

UNIVERSITÀ
DEGLI STUDI
DI PADOVA

Network Science

A.Y. 25/26

ICT for Internet & multimedia, Data science, Physics of data

Semantic networks

network science tools for their study

Data collection

how to get data from the Internet using APIs

kaggle

Search

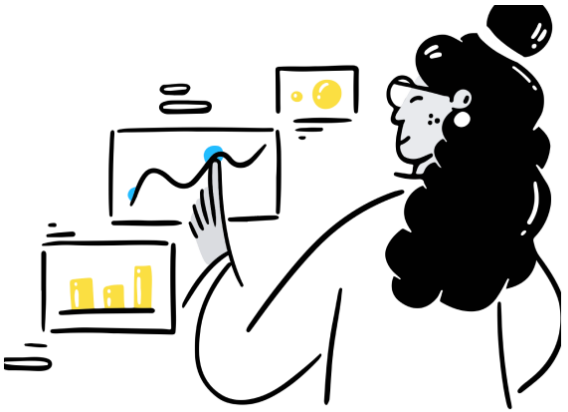
Sign In

Register

Datasets

Explore, analyze, and share quality data. Learn more about data types, creating, and collaborating.

+ New Dataset



Trending Datasets

See All



Air Quality Data

Yakhyojon · Updated a day ago
Usability **10.0** · 7 kB
1 File (CSV)



Airlines Dataset

Saad Haroon · Updated 15 hours ago
Usability 7.6 · 31 MB
1 File (SQLITE)



Electric Vehicle Population in Washington State

Zohair ahmed · Updated 3 days ago
Usability **10.0** · 5 MB
1 File (CSV)



UNIVERSITÀ DEGLI STUDI DI PADOVA

Twitter

no longer available unless you pay 5k$ per month ☹

https://developer.twitter.com/en/portal/dashboard

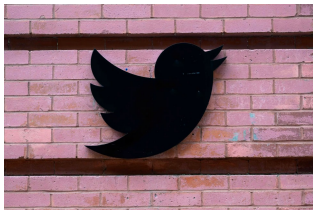
Twitter's plan to cut off free data access evokes 'fair amount of panic' among scientists

Social media platform's intent to increase revenue could end or limit many research projects

8 FEB 2023 · 4:35 PM ET · BY KAI KUPFERSCHMIDT

Twitter's plan to charge researchers for data access puts it in EU crosshairs

Elon Musk's social media giant plans to charge academics to access its data – in potential violation of Europe's content rules



BY MARK SCOTT
MARCH 22, 2023

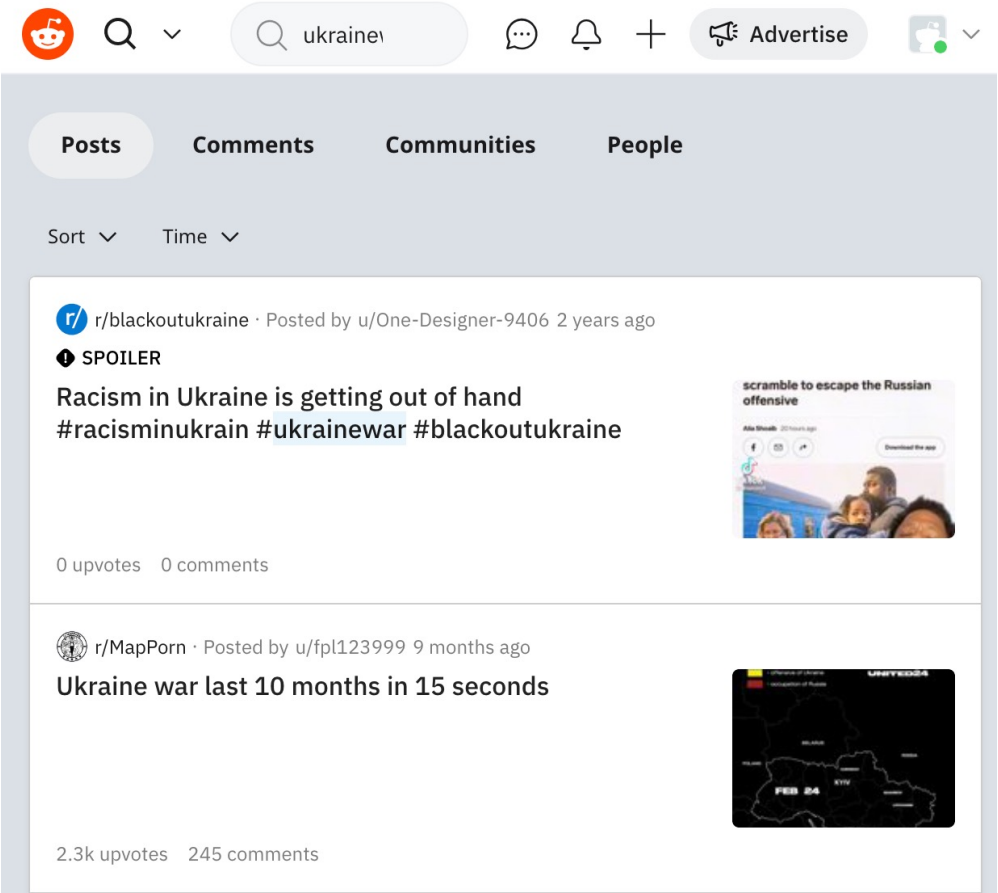
Academic researchers blast Twitter's data paywall as 'outrageously expensive'



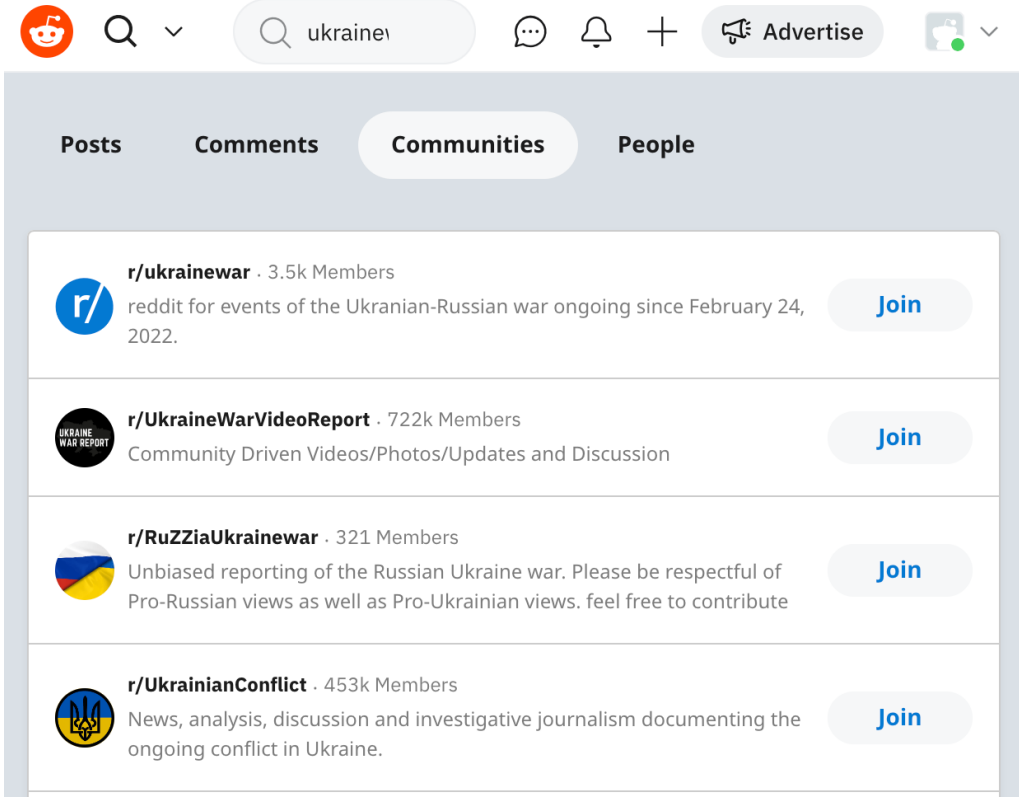
By Brian Fung, CNN

Published 11:40 AM EDT, Wed April 5, 2023

Reddit



Subreddit



Reddit apps

https://www.reddit.com/prefs/apps

GET NEW REDDIT | MY SUBREDDITS | HOME - POPULAR - ALL - RANDOM - USERS | | ASKREDDIT - GAMING - PICS - TODAYILEARNED - FUNNY - WORLDNEWS - NEWS - MOVIES - N

reddit PREFERENCES options apps RSS feeds friends blocked password/email delete Upbeat-Lychee-6630 (1) |

developed applications

ns2023

personal use script

Qbdk-FkA9jSQB9T7drY8UQ

client_id

download reddit context for the network science course, at the unioversity of Padova, of which I am the instructor

change icon

username

secret jtGPdqiaTj6hCWcvRPeS_nMNEnVkxw

client_secret

developers Upbeat-Lychee-6630 that's you!) remove

name ns2023

add developer:

description download reddit context for the network science course, at the unioversity of Padova,

about url

redirect uri https://localhost:8080

update app delete app

create another app...

register asap… will be using this in the 1st lab

Python Reddit API Wrapper – PRAW

https://praw.readthedocs.io/en/stable/ (v7.7.1)

```
!pip install praw
```

```
import pandas as pd
import praw
reddit = praw.Reddit(client_id='Qbdk-FkA9jSQB9T7drY8UQ',
                     client_secret="jtGPdqiaTj6hCWcvRPeS_nMNEnVkxw",
                     user_agent='reddit scraper 1.0 by u/Upbeat-Lychee-6630',
                     check_for_async=False)
print(reddit.read_only)
```

```
True
```

```
df = pd.DataFrame([vars(post) for post in reddit.subreddit("all")
                   .search("#ukrainewar", sort='top', limit=10)])
```

```
df.to_excel('drive/MyDrive/Colab Notebooks/samples.xlsx',
            index=True)
```



Reddit data
an example

there is a list of 116 entries per post, on which you can choose!!!

from this you extract the date

	title	created	score	upvote_ratio	ups	num_comments	selftext
0	Damn...we blinked and missed the T-34 stage of...	1.666899e+09	10394	0.99	10394	738	
1	Finnish🇫🇮 volunteer sends greetings home from ...	1.680237e+09	2095	1.00	2095	57	
2	Guess having 5 trucks fall into your office ca...	1.663341e+09	1974	1.00	1974	88	
3	[META] Important - Russia-Ukraine Crisis/War: ...	1.645712e+09	1284	0.89	1284	1	Hi, /u/Anonim97 here.\n\nWe - as a mods of 40k...
4	V*tniks coping hard Over the counter offensive...	1.662956e+09	1081	1.00	1081	86	
...	...	...	...	...	...	...	...
0	Russia Ukraine War.	1.695349e+09	1	1.00	1	0	

TikTok for developers

Products Docs Support Blog

- Content Posting API
- Display API
- Research API
  - About Research API
  - Getting Started
  - Frequently Asked Questions
  - Codebook
  - API Reference
    - Query Videos
    - Query User Info
    - Query Video Comments

Query Videos

Request

HTTP URL	https://open.tiktokapis.com/v2/research/video/query/
HTTP Method	POST
Scopes	research.data.basic

Headers

Key	Type	Description

TikTok for developers

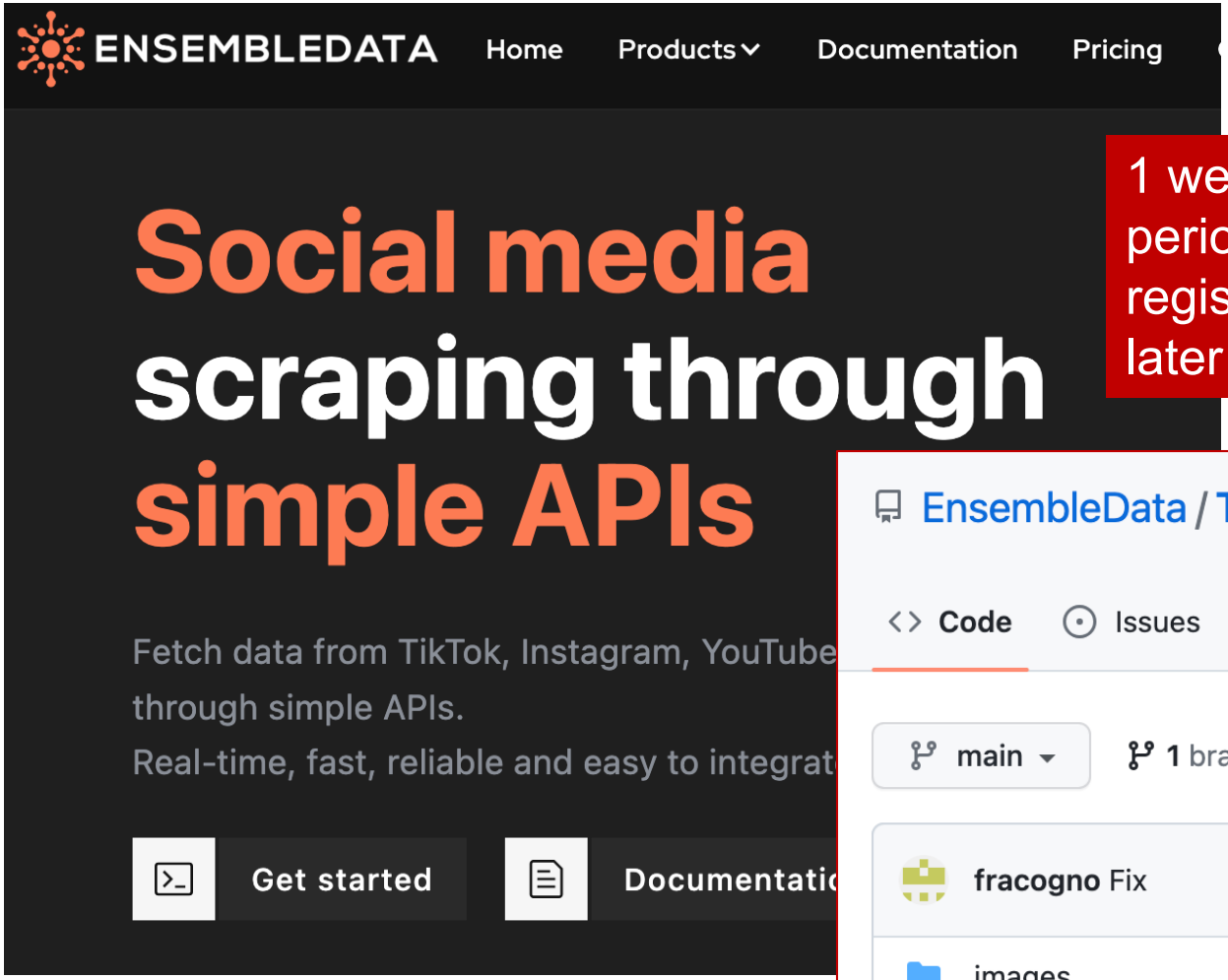


TikTok Research API application received

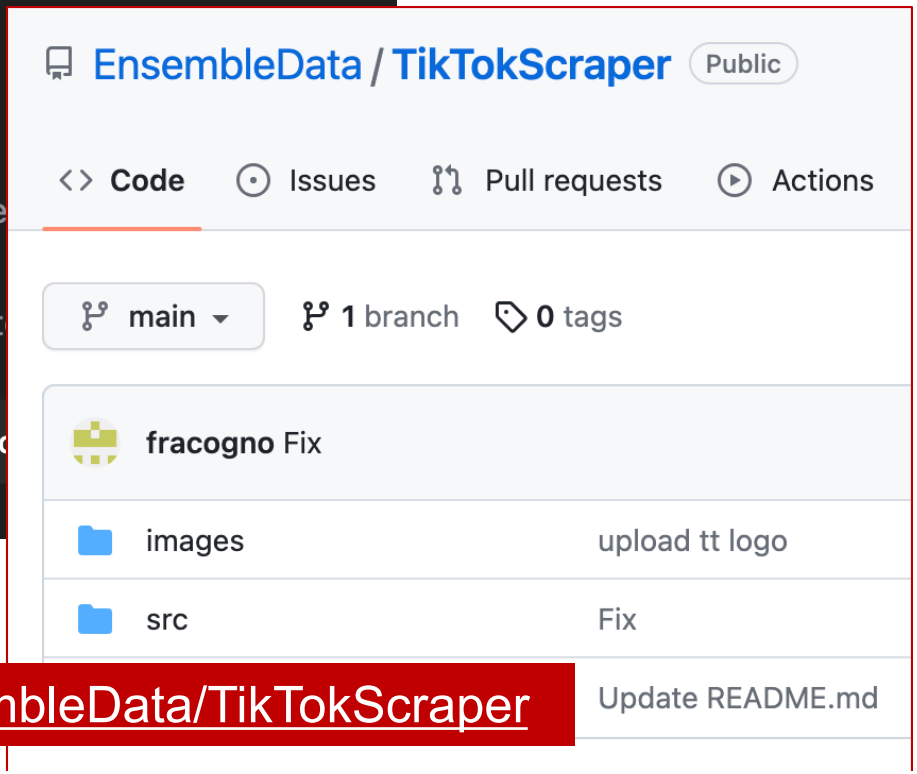
Thank you for applying to gain access to TikTok's research APIs. We have received your application and will proceed with the review process. You will be notified of the result via email within 3-4 weeks. If you have any questions, please contact our Support team.

TikTok

https://www.ensembledata.com/



1 week trial period… register later on



https://github.com/EnsembleData/TikTokScraper

Data preprocessing

how to polish raw data from the Internet

NLP cleaning process
of text sources

1. Superficial cleaning

Removing website links
Removing accented characters
Removing text inside square brackets
Removing moderator messages
Removing double spaces
Removing non-text special words and characters
Removing extra-used new lines
Limiting all the repetitions to two characters and removing the extra characters
Removing punctuation except main sentence punctuation
Removing sentences that represent the rules of the community

Fixing **contractions**
Removing **emoji**
Removing **hashtags** and **mentions**
Removing **numbers**
Lowercasing
Correct **spellings**

2. Subsentence

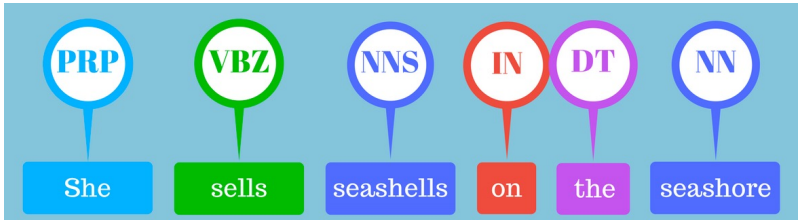
Tokenise subsentences

useful for long text samples (e.g., Reddit)

3. Deep cleaning

Stop word removal
Word tokenization
POS tagging
Lemmatization

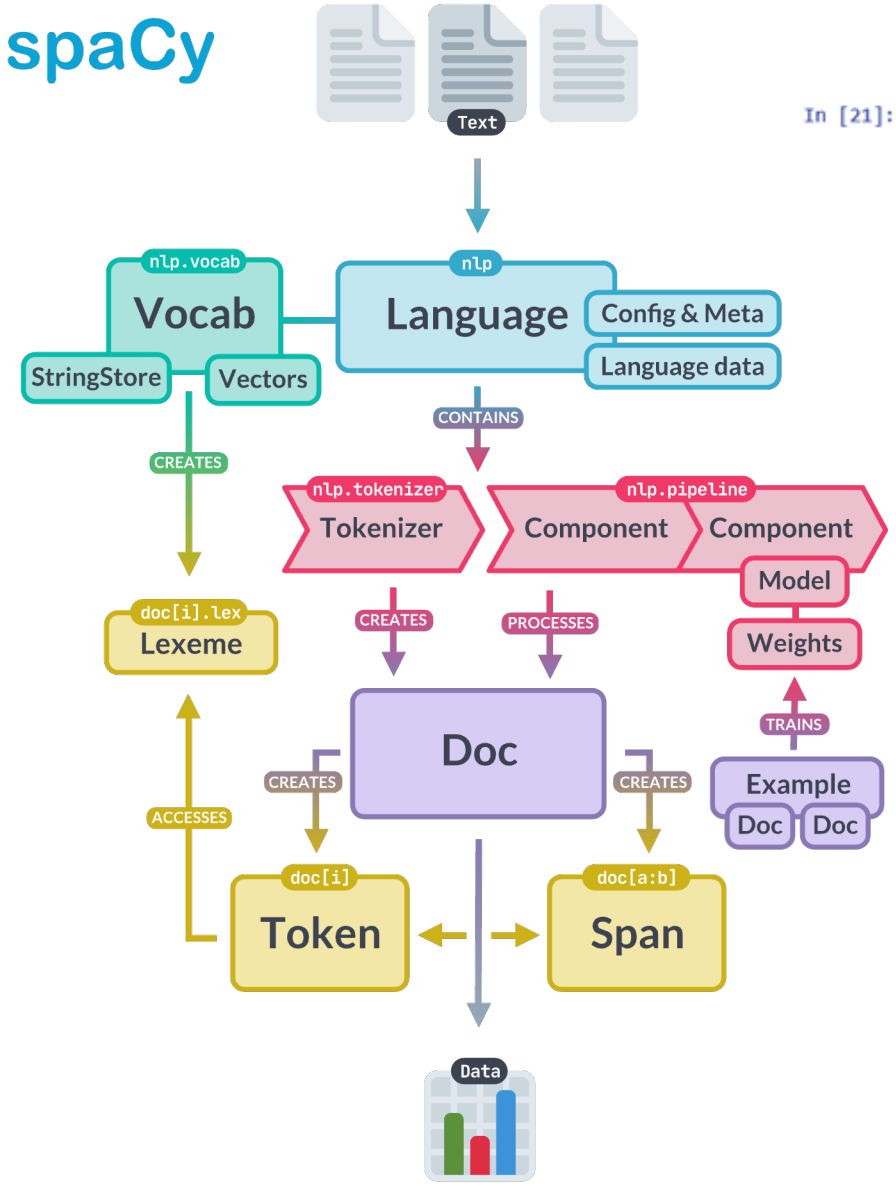
truly polished text, useful for building a semantic network



the bare minimum to polish the text, useful as an input to sentiment analysis

SpaCy

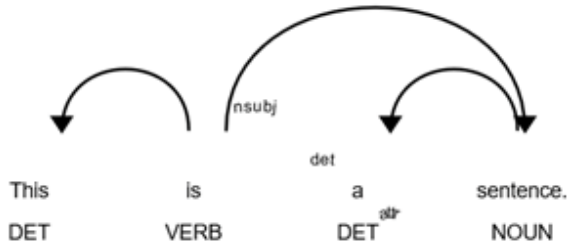
https://spacy.io/

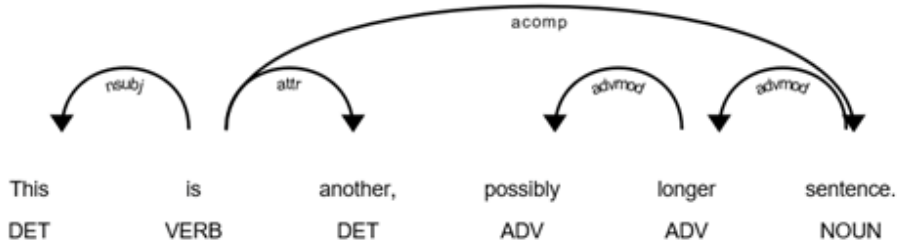



```
In [21]: doc2 = nlp(u"This is a sentence. This is another, possibly longer sentence.")

         # Create spans from Doc.sents:
         spans = list(doc2.sents)

         displacy.render(spans, style='dep',jupyter=True, options={'distance': 110})
```







UNIVERSITÀ DEGLI STUDI DI PADOVA

SpaCy part-of-speech (POS) tags

https://spacy.io/

POS	description	example
ADJ	**adjective**	big, old, green, incomprehensible, first
ADP	adposition	in, to, during
ADV	**adverb**	very, tomorrow, down, where, there
AUX	auxiliary	is, has (done), will (do), should (do)
CONJ	conjunction	and, or, but
CCONJ	coordinating conjunction	and, or, but
DET	determiner	a, an, the
INTJ	interjection	psst, ouch, bravo, hello
NOUN	**noun**	girl, cat, tree, air, beauty
NUM	numeral	1, 2017, one, seventy-seven, IV, MMXIV

POS	description	example
PART	particle	’s, not,
PRON	**pronoun**	I, you, he, she, myself, themselves, somebody
PROPN	**proper noun**	Mary, John, London, NATO, HBO
PUNCT	punctuation	., (,), ?
SCONJ	subordinating conjunction	if, while, that
SYM	symbol	$, %, §, ©, +, −, ×, ÷, =, :), 😝
VERB	**verb**	run, runs, running, eat, ate, eating
X	other	sfpksdpsxmsa
SPACE	space	

spaCy

Example

ukrainewar from Reddit

true date

superficial cleaning

deep cleaning

title	created	score	upvote_ratio	selftext	title_sup_clean	title_deep_clean	title_deep_clean_pos
Damn...we blinked and missed the T-34 stage of...	2022-10-27	10390	0.99	NaN	damn we blinked and missed the t stage of the ...	damn blink miss t stage war	[damn ADV, blink VERB, miss VERB, t PROPN, sta...
Finnish🇫🇮 volunteer sends greetings home from ...	2023-03-31	2095	1.00	NaN	finnish volunteer sends greetings home from so...	finnish volunteer send greeting home	[finnish ADJ, volunteer NOUN, send VERB, greet...
Guess having 5 trucks fall into your office ca...	2022-09-16	1980	1.00	NaN	guess having trucks fall into your office can ...	guess have truck fall office significant emoti...	[guess VERB, have VERB, truck NOUN, fall VERB,...
[META] Important - Russia-Ukraine Crisis/War: ...	2022-02-24	1280	0.89	Hi, /u/Anonim97 here.\n\nWe - as a mods of 40k...	important russia ukraine crisis war info and ...	important russia ukraine crisis war info way help	[important ADJ, russia PROPN, ukraine PROPN, c...
V*tniks coping hard Over the counter offensive...	2022-09-12	1076	1.00	NaN	v tniks coping hard over the counter offensive...	tnik cope hard counter offensive traitor pfp lmao	[tnik NOUN, cope VERB, hard ADJ, counter NOUN,...
...	...	...	...	...	...	...	...

only ADJ, ADV, NOUN, PRON, PROPN, VERB kept

Sentiment analysis

adding useful insights to your data

Socio-psychological markers
beyond simple sentiment

- Sentiment – e.g., positive, negative, neutral
  enduring cognitive content that defines the affective state
- Emotion – e.g., anger, disgust, fear, joy, sadness
  intense affective state of short duration with a precise cause
- Ingroup bias – e.g., use of pronouns I, we, us
  tendency to favor one's own group over other groups
- Outgroup bias – e.g., use of pronoun they
  tendency to dislike members of groups we don't identify with
- Agency – e.g., use of action verbs do, take, make
  perception that an individual is able to contribute to/a group can collectively reach a social change

LIWC linguistic inquiry and word count

Tausczik, Pennebaker. "The psychological meaning of words: LIWC and computerized text analysis methods." (2010)

https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=79d2494cc10a9633c42115df84bb74ed447080f6



https://www.liwc.app/

- word count (or dictionary) methodology
- over 60 dictionaries coded and validated for their accuracy in reflecting psychological content
- simplicity of implementation and usage
- state-of-the-art in psychology
- one licence available in the instructor's PC ☺

LIWC categories
ingroup and outgroup

Category	Examples	Words in Category	Psychological Correlates
Linguistic processes			
Word count			Talkativeness, verbal fluency
Words/sentence			Verbal fluency, cognitive complexity
Dictionary words	(Percentage of all words captured by the program)		Informal, nontechnical language
Words >6 letters	(Percentage of all words longer than 6 letters)		Education, social class
Total function words		464	
Total pronouns	I, them, itself	116	Informal, personal
Personal pronouns	I, them, her	70	Personal, social
First-person singular	I, me, mine	12	Honest, depressed, low status, personal, emotional, informal
First-person plural	We, us, our	12	Detached, high status, socially connected to group (sometimes)
Second person	You, your, thou	20	Social, elevated status
Third-person singular	She, her, him	17	Social interests, social support
Third-person plural	They, their, they'd	10	Social interests, out-group awareness (sometimes)

ingroup

outgroup

LIWC categories

goal orientation, aggression, social concern, emotionality

Category	Examples	Words in Category	Psychological Correlates
Indefinite pronouns	It, it's, those	46	
Articles	A, an, the	3	Use of concrete nouns, interest in objects and things
Common verbs	Walk, went, see	383	
Auxiliary verbs	Am, will, have	144	Informal, passive voice
Past tense	Went, ran, had	145	Focus on the past
Present tense	Is, does, hear	169	Living in the here and now
Future tense	Will, gonna	48	Future and goal oriented
Adverbs	Very, really, quickly	69	
Prepositions	To, with, above	60	Education, concern with precision
Conjunctions	And, but, whereas	28	
Negations	No, not, never	57	Inhibition
Quantifiers	Few, many, much	89	
Numbers	Second, thousand	34	
Swear words	Damn, piss, fuck	53	Informal, aggression,
Psychological processes			
Social processes	Mate, talk, they, child	455	Social concerns, social support
Family	Daughter, husband	64	
Friends	Buddy, friend, neighbor	37	
Humans	Adult, baby, boy	61	
Affective processes	Happy, cried, abandon	915	Emotionality

focus on past, present or future

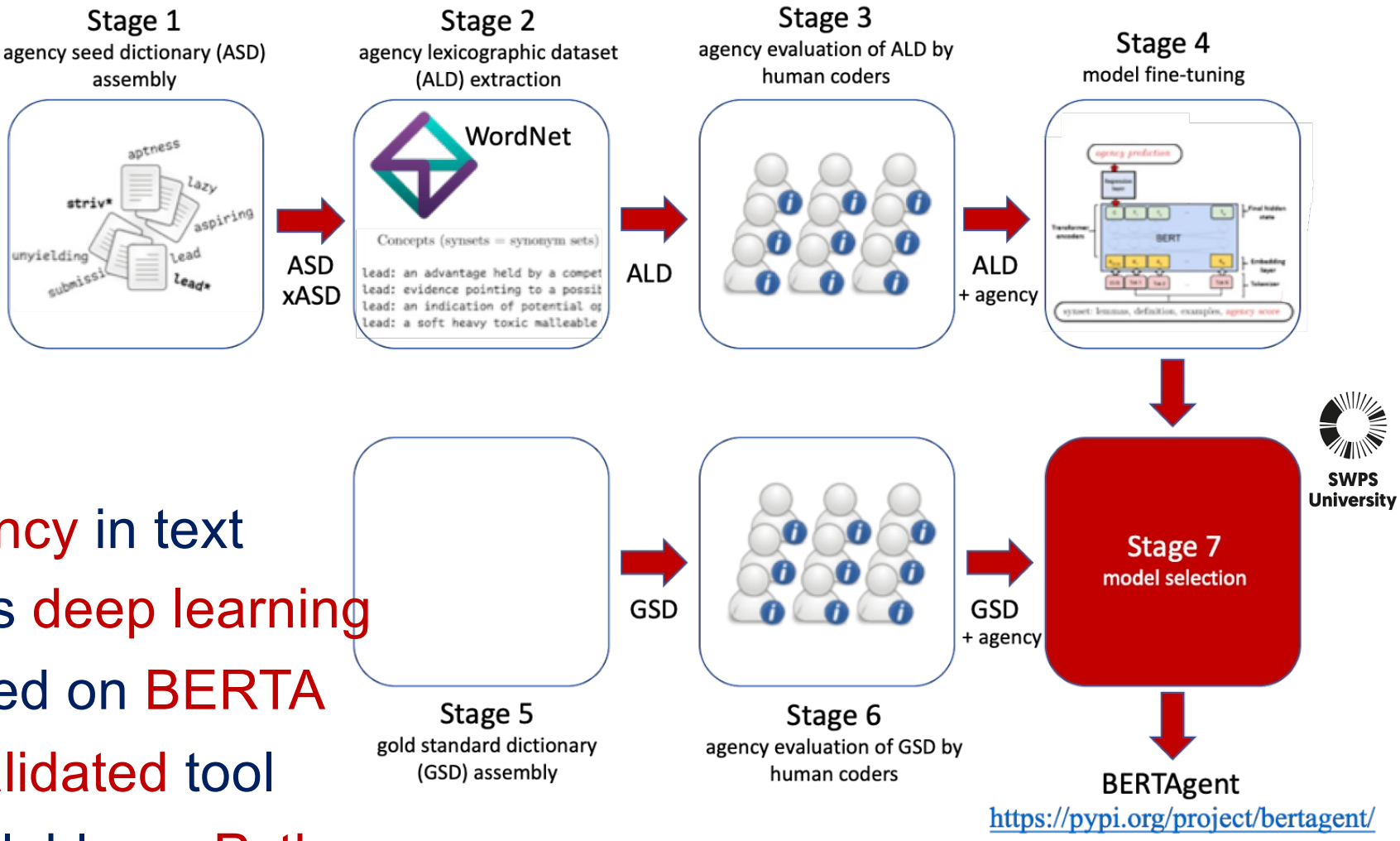
LIWC categories
a full list

WC	Analytic	Clout	Authentic	Tone	WPS	Sixltr	Dic	function	pronoun
ppron	**i**	**we**	**you**	**shehe**	**they**	ipron	article	prep	auxverb
adverb	conj	negate	verb	adj	compare	interrog	number	quant	affect
posemo	**negemo**	**anx**	**anger**	**sad**	social	family	friend	female	male
insight	cause	discrep	tentat	certain	differ	percept	see	hear	feel
bio	**body**	**health**	**sexual**	ingest	drives	**affiliation**	**achieve**	**power**	**reward**
risk	**focus past**	**focus present**	**focus future**	relativ	motion	space	time	work	leisure
home	money	**relig**	death	**informal**	**swear**	netspeak	assent	nonflu	filler
AllPunc	Period	Comma	Colon	SemiC	QMark	Exclam	Dash	Quote	Apostro
Parenth	cogproc								

Choose the ones of interest to your project!

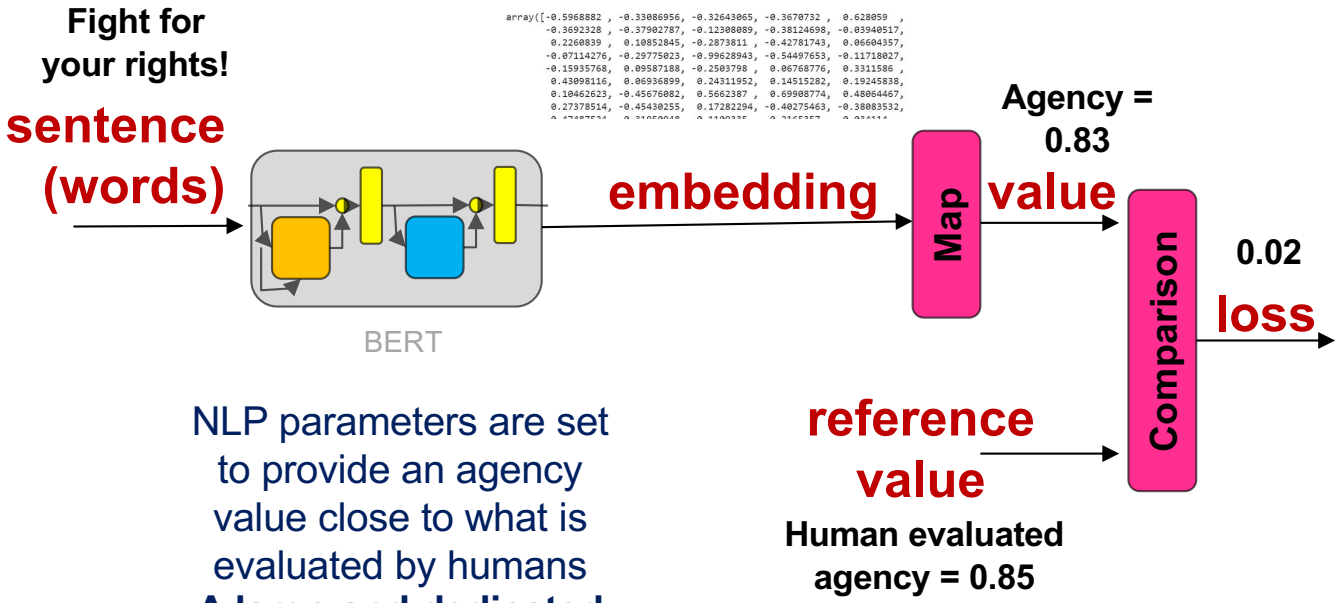
BERTAgent

Nikadon et al., «BERTAgent: A novel tool to quantify agency in textual data,» (2023) https://psyarxiv.com/qw6u3



- agency in text
- uses deep learning
- based on BERTA
- a validated tool
- available on Python

BERT
Training a NLP tool



NLP parameters are set to provide an agency value close to what is evaluated by humans

A large and dedicated dataset is needed!



Validation of BERTAgent

deep learning wins versus DWC = dictionary word count

Variable	*M*	*SD*	1	2	3	4	5	6	7	8
1. HumEval	0.12	1.54								
2. PietA	0.05	0.05	.17** [.06, .28]		-1.25	0.28	0.05	5.35**	-1.78	-10.95**
3. PietB	0.02	0.03	.25** [.14, .35]	.40** [.30, .49]		1.27	1.16	6.58**	-0.70	-10.00**
4. PietC	0.05	0.05	.17** [.06, .28]	.99** [.99, 1.00]	.40** [.30, .49]		0.03	5.34**	-1.80	-10.93**
5. NicoPos	0.03	0.04	.17** [.05, .27]	.18** [.07, .29]	.23** [.12, .34]	.17** [.06, .28]		5.49**	-3.81**	-11.08**
6. NicoNeg	0.01	0.03	-.28** [-.38, -.17]	-.10 [-.21, .01]	-.01 [-.12, .11]	-.10 [-.21, .02]	-.03 [-.14, .09]		-5.73**	-13.40**
7. NicoCom	0.02	0.05	.30** [.19, .40]	.20** [.09, .31]	.19** [.08, .30]	.19** [.08, .30]	.82** [.78, .85]	-.60** [-.67, -.52]		-10.38**
8. BATot	0.09	0.35	.78** [.73, .82]	.21** [.10, .31]	.24** [.13, .34]	.20** [.09, .31]	.22** [.11, .33]	-.42** [-.51, -.33]	.42** [.33, .51]	

Human evaluation

BERTAgent

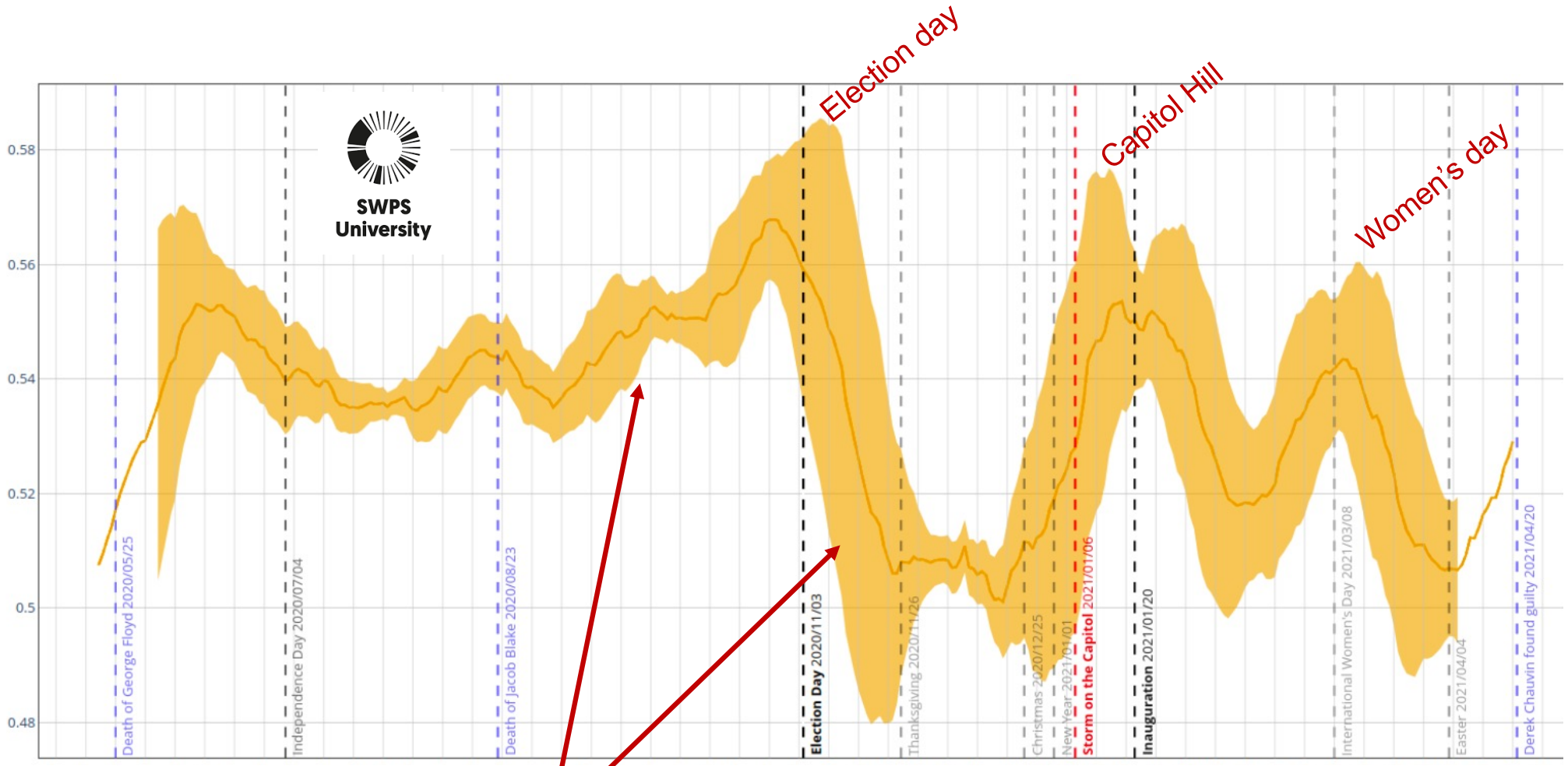
best correlation with Human evaluation

Z-statistics: correlation is statistically more relevant that DWC

Agency in US elections

Twitter, 2020-2021

by Jan Nikadon @ swps



twitter

agency raises before elections than drops

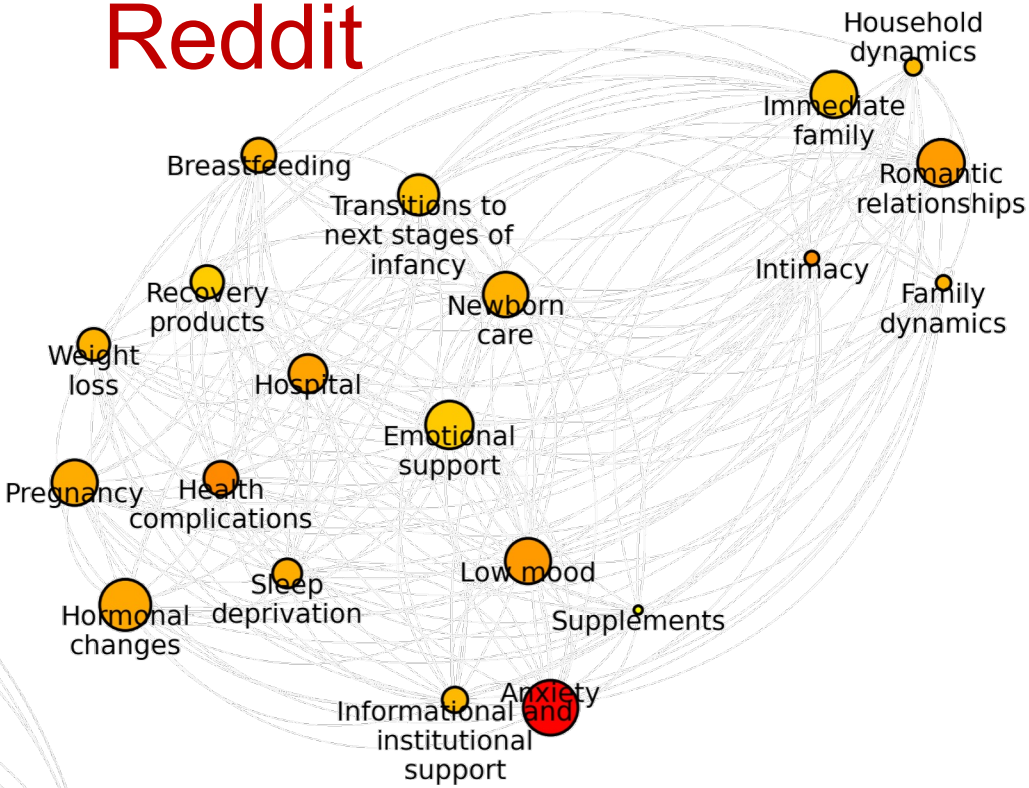
BERTAgent
at work

	translated	agency
0	I MISSED YOU GUYSS	-0.331446
1	we love you so much namjoon🥰💜💜💜	0.009748
2	haters: 🙄😡😤🦣🎃\nBP: 💗👄🎀💸💋	-0.261824
3	BLACKPINK BEST GROUP IN THE WORLD💯🔥🔥	0.058124
4	I love Changbin's humor 😭😭❤️	0.077732
...	...	...
153	My legs are so big	-0.065885
154	ne ne	-0.124783
155	I think not being able to see it live will be ...	-0.230997
156	MY BABIES	0.024777
157	I learned it from the foot steps.	0.133080

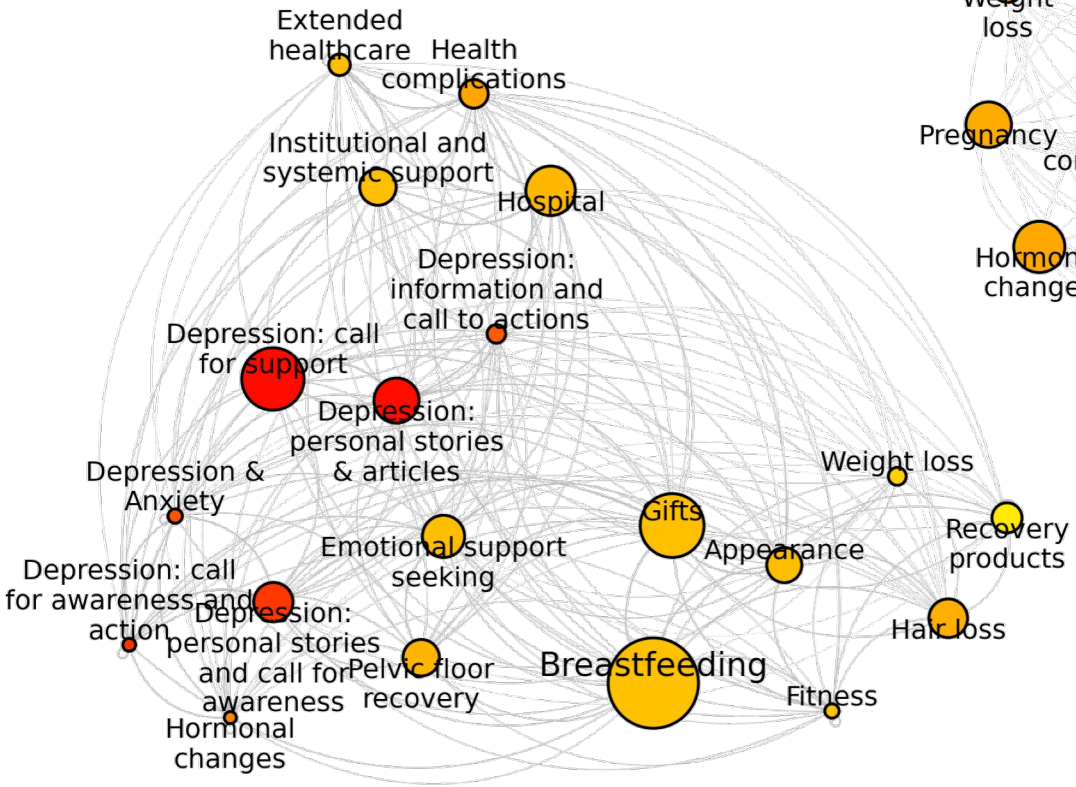
Agency in postpartum depression

Topics view

Reddit



Twitter





UNIVERSITÀ DEGLI STUDI DI PADOVA

Cardiff NLP on emotions

https://huggingface.co/cardiffnlp/twitter-roberta-large-emotion-latest

	translated	anger	anticipation	disgust	fear	joy	love	optimism	pessimism	sadness	surprise	trust
0	I MISSED YOU GUYSS	0.0121	0.0890	0.0381	0.0474	0.1375	0.2628	0.0180	0.1967	0.9927	0.0232	0.0251
1	we love you so much namjoon🥰💜💜💜	0.0183	0.0224	0.0183	0.0096	0.9826	0.9934	0.4805	0.0186	0.0637	0.0150	0.0883
2	haters: 🙄😡😤🦡🎃 \nBP: 💖👄🎀🎁💋	0.9528	0.0583	0.4327	0.0054	0.8497	0.0789	0.5624	0.0130	0.0477	0.0095	0.0380
3	BLACKPINK BEST GROUP IN THE WORLD💯🔥🔥	0.0203	0.2438	0.0112	0.0068	0.9874	0.8841	0.6740	0.0142	0.0147	0.1182	0.6141
4	I love Changbin's humor 😭😭❤️	0.0136	0.0297	0.0149	0.0113	0.9856	0.9942	0.5646	0.0188	0.0732	0.0193	0.1168
...	...	...	...	...	...	...	...	...	...	...	...	...
153	My legs are so big	0.0138	0.1850	0.0256	0.2976	0.5866	0.0316	0.1099	0.0345	0.2241	0.2603	0.0221
154	ne ne	0.0241	0.0774	0.0888	0.0284	0.0159	0.0024	0.0128	0.0415	0.1520	0.0087	0.0044
155	I think not being able to see it live will be ...	0.0114	0.1195	0.0499	0.0344	0.0823	0.0305	0.0493	0.4038	0.9966	0.0183	0.0148
156	MY BABIES	0.0137	0.0416	0.0367	0.0604	0.7553	0.9666	0.0974	0.0536	0.7570	0.0205	0.0484
157	I learned it from the foot steps.	0.0061	0.4315	0.0093	0.0165	0.0962	0.0099	0.5374	0.0292	0.0271	0.0167	0.1124

158 rows × 12 columns

Cardiff NLP on sentiment

https://huggingface.co/cardiffnlp/twitter-roberta-base-sentiment-latest

	translated	sentiment	score
0	I MISSED YOU GUYSS	positive	0.9299
1	we love you so much namjoon🥰💜💜💜	positive	0.9859
2	haters: 🙄😡😤🦣🎃\nBP: 💗👄🎀💸💋	negative	0.4806
3	BLACKPINK BEST GROUP IN THE WORLD💯🔥🔥	positive	0.9798
4	I love Changbin's humor 😭😭❤️	positive	0.9793
...	...	...	...
153	My legs are so big	neutral	0.4214
154	ne ne	neutral	0.5957
155	I think not being able to see it live will be ...	negative	0.8737
156	MY BABIES	positive	0.9029
157	I learned it from the foot steps.	neutral	0.6420

Building the semantic network

bipartite and projected counterparts

Probability matrices
linking words to documents

number of occurrences of words in documents

N_{wd} =

0	1	1	1	#globalwarming
1	1	1	1	#climatechange
1	0	1	0	#climateaction
0	0	1	1	#gretathunberg
1	0	1	1	#environment

probability of words given a documents

$P_{w|d}$ =

0	½	⅕	¼
⅓	½	⅕	¼
⅓	0	⅕	0
0	0	⅕	¼
⅓	0	⅕	¼

we capture the statistical properties by normalizing by columns

we identify a document probability

$$p_d = \begin{cases} \frac{1}{D} & \text{equally likely} \\ \frac{n_d}{\sum_d n_d} & \text{custom} \end{cases}$$

Probability matrices
projecting to words or documents

bipartite network

joint probability of words and documents

$\boldsymbol{P}_{wd} = \boldsymbol{P}_{w|d}\,\mathrm{diag}(\boldsymbol{p}_d)$

0	$1/8$	$1/20$	$1/16$
$1/12$	$1/8$	$1/20$	$1/16$
$1/12$	0	$1/20$	0
0	0	$1/20$	$1/16$
$1/12$	0	$1/20$	$1/16$

marginal probabilities

$\boldsymbol{p}_w = \boldsymbol{P}_{wd}\,\boldsymbol{1}$ $\quad \boldsymbol{p}_d = \boldsymbol{P}_{wd}^{\mathsf{T}}\,\boldsymbol{1}$

$$p_{w_1,w_2} = \sum_d p_{w_1|d,\cancel{w_2}}\, p_{w_2,d}$$

$$\boldsymbol{P}_{ww} = \boldsymbol{P}_{wd}\,\mathrm{diag}(\boldsymbol{p}_d)^{-1}\,\boldsymbol{P}_{wd}^{\mathsf{T}}$$

$$\boldsymbol{p}_w = \boldsymbol{P}_{ww}\,\boldsymbol{1}$$

projection on words

projection on documents

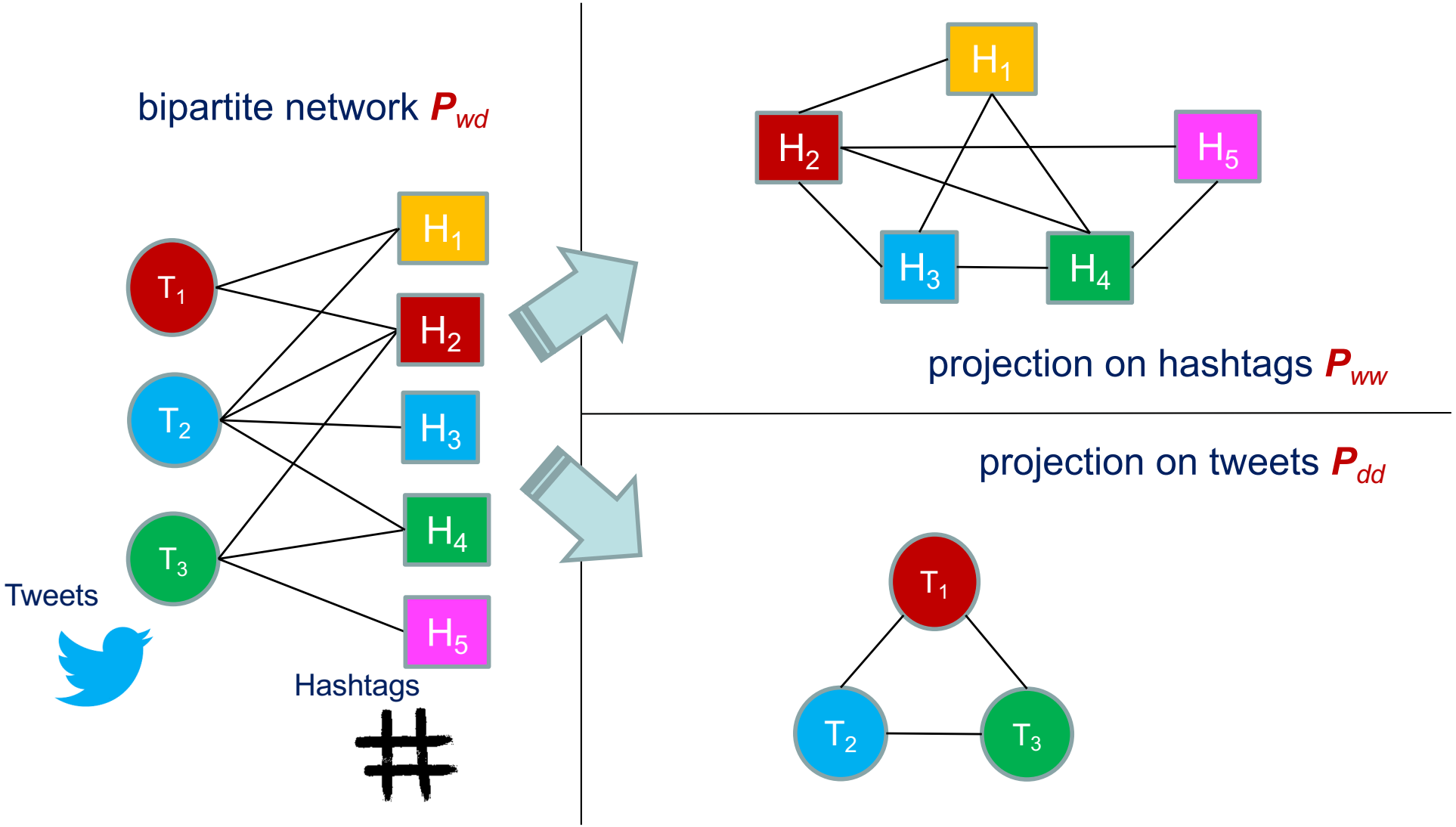
$$p_{d_1,d_2} = \sum_w p_{d_1|w,\cancel{d_2}}\, p_{d_2,w}$$

$$\boldsymbol{P}_{dd} = \boldsymbol{P}_{wd}^{\mathsf{T}}\,\mathrm{diag}(\boldsymbol{p}_w)^{-1}\,\boldsymbol{P}_{wd}$$

$$\boldsymbol{p}_d = \boldsymbol{P}_{dd}\,\boldsymbol{1}$$

Bipartite and projected networks

a comparison



Projection on words

an example


violence
work
check
harassment
say
power
continue
join
day
action
promote
meet
today
need
empowerment
discussion
ensure
country
equality
find
gender
society
director
people
discrimination
men
work
woma
make
support
health
education
achieve
year
commitment
story
help
right
want
see
time
business
stand
girl
life
world
community
celebrate
family
take
speak
activist
home
movement
hear
stereotype
inspire
voice
share
must
become
role
everyone
participation
leadership
empower
start
bring
learn
tell
give
believe
event
live
diversity
marriage
tool
change
difference
decision
position
process
disability
politics
inequality
fight
member
agree
participate
programme
policy
gap
camp
income
crisis
culture
dialogue
security
progress
worker
access
win
enter
show
interview
lead
reach
recognize
respond
village
inclusion
person
month
way
organization
responsibility
husband
daughter
protection
advocate
use
amplify
household
example
poverty
barrier
launch
think
happen
raise
leave
seek
medium
ambassador
principle
entrepreneur
student
transform
sport
bias
gender
survivor
provide
form
prevent
flee
conflict
highlight
condition
impunity
hold
workplace
ask
generation
conversation
space
play
part
goal
refugee
fact
get
know
let
child
expert
reduce
thank
leader
opportunity
council
development
session
improve
message
mutilation
commit
catch
line
weekend
participant
launch
framework
trafficking
accountability
cooperation
practice
impact
priority
follow
support
run
create
contribution
begin
journalist
mother
benefit
perspective
visit
peacekeeper
protect
prevention
preserve
idea
wife
issue
migration
innovation
video
service
book
teacher
budget
call
introduce
agency
law
parliament
name
launch
economy
week
discuss
representative
step
group
importance
population
challenge
job
claim
number
employment
objective
reality
partly
solidarity
partner
model
solution
build
break
skill
technology
labour
enable
conference
project
contribute
learn
relate
language
household
poverty
example
barrier
welcome
enter
show
end
interview
lead
reach
recognize
respond
village
inclusion
person
month
way
organization
responsibility
husband
daughter
protection
advocate
use
amplify
household
example
poverty
barrier
launch
think
happen
raise
leave
seek
medium
ambassador
principle
entrepreneur
student
transform
sport
bias
gender
tool
change
difference
decision
position
process
disability
politics
inequality
fight
member
agree
participate
programme
policy
gap
camp
income
crisis
culture
dialogue
security
progress
worker
access
win
enter
show

Example

always check words occurrencies

occurrence
(i.e., number of times it appears in the documents)

