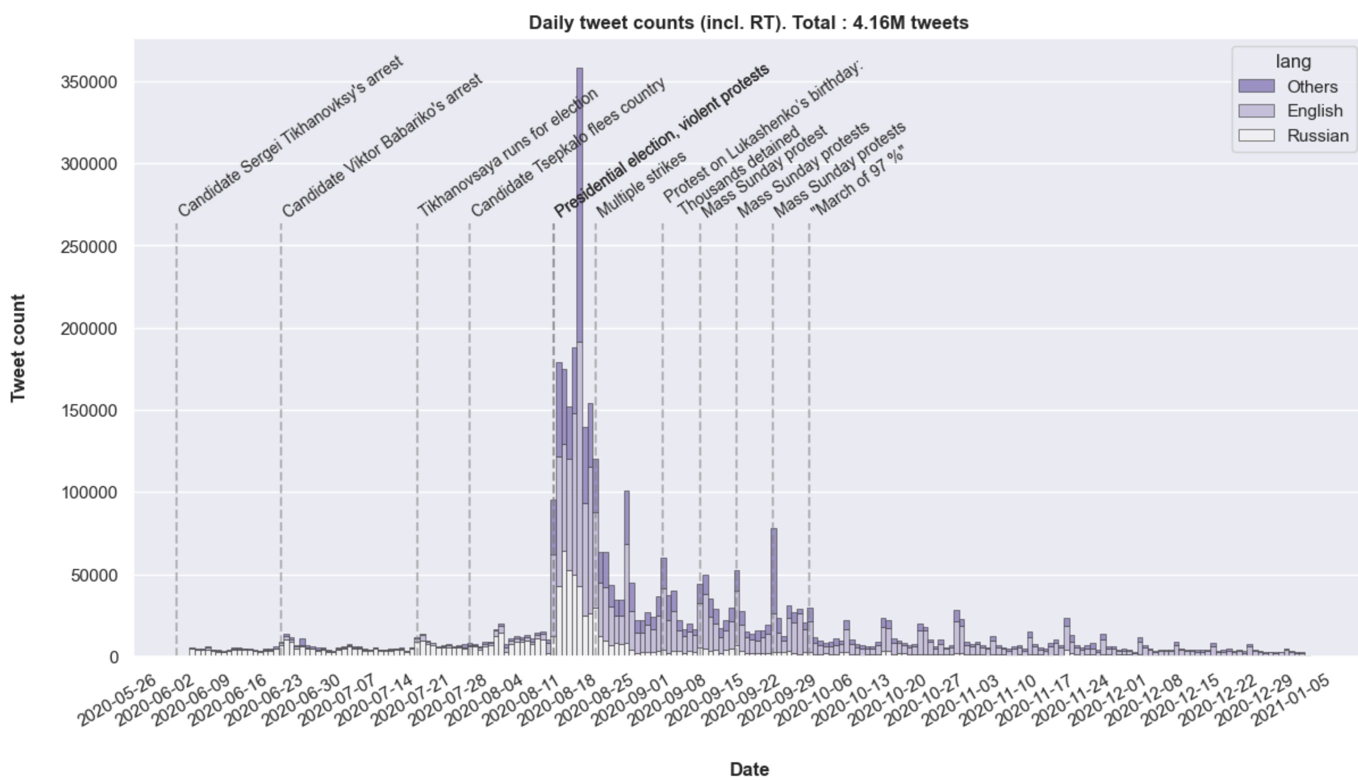

Stella Holt
@Stellalesslumpy
OK so how is whats going on in #Belarus and
#Belarusprotests any different from what we have going
on in #TrumpsAmerica in #Portland Portland and the
#PortlandProtest WE HAVE TO #VoteBidenHarris2020
#VoteBlueToSaveAmerica2020 #BelarusSolidarity
#PortlandStrong #maddow
3:57 AM · Aug 14, 2020

Figure 1- example of deleted tweets

Dataset description

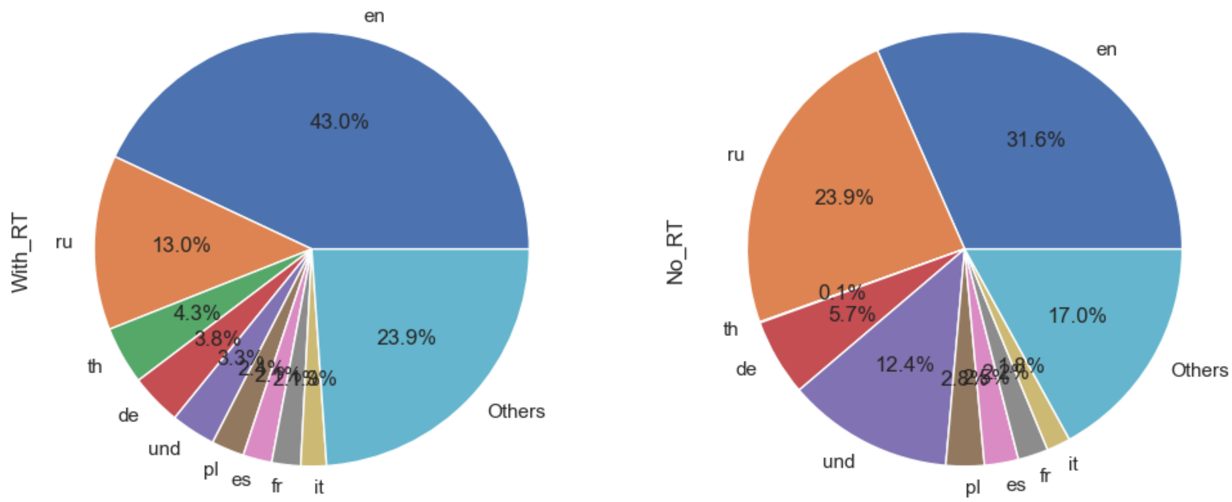
The dataset comprises 847 705 unique users, 4 160 528 tweets (including retweets) and 867 324 original tweets.

Figure below shows tweet and retweet counts during the major events.



Subset selection based on language

Tweets from our dataset are labeled with the language detected by Twitter. Our data comprises 58 languages and 152 k tweets where the language remained undetected ('und'). Note that Twitter was unable to detect Belarusian language (code : be), which was likely detected as Russian or undetected.



For the purpose of our analysis, we retain only tweets detected as English and Russian. However, in the case of tweets where language was undetected, when their author has tweeted in one single other language with sufficient volume, we assume their 'undetected' tweets are also in their main language. More precisely, we make this assumption when there is at least a 3:1 ratio between that main lang and their 'undetermined' tweets, *and* a positive z-score indicating this user's number of tweets in ``lang`` is superior to the mean tweet count in ``lang``. This rather conservative method allows us to infer language for 8.5K tweets in Russian and English.

NLP analysis

Our analysis relies on topic modelling, an unsupervised natural language processing (NLP) technique aiming to infer meaningful collections of words (i.e topics) based on their distribution across a collection of documents. Topic models have been widely used in humanities research to elicit what events or concepts are being discussed within corpora (Vayansky et Kumar 2020). One seminal method, Latent Dirichlet Allocation (LDA, Blei 2003), aims to do so by hypothesizing documents are generated through sampling from a mixture of topics, where the mixture proportions are distributed as a random Dirichlet variable ; then sampling words from that mixture. Through variational Bayesian optimization, LDA learns the topic-term distribution that would best recreate the corpus. This generative model, and other "bag-of-words" models such as Non-negative Matrix Factorization (NMF), rely purely on word co-occurrence for inference, i.e. the ordering of words is not taken into account when determining topics.

Our corpus presents three challenges when using such methods: (1) the high volume and short length of tweets result in sparse term-document matrices and limit the co-occurrences of rare words. (2) we want to compare discourse between at least two languages or language families, English and Slavic.

When taking multilingual corpora as input, “bag-of-word” models will naturally learn separate topics for each language, even though some of those topics might be semantically similar. And finally (3) we need to take into account the dynamics over several months of data. Each of these challenges have been identified in the context of topic modeling, and numerous models have been proposed to address the shortcomings of LDA (Jipeng et al. 2019; Albalawi, Yeap, et Benyoucef 2020). However, no single model has been developed to take as input both multilingual and time dynamic features. We experimented with several of them to get the best dynamics and cross-lingual comparisons.

Some of them lay in the same realm of generative, “bag-of-words” models as LDA and NMF, relying on different strategies of document aggregation to address the sparsity problem caused by the short length of tweets. We experimented with the biterm topic model (BTM, Yan et al., s. d.) which predicts the single most probable topic per document, based on the aggregation of word pairs within the whole corpus. We also tested GSDMM, (Yin et Wang 2014).

For the last decade, numerous other topic models have been relying on word embeddings, that is, distributed representations of words as vectors, whose distance similarity (such as cosine similarity) reflect semantic similarity. Among those, we experimented with the recent Contextualized Topic Model (CTM, Bianchi, Terragni, et Hovy 2021) in which LDA-generated topics are weighted by sentence embeddings from a large, pre-trained language model (LLM). These embeddings encapsulate not only semantic, but contextual similarity, i.e. a word will be attributed several vectors depending on its context.

This model also presents an interest for the sake of challenge (2), that is, multi-lingual text modelling. When using embeddings from a LLM trained on multi-lingual corpora such as “XLM-RoBERTa” (Barbieri, Anke, et Camacho-Collados 2022), a trained CTM may infer cross-lingual topics in a “zero-shot” fashion, that is, predict the topic mixture from language A that would be most similar to a document in target language B (Bianchi et al. 2021).

Finally, for the sake of challenge (3), dynamics examination, we ran models for different time segmentations of our data. We first ran the topic models on the overall 23-day period between Aug. 8, eve of the election, and Aug.31, and visualized the mixture of topics for each 24-hour time slice during that window. We then split our dataset in 24-hour time slices (at the experiment stage, we used days between August 9 and August 12, 2020) and 7-day slices and ran topic models on each of those slices. Some models also specifically take into account time data; whether by temporal tweet pooling (Culmer et Uhlmann 2021) or learning time labels for each document (Twitter-TTM, Sasaki, Yoshikawa, et Furuhashi 2014).

Benchmark and optimization of topic models

From models discussed above, we experimented with **LDA, NMF, CTM** and **BTM** ; in addition, we ran **NMF-tfidf** and **LDA-tfidf**; that is, those same two models, where the document-term co-occurrence matrix taken as input is weighted by Term frequency-Inverse document frequency, which penalizes the frequency of terms based on how many documents they occur.

We used the open-source python library OCTIS (Terragni et al. 2021) to run those models through different pipelines of pre-processing, time-slicing of data, separating languages vs entire dataset. We aimed to compare the performance of models, but also to gauge how a single model’s performance would vary for different values of parameter K (number of topics). We also use OCTIS to perform hyperparameter search & optimization prior to testing K.

Metrics

We used three metrics as a benchmark of model performance : **(1) C_V coherence** (Röder, Both, et Hinneburg 2015), which reflects how strongly pairs taken from the top-10 words of a topic support each other within the corpus, and is reported to correlate best with human evaluation of topic coherence (Chang et al. 2009; Syed et Spruit 2017). **(2) Topic diversity**, defined as the proportion of unique top-10 words within topics; and finally, **(3) qualitative evaluation** of the topics based on human interpretability, but also saliency in terms of our RQ. Qualitative evaluation was supported by LDAvis (Sievert et Shirley 2014)[2] which allows ranking of topic words according to their weight over the corpus, exclusivity to the topic, as well as visual clustering of topics based on mutual information through PCA.

Pre-processing

Our final pre-processing pipeline for tweet texts consists of : removal of URLs, emojis, user quotes, punctuation and whitespace; removal of stop-words; part-of-speech (POS) filtering to keep only nouns, named entities, adjectives, verbs, adverbs, interjections, numerals ; lemmatization, while preserving certain named entities such as лукашенка/лукашенко (Lukashenka/Lukashenko) which convey cultural meaning (Noubel et Edwards 2020). Further pre-processing steps are model-dependent, such as vocabulary filtering: we keep the entire vocabulary for BTM, only 10K most frequent tokens for LDA and NMF, 2K most frequent tokens for CTM. Finally, we exclude tokens occurring less than 5 times, and tokens occurring in more than 95% of the documents. Aggregation of tokens into bigram and trigram phrases might improve results. In parallel, we output a lightly pre-processed corpus (replacement of @ and URLs) for the purpose of fine-tuning the LLM component of the CTM model, as per the authors' recommendation.

Hyperparameter search

For LDA and CTM, we first perform Bayesian optimization of some hyperparameters, with the goal of maximizing C_V coherence through at least 50 iterations and 3 runs of each model per iteration (since for all models, training and inference is initialized with random weights and results vary at each run). We select hyperparameters that produced the best median result.

LDA : Dirichlet $\alpha = 0.112$ (distribution of topics over documents). For Russian, $\alpha = 0.26$

CTM : Number of embedding layers : 2 for English Language, 3 for Russian.

BTM : $\alpha = \frac{50}{K}$ (distribution of topics over documents), $\beta = 0.01$ (distribution of terms over topics)

Selection of the optimal number of topics

[2] We used two implementations, mainly PyLDAVis (https://github.com/bmabey/pyLDAvis) and TMplot (https://github.com/maximtrp/tmplot)

Then, rather than to determine K (number of topics) through Bayesian optimization, we systematically run 3 models for $K = 3,\ldots,12$ (for 1-day time slices), and $K = 5,\ldots,25$ (for the entire month of August). For CTM models, we also compare 3 different variants of the LLM (BERT) component, for each language[3].

Finally, we visualize the topics for the lowest value of K where C_V score is higher than topic diversity score ; that is, the best trade-off between those two metrics, as illustrated in the example result below.

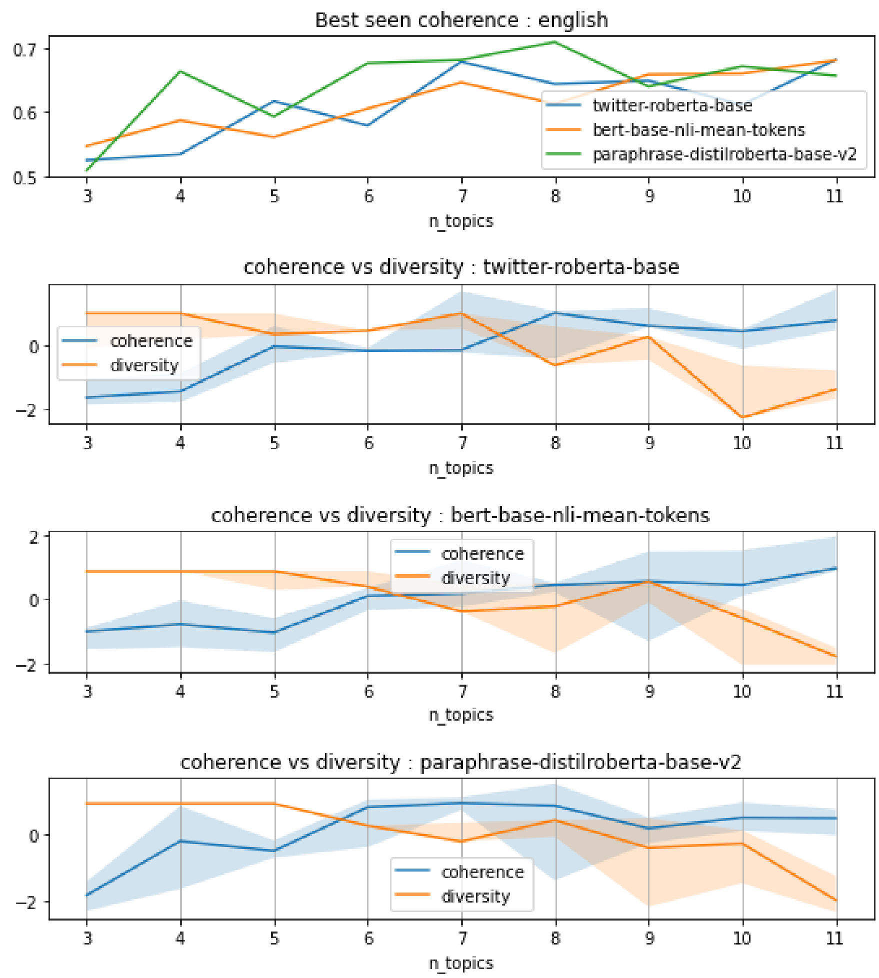


Figure 2 : CTM model benchmark, english tweets from August 10.
Top : Comparison of C_V coherence for three LLM components.
For each LLM we select the lowest K where median C_V > median diversity score, after standardization.

[3] Note that during the benchmarking and optimization stage, the heavily pre-processed text was being used as an input to the LLM component of CTM. Final CTM results benefit from using un-preprocessed tweets for training the LLM component, and preprocessed tweets for training the LDA component.

In some cases (LDA, LDA-tfidf on monthly data), there appeared to be no such trade-off: topic diversity increases with the number of topics. In those cases, we visualize the topics for the value of K which provides the highest median C_V.

Results for monthly data

Mixture models

The NMF models were found to consistently outperform LDA models for every value of K, when comparing C_V coherence (see fig. below).

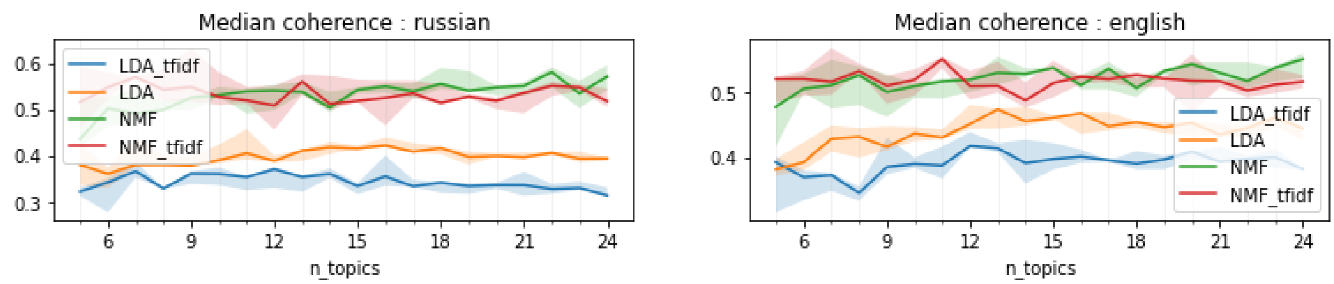
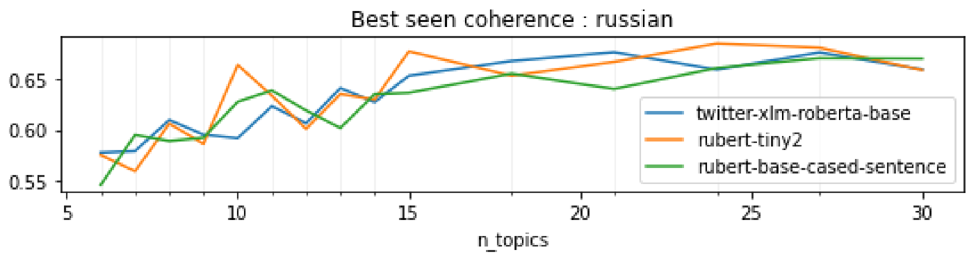


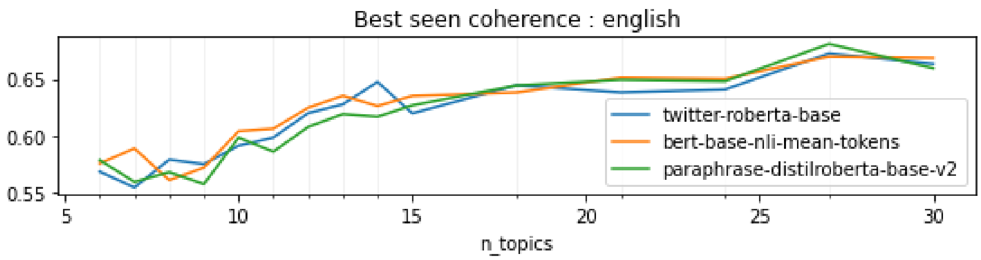
Figure 3 - Comparison of mixture models ; median C_V for K topics

CTM

All mixture models were outperformed by CTM, both in terms of C_V coherence and human evaluation. Variation in the LLM component had little impact on C_V coherence. Researcher's examination found sets of topics to be consistent with each other when varying LLM components.

For English language, the best result was attained using Twitter-RoBERTa (Barbieri, Anke, et Camacho-Collados 2022), a LLM trained on English Twitter data. For Russian, RuBERT-tiny (Dale 2021) provided the best C_V coherence, but its predecessor RuBERT (Kuratov et Arkhipov 2019) yielded more legible topics for monthly data (presented in supplements).





We present Russian and English CTM topics with $K = 15$.

Results for 30-day data (Aug 01 to Aug 31, 2020)

CTM, Standalone version

Russian topics (visualisation)	English topics (visualisation)
['правый (right)', 'the (the)', 'отец (father)', 'видеть (see)', 'лента (tape)', 'вклад (contribution)', 'народный (folk)', 'маленький (small)', 'вечный (eternal)', 'святой (St)', 'сын (son)', 'belarus (belarus)', 'текст (text)', 'мать (mother)', 'дочь (daughter)']	['protest', 'lukashenko', 'alexander', 'president', 'thousand', 'ten', 'election', 'alexanderlukashenko', 'rally', 'sunday', 'capital', 'mass', 'dispute', 'dictator', 'rig']
['корж (korzh)', 'макс (max)', 'слушать (listen)', 'нравиться (like)', 'пост (post)', 'песня (song)', 'пацан (boy)', 'тип (type)', 'написать (write)', 'петь (sing)', 'тепло (warm)', 'музыка (music)', 'понять (understand)', 'нахуй (fuck)', 'говно (shit)']	['stand', 'call', 'people', 'belarus2020', 'belarusprot', 'new', 'violent', 'europe', 'repression', 'беларусь', 'free', 'tell', 'place', 'peaceful', 'belarusian']
['человек (human)', 'народ (people)', 'власть (power)', 'должный (due)', 'избивать (beat)', 'ребёнок (child)', 'убивать (kill)', 'мирный (peaceful)', 'страна (country)', 'защищать (protect)', 'хотеть (to want)', 'право (right)', 'лукашенкоуходи (Lukashenka go away)', 'жизнь (life)', 'омон (riot police)']	['late', 'destroy', 'link', 'daily', 'website', 'thank', 'non', 'outlet', 'twitter', 'access', 'reach', 'available', 'coverage', 'online', 'read']
['belarusprotest (belarusprotest)', 'минск (Minsk)', 'бчб (white-red-white flag)', 'долой (Down with)', 'кровь (blood)', 'bialorus (bialorus)', 'belarus (belarus)', 'рука (hand)', 'прощать (forgive)', 'диктатура (dictatorship)', 'подписывать (to sign)', 'лукашэнка (Lukashenka)', 'беларусь (Belarus)', 'нехта (Nexta)', 'electby (Electby)']	['break', 'chain', 'add', 'жывебеларусь', 'лукашенкоуходи', 'живебеларусь', 'help', 'electby', 'минск', 'живэбеларусь', 'беларусь', 'lukashenkogoaway', 'минске', 'bialorus', 'personal']
['верить (believe)', 'беларусы (Belarusians)', 'белорус (Belarusian)', 'получиться (turn out)', 'надеяться (hope)', 'сила (force)', 'молодец (well	['russian', 'russia', 'putin', 'troop', 'kremlin', 'border', 'intervention', 'military', 'ukraine', 'vladimir', 'invade', 'moscow', 'nato', 'western', 'intervene']

done)', 'сильный (strong)', 'держаться (hold on)', 'хороший (good)', 'сдаваться (give up)', 'брат (brother)', 'свобода (freedom)', 'смелый (bold)', 'удача (luck)']	
['происходить (take place)', 'мочь (be able)', 'очень (very)', 'страшный (scary)', 'смотреть (watch)', 'читать (read)', 'ситуация (situation)', 'знать (know)', 'человек (human)', 'хотеть (to want)', 'жить (live)', 'страшно (scary)', 'пиздец (fucked up)', 'день (day)', 'страна (side)']	['humanity', 'great', 'worth', 'standwithbelarus', 'gas', 'together', 'savebelarus', 'stay', 'tyranny', 'fighter', 'fightforfreedom', 'strong', 'safe', 'freebelarus', 'highlightbelarus']
['окрестина (prison)', 'сизо (pre-trial detention center)', 'тюрьма (prison)', 'выпустить (release)', 'отпустить (let go)', 'рассказ (story)', 'выпускать (release)', 'освободить (release)', 'пытка (torture)', 'изолятор (insulator)', 'задержать (detain)', 'врач (doctor)', 'волонтёр (volunteer)', 'камера (camera)', 'рассказать (tell)']	['people', 'world', 'fight', 'democracy', 'get', 'freedom', 'need', 'internet', 'know', 'country', 'love', 'hope', 'brave', 'want', 'right']
['рабочий (worker)', 'метод (method)', 'совет (advice)', 'сопротивление (resistance)', 'гражданский (civil)', 'мзкт (mzkt)', 'налог (account)', 'культура (culture)', 'кандидат (candidate)', 'завод (plant)', 'работник (worker)', 'координационный (coordinating)', 'заявить (declare)', 'директор (Director)', 'неповиновение (disobedience)']	['strike', 'worker', 'employee', 'plant', 'today', 'factory', 'chant', 'tractor', 'square', 'independence', 'tv', 'city', 'rally', 'join', 'state']
['большой (big)', 'убийца (killer)', 'человечество (humanity)', 'усатый (whiskered)', 'таракан (cockroach)', 'жывэбелорусь (livebelarus)', 'европейский (European)', 'создать (create)', 'ус (water)', 'лукашенкокровавыйубийца (lukashenkabloodykiller)', 'лукашенкоубийца (Lukashenka killer)', 'ублюдок (bastard)', 'сообщество (community)', 'кровавый (bloody)', 'тварь (creature)']	['grenade', 'riot', 'bullet', 'police', 'force', 'rubber', 'shoot', 'beat', 'car', 'stun', 'man', 'video', 'officer', 'security', 'arrest']
['брест (elm)', 'мвд (Ministry of Internal Affairs)', '2020 (2020)', 'сайт (site)', 'полиция (police)', '24 (24)', '19 (19)', 'август (August)', 'память (memory)', 'гомель (Gomel)', 'область (region)', 'рубль (rubles)', 'полицейский (police officer)', 'brest (brest)', '08 (08)']	['eu', 'estonia', 'czech', 'sweden', 'latvia', 'france', 'sanction', 'germany', 'spain', 'meeting', 'foreign', 'discuss', 'netherlands', 'poland', 'turkey']
['стране (countries)', 'вновь (again)', 'занять (to occupy)', 'связаться (contact)', 'реагировать (to react)', 'остальные (rest)', 'положить (put)', 'волна (wool)', 'игнорировать (ignore)', 'тг (tg)', 'практически (practically)', 'вывод (conclusion)', 'ловить (catch)', 'врать (lie)', 'ресурс (resource)']	['breaking', 'convince', 'denounce', 'trigger', 'intimidate', 'previously', 'repress', 'thus', 'confrontation', 'publicly', 'elsewhere', 'opposite', 'heavily', 'damage', 'punish']
['продолжать (continue)', 'цепочка (chain)', 'продолжить (proceed)', 'belarusfreedom (belarusfreedom)', 'пожалуйста (please)', 'minsk (minsk)', 'живэбеларусь (livebelarus)', 'важный (important)', 'belaruspresidentialelection (belaruspresidentialelection)', 'отметить (Mark)', 'живебелорусь (livebelarus)', 'отмечать (note)', 'простить (forgive)', 'извинить (sorry)', 'belarus (belarus)']	['commission', 'result', 'opposition', 'candidate', 'tikhanovskaya', 'sviatlana', 'presidential', 'vote', 'svetlana', 'election', 'flee', 'coordination', 'leader', 'poll', 'winner']
['молотов (molotov)', 'коктейль (cocktail)', 'баррикада (barricade)', 'машина (machine)', 'бросать (throw)', 'протестовать (protest)', 'кидать (throw)', 'омон (riot police)', 'цветок (flower)', 'резиновый (rubber)', 'пушкинский (Pushkin)', 'камень (stone)', 'граната (grenade)', 'пуля (bullet)', 'строить (build)']	['belarusrevolution', 'belarusprotest', 'belarusfreedom', 'minskmaidan', 'tichanowskaja', 'white', 'limit', 'wall', 'flag', 'song', 'belaruselection', 'belarussolidarity', 'red', 'woman', 'humanrights']

['протест (protest)', 'минск (Minsk)', 'акция (action)', 'белоруссия (Belarus)', 'митинг (rally)', 'беларусь (Belarus)', 'новость (news)', 'belorussia (belorussia)', '97 (97)', 'белорусский (Belorussian)', 'эфир (ether)', 'хабаровск (Khabarovsk)', 'август (August)', 'прямой (straight)', 'правачеловека (human rights)']	['think', 'trump', 'look', 'bad', 'america', 'dictator', 'go', 'thing', 'make', 'know', 'see', 'well', 'mind', 'want', 'really']
['лукашенко (Lukashenka)', 'россия (russia)', 'беларусь (Belarus)', 'путин (Putin)', 'президент (the president)', 'выбор (choice)', 'украина (Ukraine)', 'страна (side)', 'власть (power)', 'тихановская (Tikhanovskaya)', 'республика (republic)', 'народ (people)', 'год (year)', 'диктатор (dictator)', 'белоруссия (Belarus)']	['violence', 'end', 'sign', 'brutal', 'demand', 'peaceful', 'right', 'crackdown', 'petition', 'happen', 'torture', 'protester', 'human', 'condemn', 'police']

BTM

(incoming)

Results for 1-day data split

CTM (Version : B6 CTM Days Octis)

RUSSIAN (CTM)	ENGLISH (CTM)
august 9, 4 topics found	**august 9, 5 topics found**
лукашенко выбор беларусь белоруссия тихановская belarus омон минск belorussia милиция ['Lukashenka', 'election', 'Belarus', 'Belarus', 'Tikhanovskaya', 'belarus', 'riot police', 'Minsk', 'belorussia', 'militia']	protester minsk police riot use city peaceful force grenade protestor
омоновец корж группа ехать атака также пустить беларуского мужчина ветер [riot policeman, 'Korzh', 'Group', 'drive', 'attack', 'also', 'let', 'Belarusian', 'the male', 'wind']	presidential election poll exit vote result lukashenko opposition president tikhanovskaya
сябрысила нереально беларуского единение выбараў таракан футбол забывать участие лукашеску ['syabrysila', 'unreal', 'Belarusian', 'unity', 'elections', 'cockroach', 'football', 'forget', 'participation', 'Lukashenko']	ukraine russia like guy look still much not do good
свой это страна народ белорус надеяться очень жывебеларусь просто происходить ['own', 'This', 'side', 'people', 'Belarusian', 'hope', 'very', 'live in Belarus', 'simply', 'take place']	people belarus belarusian freedom democracy lukashenko solidarity change dictator color
	eupol shameful organisation student escalation reject address director unity baton
august 10, 6 topics found	**august 10, 5 topics found**
коктейль молотов пушкинский омон граната протестовать пуля силовик баррикада площадь ['cocktail', 'molotov', 'Pushkin', 'riot police', 'grenade', 'protest', 'bullet', 'security officer', 'barricade', 'square']	police force report grenade city riot minsk protester car video
свой просто это страна народ человек происходить хотеть такой мочь ['own', 'simply', 'This', 'side', 'people', 'Human', 'take place', 'to want', 'such', 'be able']	legitimacy possibility resist conference path politics realise view guarantee disruption
навсегда вперёд напрасный тренд грустный радоваться восторжествовать вера вечнаяпамять очевидно ['forever and ever', 'forward', 'vain', 'trend', 'sad', 'rejoice', 'triumph', 'faith', 'everlasting memory', 'obviously']	break we жывебеларусь chain help minsk минске do bialorus white
диктатура bialorus долой electby кровь рука подписывать прощать беларусь жывебеларусь	get try people world internet fight would right беларусь allow

['dictatorship', 'bialorus', 'Down with', 'electby', 'blood', 'hand', 'to sign', 'forgive', 'Belarus', 'live in Belarus']	
minsk пожалуйста chain цепочка belarus belarusfreedom продолжать break don продолжить ['minsk', 'please', 'chain', 'chain', 'belarus', 'belarusfreedom', 'continue', 'break', 'don', 'proceed']	lukashenko election president result vote victory protest tikhanovskaya candidate opposition
лукашенко выбор белоруссия коррумпированного протест год сосать братский 14 беларусь ['Lukashenka', 'election', 'Belarus', 'corrupt', 'protest', 'year', 'suck', 'brotherly', '14', 'Belarus']	

august 11, 7 topics found	**august 11, 5 topics found**
молотов коктейль протестовать пушкинский силовик омон машина баррикада задержать минск ['molotov', 'cocktail', 'protest', 'Pushkin', 'security officer', 'riot police', 'machine', 'barricade', 'detain', 'Minsk']	space freedomforbelarus stuff shameful 21st mobile facebook asap unitedstate commit
тихановская выбор belarusprotest лукашенко беларусь белоруссия протест президент russia литва ['Tikhanovskaya', 'election', 'belarusprotest', 'Lukashenka', 'Belarus', 'Belarus', 'protest', 'the president', 'russia', 'Lithuania']	we break chain жывебеларусь do help not twibbon tag живебеларусь
корж макс пост песня коржа нравиться трек слушать месяц написать ['Korzh', 'max', 'post', 'song', 'crust', 'like', 'track', 'listen', 'month', 'write']	tangodown минск belaruselection belarusfreedom лукашенко police belarusprotest beat city minsk
minsk цепочка пожалуйста belarus belarusfreedom продолжить продолжать chain жывебеларусь the ['minsk', 'chain', 'please', 'belarus', 'belarusfreedom', 'proceed', 'continue', 'chain', 'live in Belarus', 'the']	people please eu riot belarusian internet stop support call right
человек свой народ лукашенкоуходи страна такой это просто власть которые ['Human', 'own', 'people', 'Lukashenka go away', 'side', 'such', 'This', 'simply', 'power', 'which']	election tikhanovskaya svetlana lithuania opposition candidate presidential flee president protest
верный пролить допустить дикий счёт сломать человечность превратиться оставлять противостоять ['loyal', 'shed', 'allow', 'wild', 'check', 'break', 'humanity', 'turn into', 'leave', 'resist']	
очень надеяться вами верить белорус происходить держаться весь всё получиться ['very', 'hope', 'you', 'believe', 'Belarusian', 'take place', 'hold on', 'the whole', 'all', 'turn out']	

august 12, 4 topics found	**august 12, 5 topics found**
свой человек просто это народ страна происходить такой которые этот ['own', 'Human', 'simply', 'This', 'people', 'side', 'take place', 'such', 'which', 'this']	election belarus eu president foreign violence belarusian lukashenko lithuania sanction
мурашка восхищение несправедливость придумать изверг лукаш скрутить другим доверие великий ['ant', 'Delight', 'injustice', 'come up with', 'monster', 'Lukash', 'twist', 'others', 'trust', 'large']	goal administration task introduce chaos regard decade voting seek present
minsk пожалуйста цепочка belarusfreedom belarus жывебеларусь продолжить живэбеларусь belarusprotest важный ['minsk', 'please', 'chain', 'belarusfreedom', 'belarus', 'live in Belarus', 'proceed', 'livebelarus', 'belarusprotest', 'important']	minsk police car woman protester force belarusprotest protest beat white
лукашенко протест беларусь белоруссия выбор минск задержать коктейль тихановская президент ['Lukashenka', 'protest', 'Belarus', 'Belarus', 'election', 'Minsk', 'detain', 'cocktail', 'Tikhanovskaya', 'the president']	people love freedom go not country internet see kill day
	жывебеларусь help please twibbon беларусь add break chain belarusfreedom we