Università degli Studi di Padova

Social Network Analysis

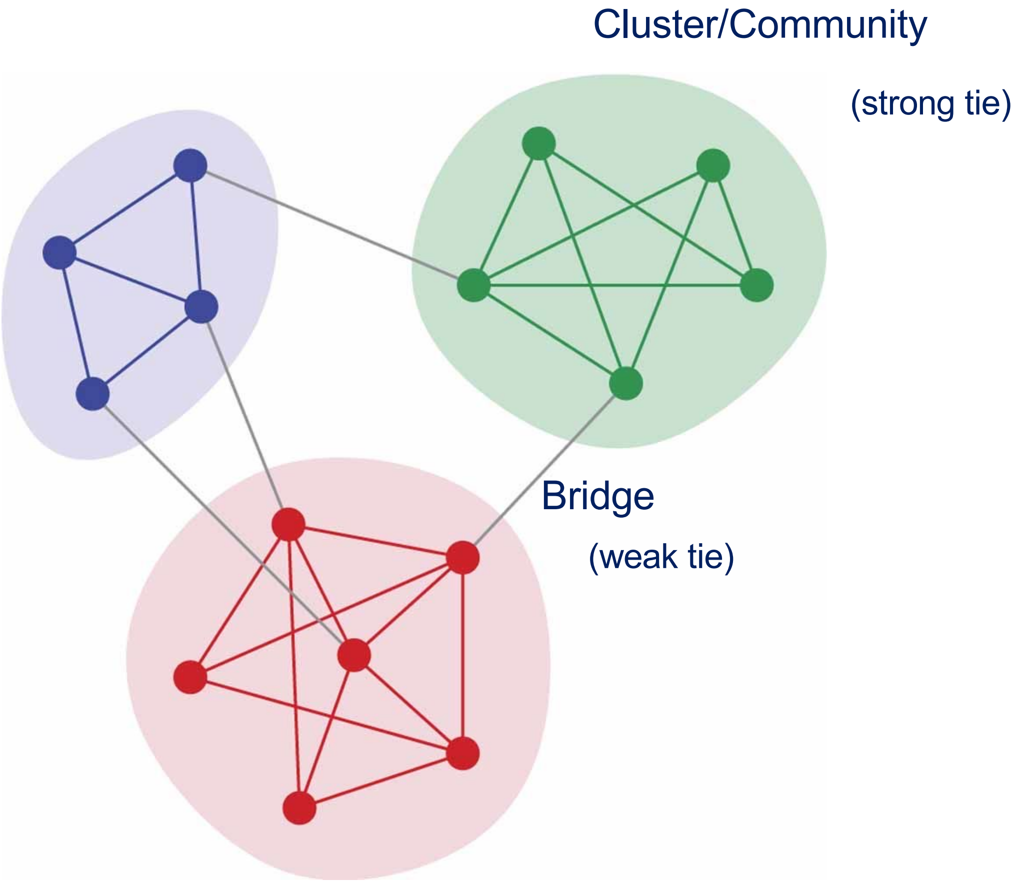
A.Y. 23/24

Communication Strategies

Community detection

a.k.a. clustering in graphs

Conceptual picture of a network
explaining the role of community detection



- We often think of networks looking like this
- But, where does this idea come from?

Granovetter's explanation

Granovetter, The strength of weak ties [1973]
https://www.jstor.org/stable/pdf/2776392.pdf

Q: How do people discovered their new jobs?

A: Through personal contacts, and mainly through acquaintances rather than through close friends

Local cluster/community
Strong ties

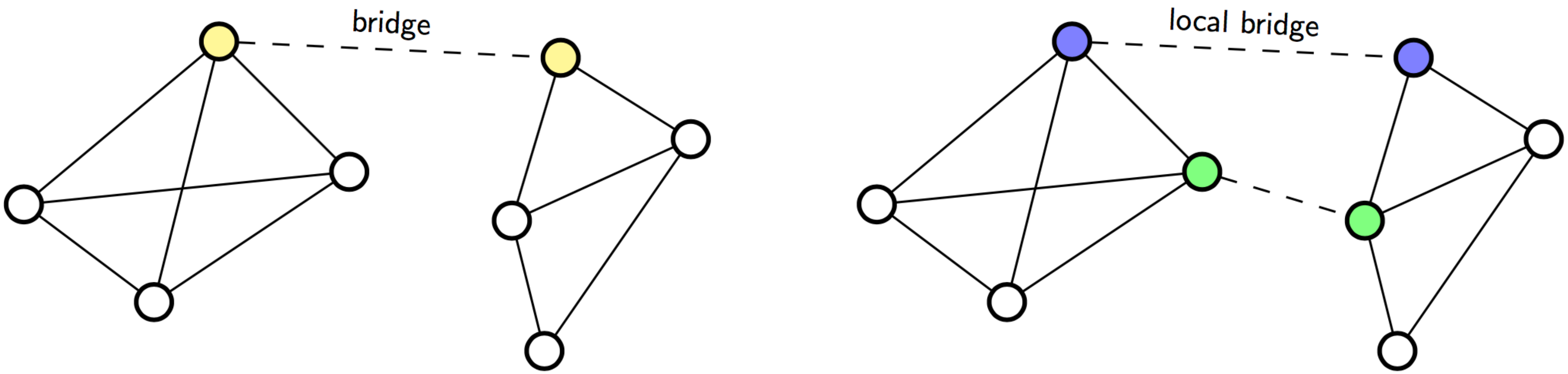
Remark: Good jobs are a scarce resource

Conclusion:

- Structurally embedded edges are also socially strong, but are heavily redundant in terms of information access
- Long-range edges spanning different parts of the network are socially weak, but allow you to gather information from different parts of the network (and get a job)

Bridges
Weak ties

Local bridges



- An edge is a bridge if deleting it *the nodes it connects* fall into different components

  this is extremely rare, e.g., because of small world properties

- An edge is a local bridge if, by deleting it, *the nodes it connects* have a span (distance) greater than 2, i.e., if *they* do not have friends in common

  common friends imply belonging to a triadic closure

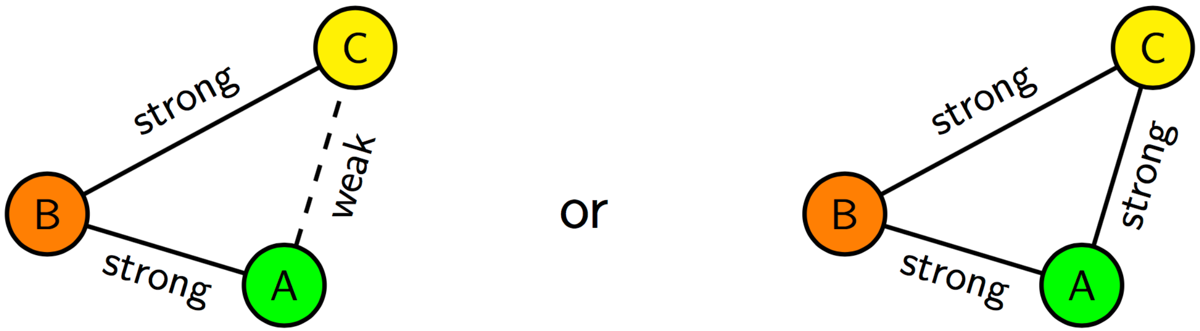
Strong triadic closure

friends/relatives and acquaintances

Assume two categories of edges:

- Strong ties (close friends)
- Weak ties (acquaintances)

Remark. If node B is strongly tied with A and C, then A and C are very likely to be connected (either weakly or strongly), that is



Strong triadic closure property – If a generic node B is strongly tied with A and C, then A and C are connected (either weakly or strongly)

Granovetter's claim

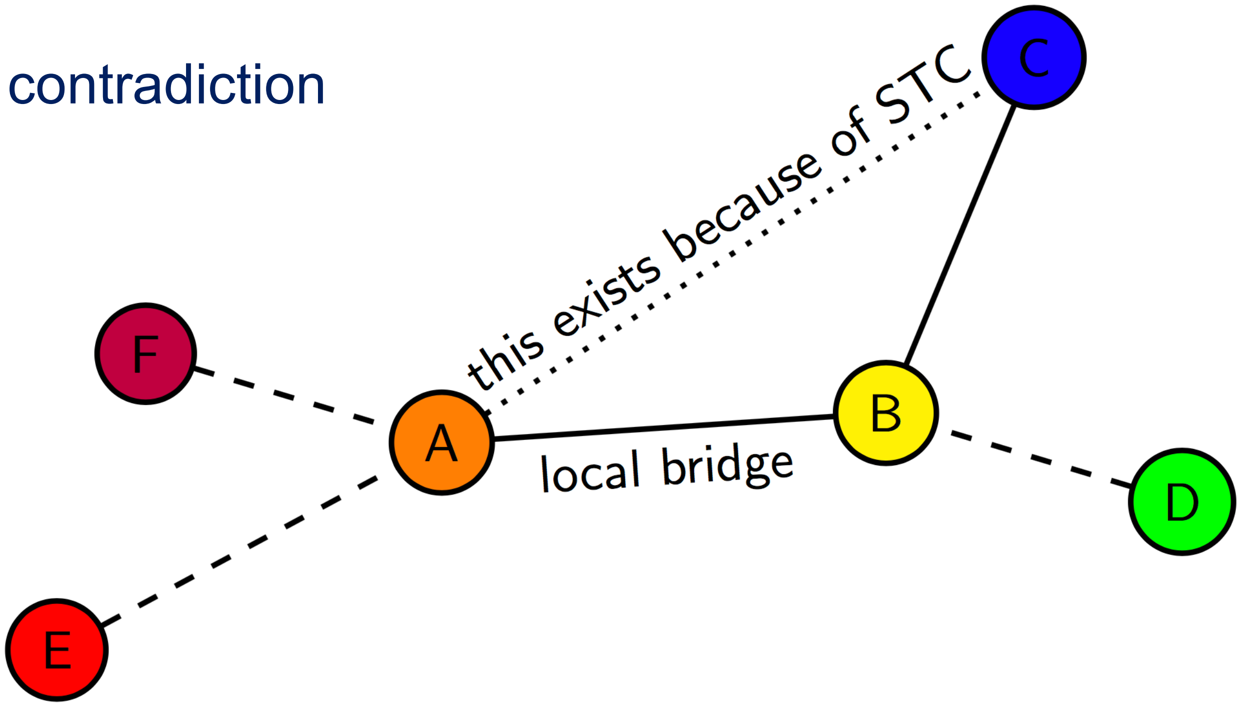
under strong triadic closure

Claim:

- Under the strong triadic closure property, local bridges are weak ties (if at least one of their nodes belongs to at least two strong ties)

Proof:

- By contradiction



Community detection

the general approach

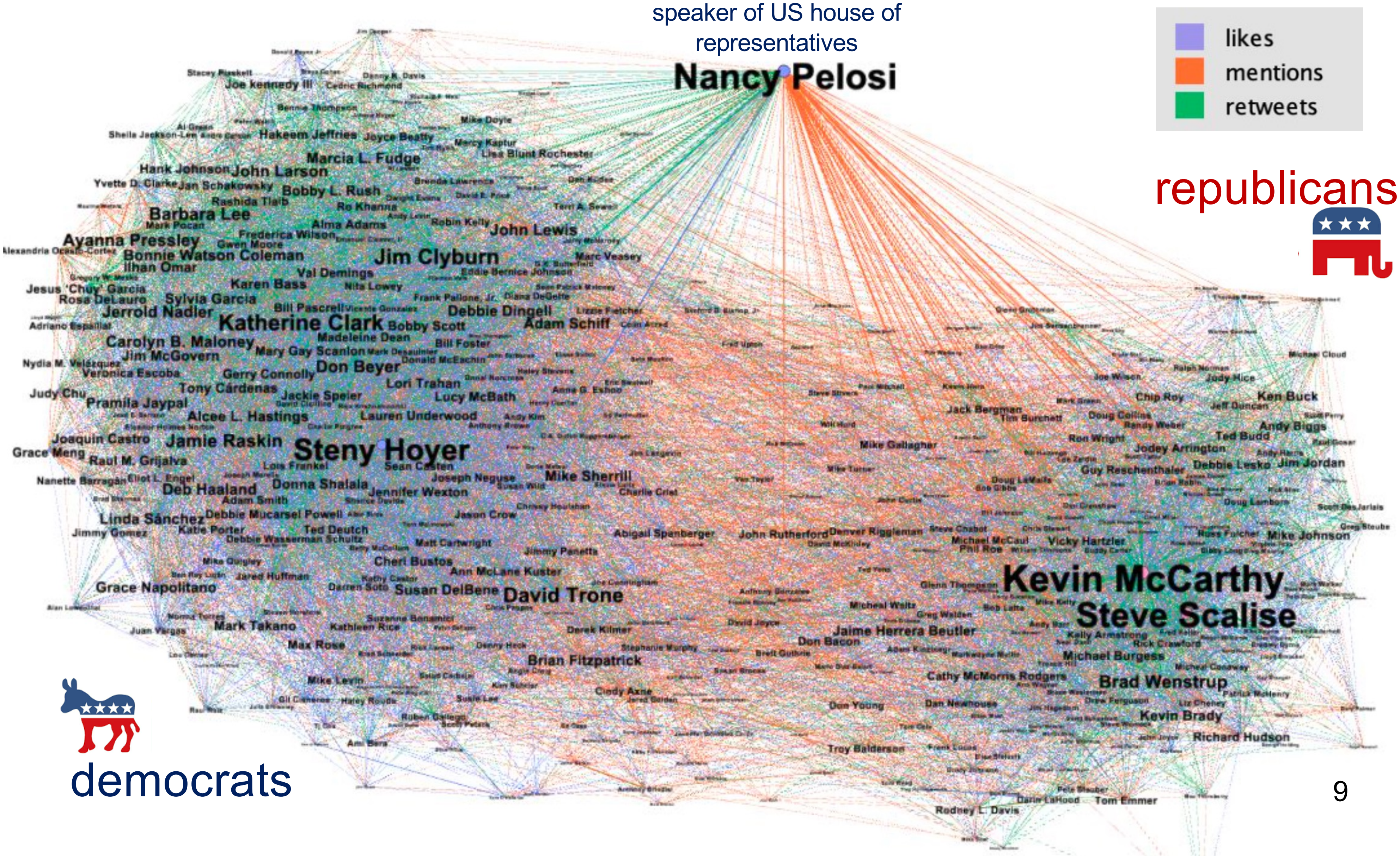
- Granovetter's theory suggests that networks are composed of tightly connected sets of nodes (i.e., communities), loosely connected between them
- We want to be able to automatically find such densely connected group of nodes
- We look for unsupervised methods, as most of the times no ground truth is available
- We look for a measure of the goodness of a community assignment, to be able to compare the performance of different algorithms
- Applications in:
  - social networks
  - functional brain networks in neuroscience
  - scientific interactions

Figure 2 | A network of collaborations among scientists at a research

Università degli Studi di Padova

Clustering political beliefs

US republican and democrats interactions on Twitter 2020



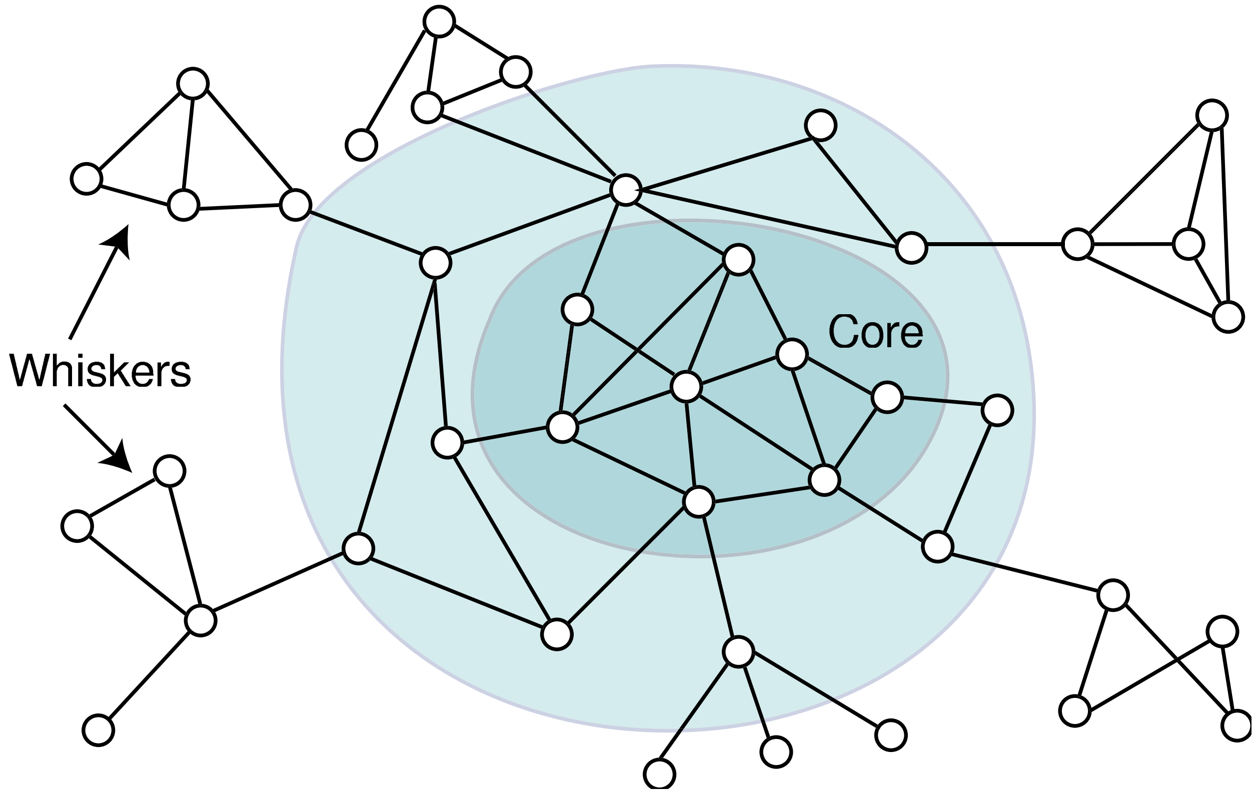
The core periphery model

Lescovec, Lang, Dasgupta, Mahoney, Community Structure in Large Networks: Natural Cluster Sizes and the Absence of Large Well-Defined Clusters (2008)

https://arxiv.org/abs/0810.1355

Can we find a justification for this?

Small, peripheral clusters

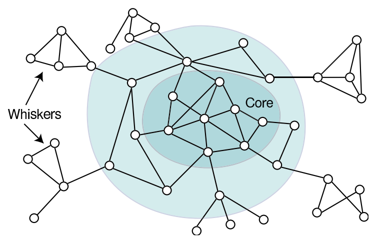


Caricature of network structure

UNIVERSITÀ
DEGLI STUDI
DI PADOVA

Overlapping communities

to explain the core periphery model

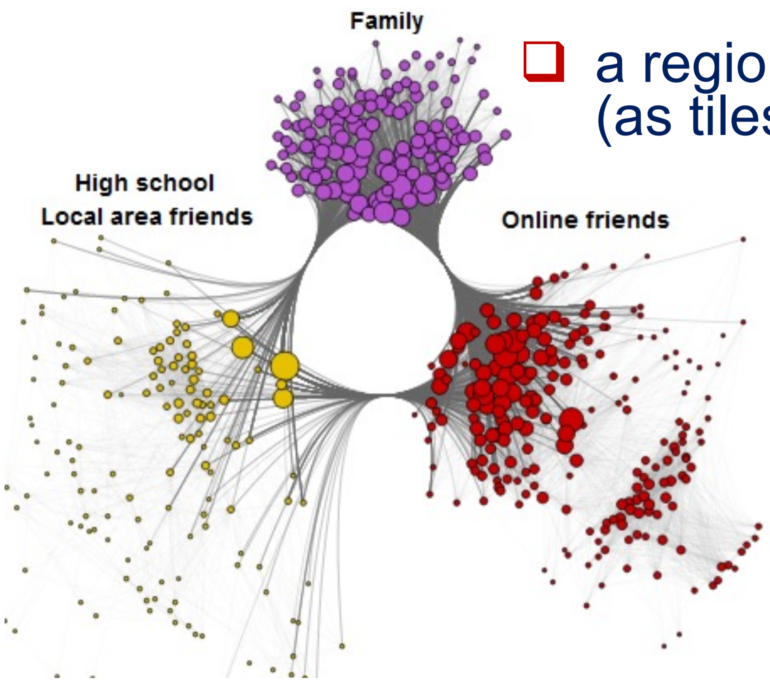


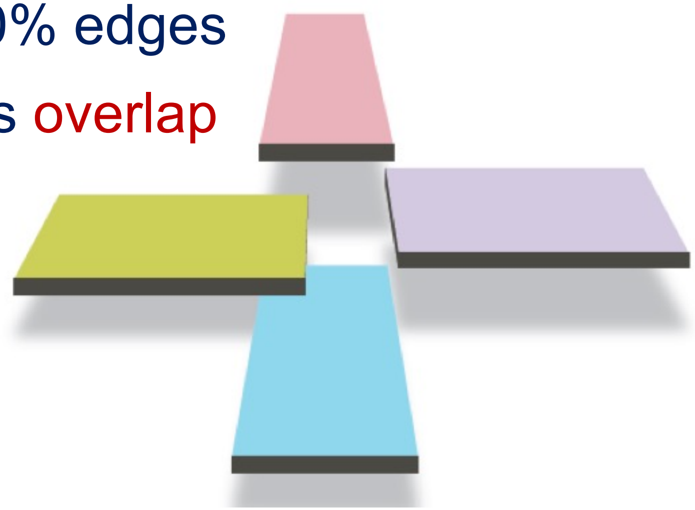
Wiskers

- are typically of size 100
- are responsible of good communities

Core

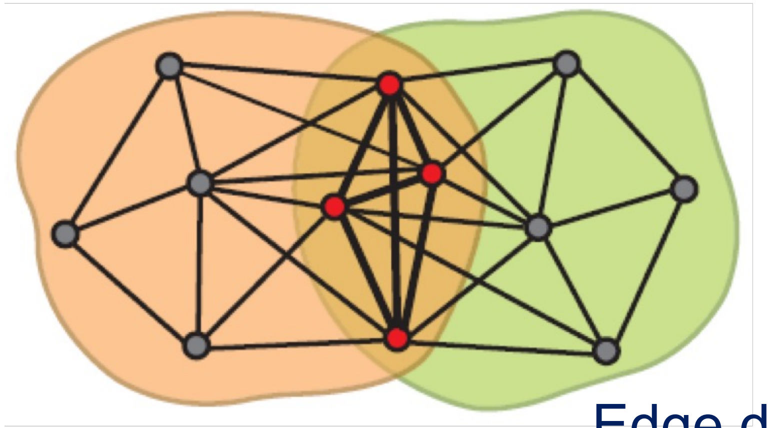
- denser and denser region
- contains 60% nodes and 80% edges
- a region where communities overlap (as tiles)



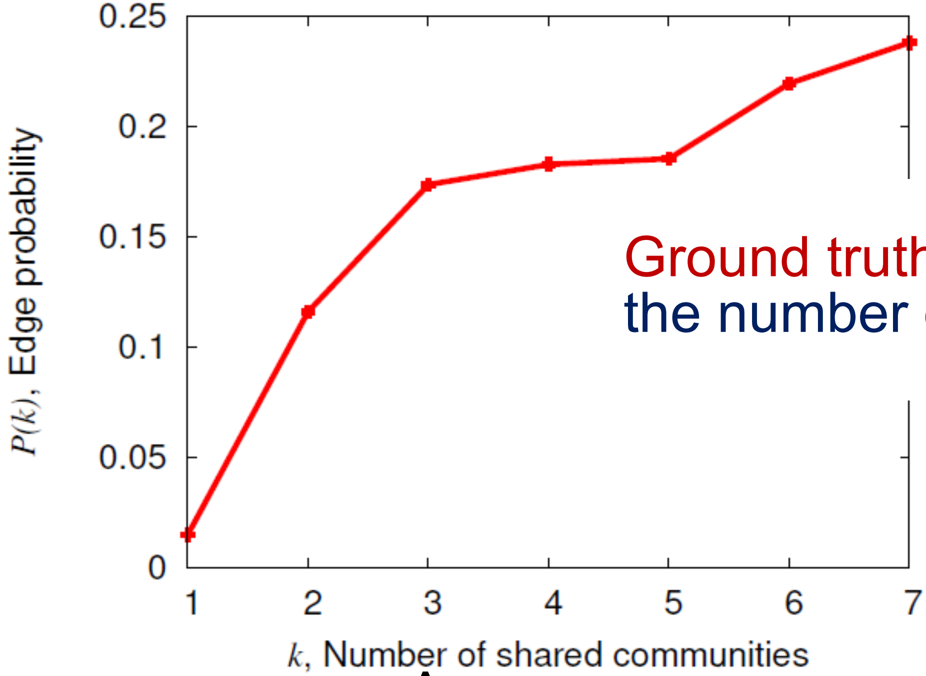


Measuring overlapping

in social networks



Edge density is bigger in the overlap



Amazon

Ground truth - Edge probability increases with the number of shared communities

Feld, The focused organization of social ties, [1981]
The more different communities that two individuals share, the more likely is that they will be tied

Clustering algorithms

i.e. community detection algorithms

Modularity

Newman, Modularity and community structure in networks (2006)
https://www.pnas.org/content/pnas/103/23/8577.full.pdf

Want to:

- measure of how well a network is partitioned into communities (i.e., sets of tightly connected nodes)

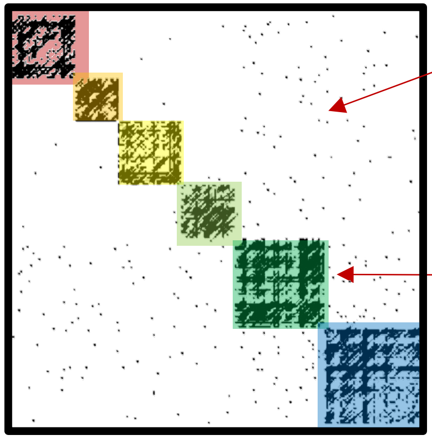
Idea:

- “If the number of edges between two groups is only what one would expect on the basis of random chance, then few thoughtful observers would claim this constitutes evidence of meaningful community structure”
- Modularity is “the number of edges falling within groups minus the expected number in an equivalent network with edges placed at random”
- The higher modularity, the better the community assignment

Number of edges falling within groups

an adjacency matrix overview

Original graph

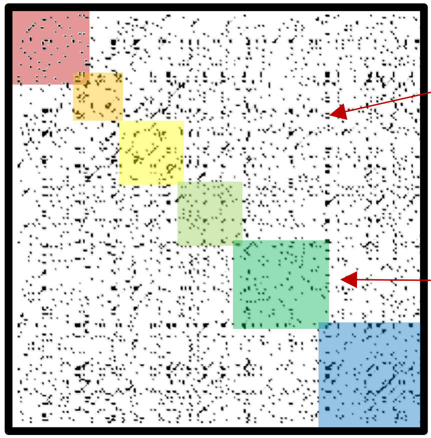


few edges linking the communities

good community assignment = many edges inside blocks

The sum **Q1** of active connections inside the boxes (inside the communities) of the original graph is **high**!

Randomly rewired graph



edges equally distributed everywhere

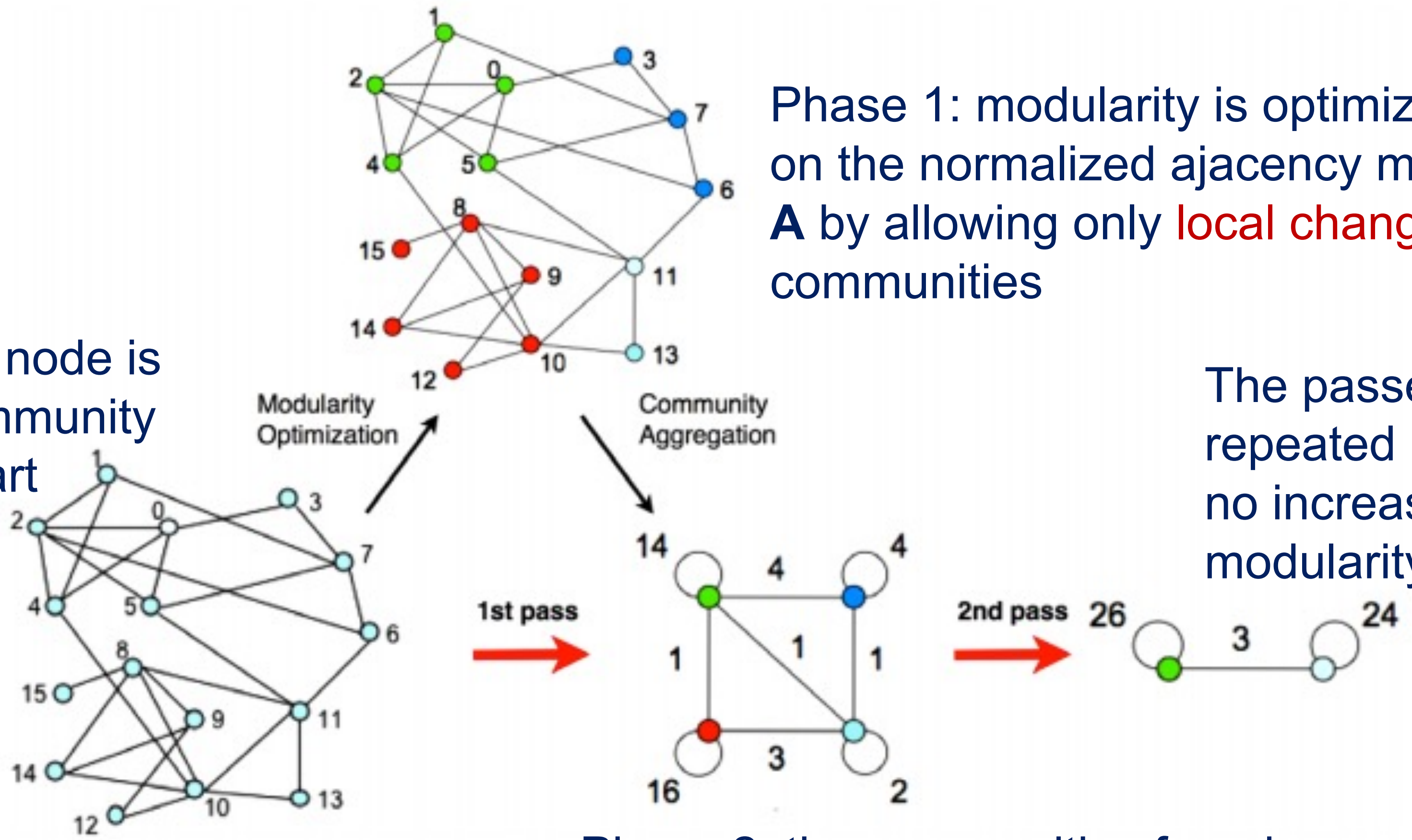
much fewer edges inside the communities!

The sum **Q2** of active connections inside the boxes (inside the communities) of the rewired graph is **low**!

Modularity: **Q = Q1 – Q2**
The higher Q the better the community assignment!

The Louvain algorithm

Blondel, Guillaume, Lambiotte, Lefebvre, Fast unfolding of communities in large networks (2008)
https://arxiv.org/abs/0803.0476



Phase 1: modularity is optimized on the normalized ajacency matrix **A** by allowing only local changes of communities

Each node is a community @ start

The passes are repeated iteratively until no increase of modularity is possible

Phase 2: the communities found are aggregated (sum of links) in order to build a new network of communities with normalized adjacency matrix $\boldsymbol{P}_{CC}$

Characteristics of Louvain

- Implements modularity optimization
- Scalable (low complexity)
- Effective
- Available as the reference implementation in any programming language
- A greedy technique (in the order the nodes are searched)

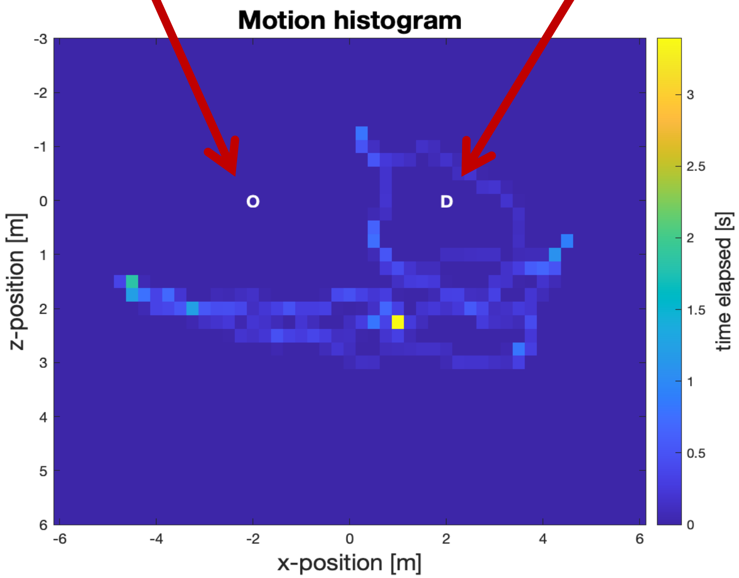
Università degli Studi di Padova

Clustering motion patterns

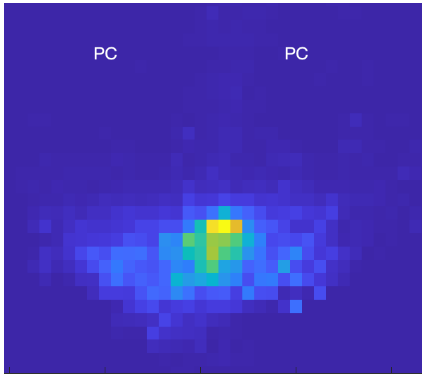
In immersive environments

original pointcloud

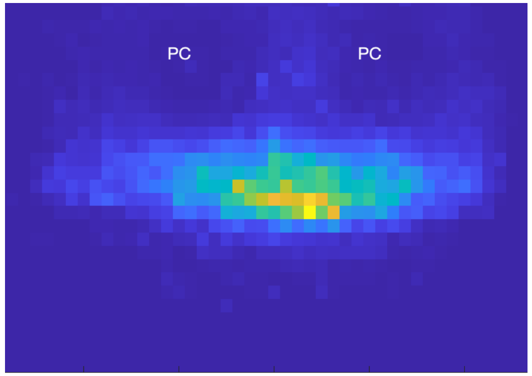
distorted pointcloud

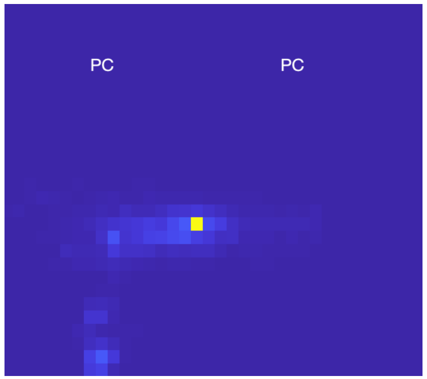


Cluster 1: walking from a distance



Cluster 2: walking closely





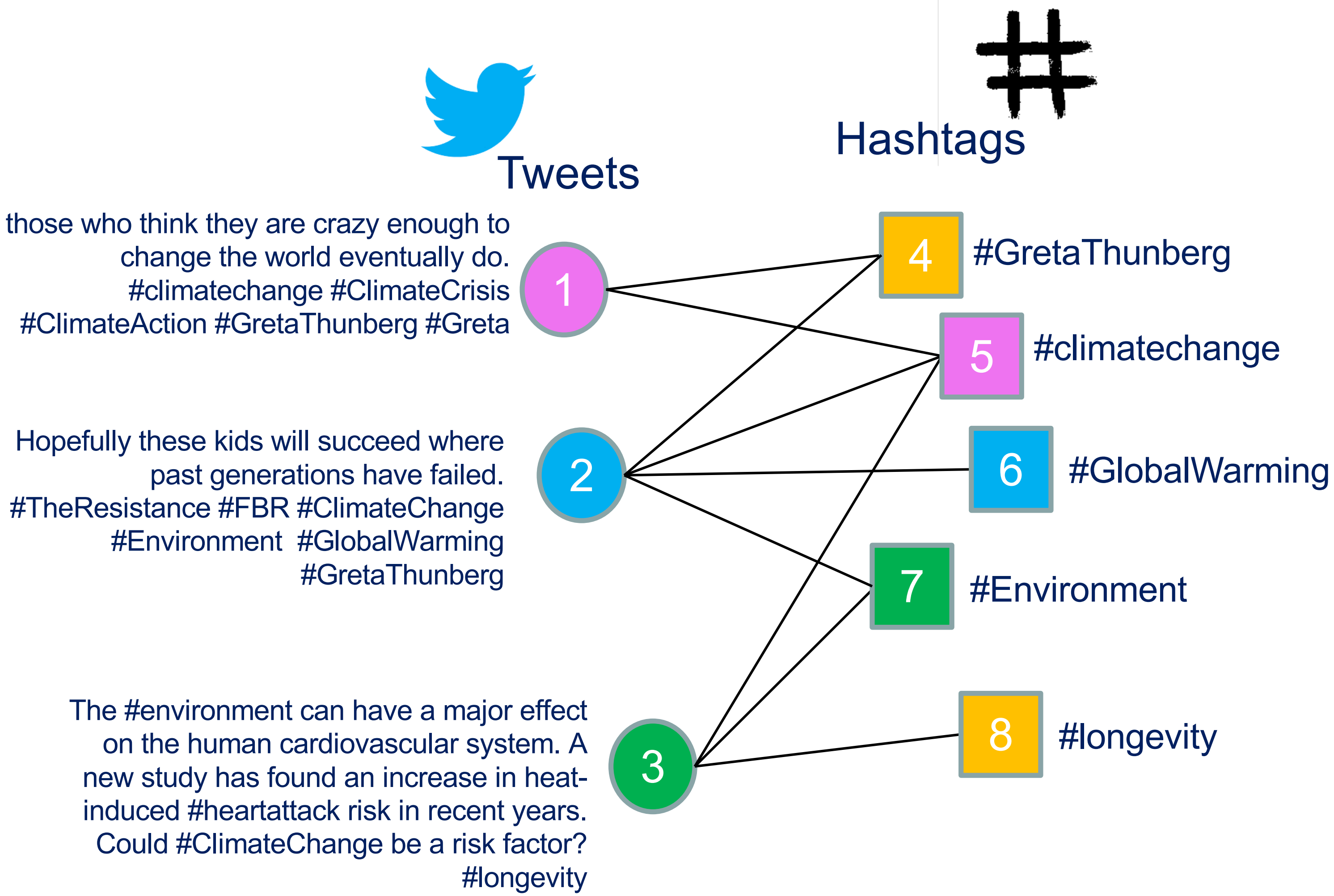
Cluster 3: standing still

motion behaviours detected by Louvain on Pearson correlations over (filtered) motion patterns

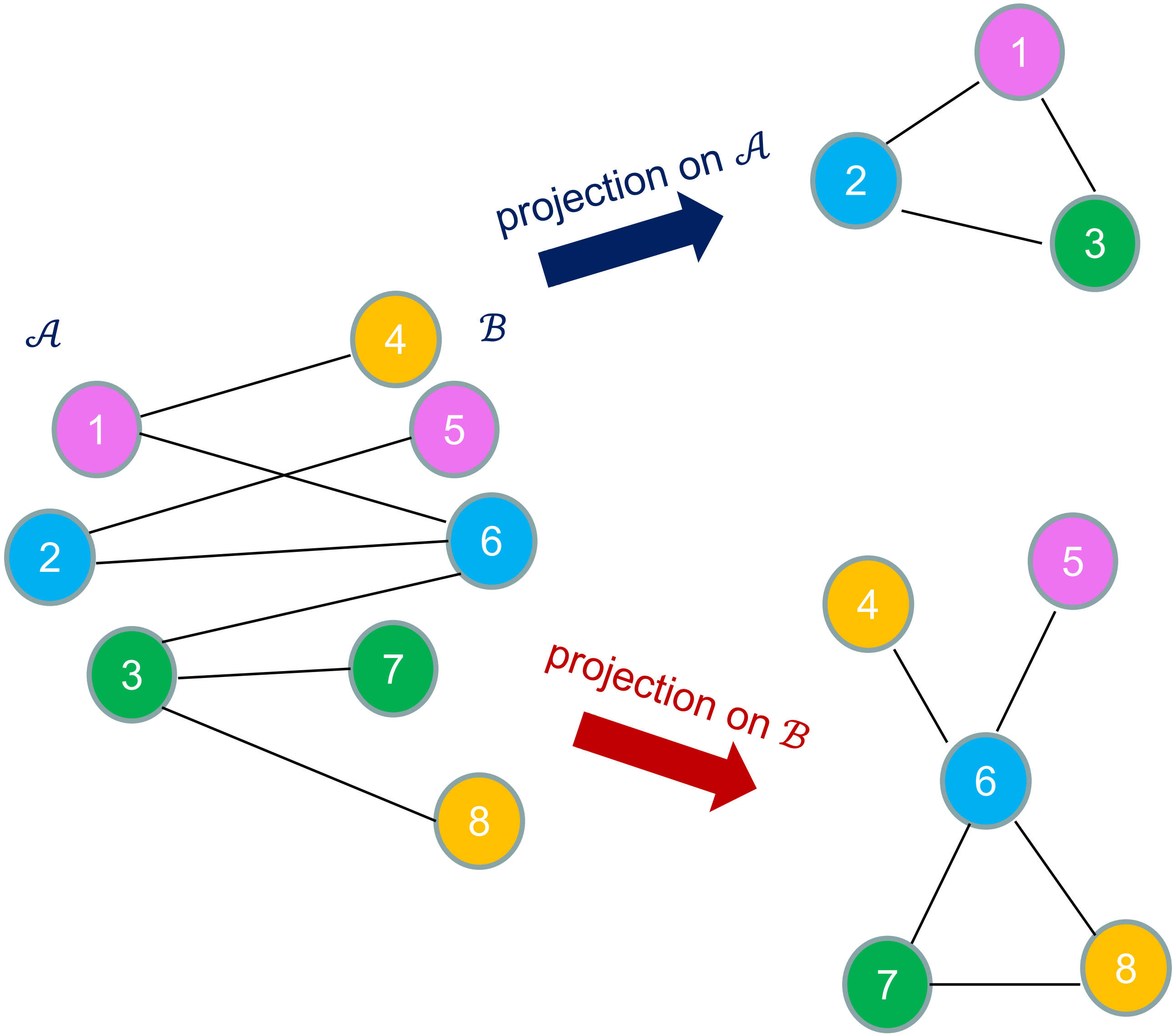
Topic Detection

i.e. community detection in semantic networks

Bipartite graph example



Projections



Nodes are linked if they have a common neighbour in $\mathcal{B}$

PS: we say that nodes *i* and *j* have a common neighbour *k* if both *i* and *j* are connected to *k*

Nodes are linked if they have a common neighbour in $\mathcal{A}$

UNIVERSITÀ DEGLI STUDI DI PADOVA

Word clustering with Louvain

community detection applied to semantic networks = topic detection

#metoo tweets

Università degli Studi di Padova

Document clustering with BERTopic

clustering documents into different topics

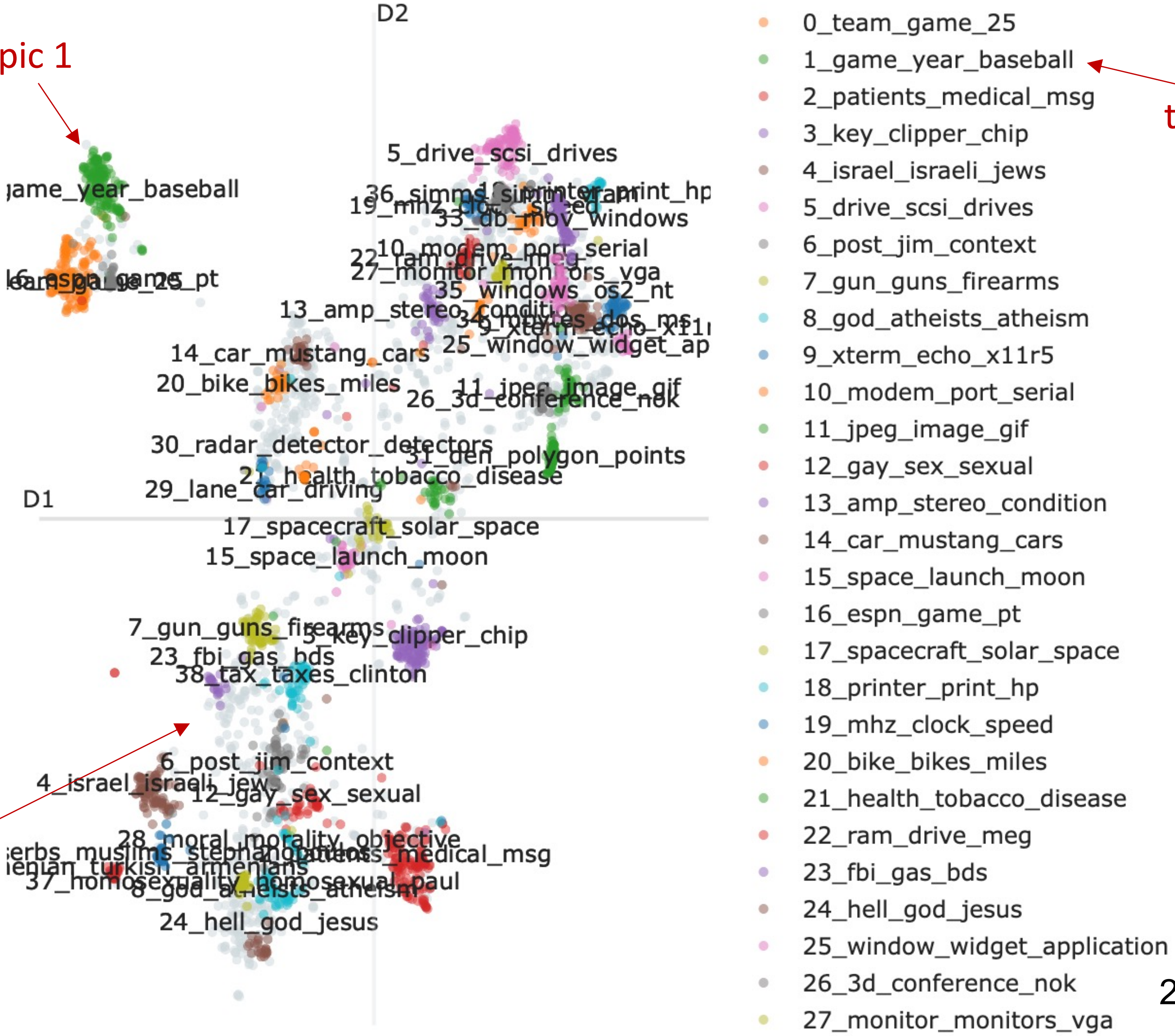
1. each document is mapped into an embedding (vector) by BERT

2. cosine metric is used to identify distances among documents

3. UMAP maps into a simpler space (e.g. 2 dimensional)

4. HDBSCAN is run to identify topics

topic 1



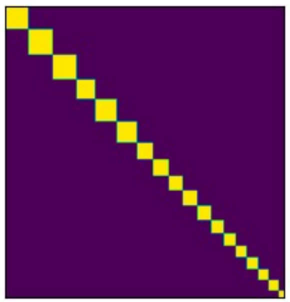
topic 1

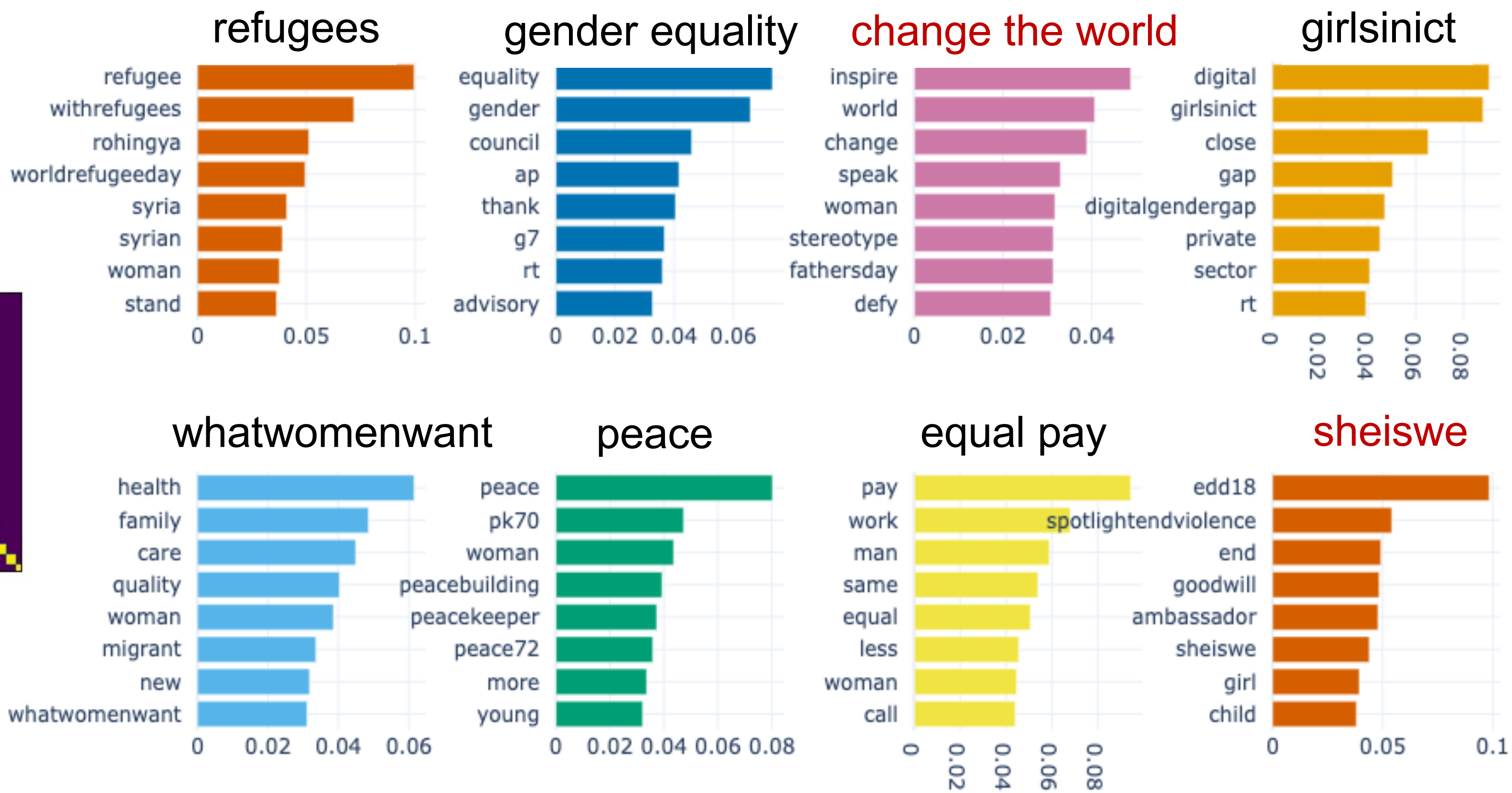
outliers in gray

UNIVERSITÀ
DEGLI STUDI
DI PADOVA

Topic description

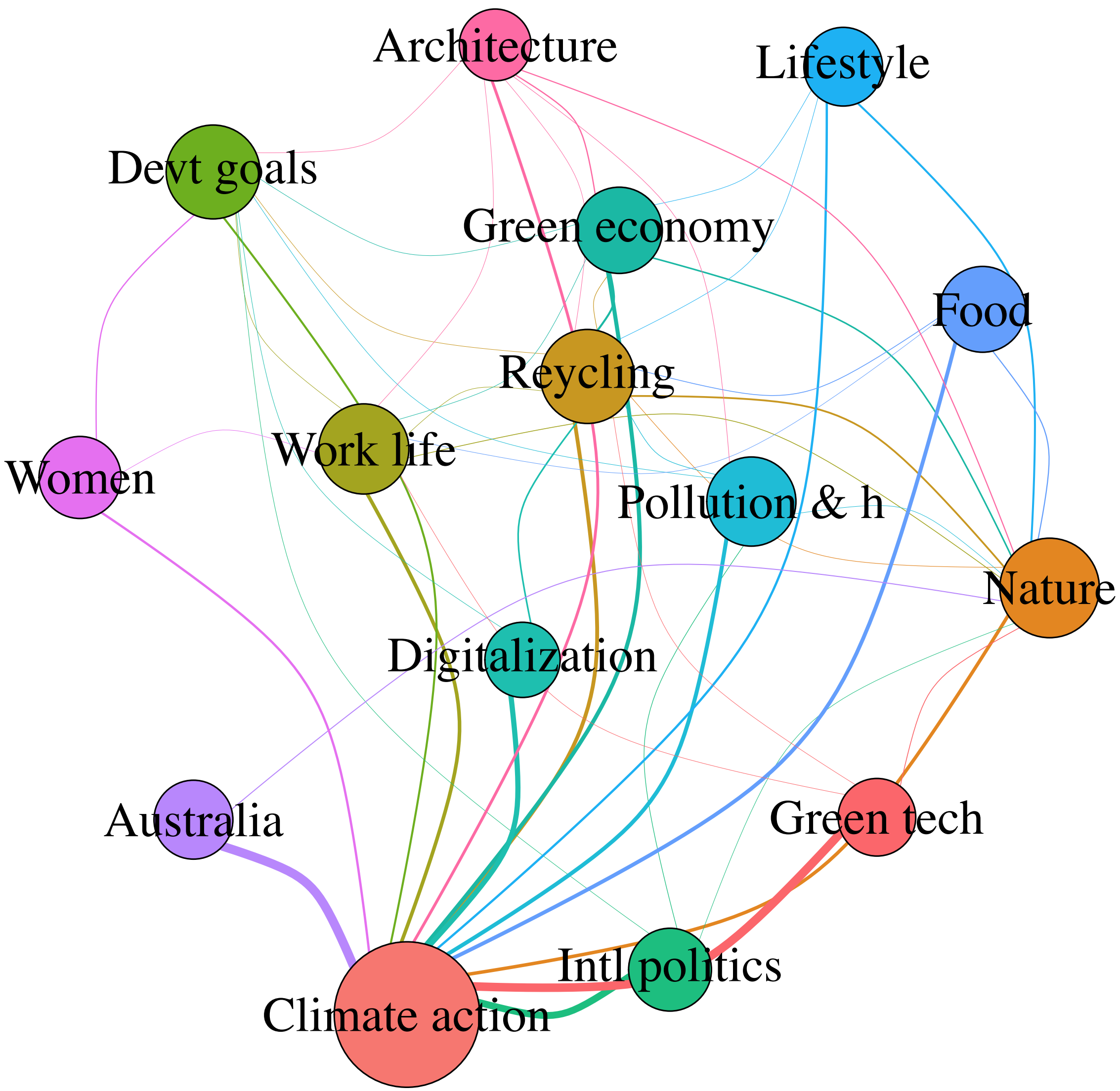
In BERTopic, #metoo2018





Topics network

More informative than plotting documents as nodes



projecting the adjacency matrix on topics

P_{11}	P_{12}	P_{13}
P_{21}	P_{22}	P_{23}
P_{31}	P_{32}	P_{33}

Extracting sentiment

And related ideas

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 ☺

Università degli Studi di Padova

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!

LIWC (in Python)
at work

	translated	i	we	body	social	past	future
0	I MISSED YOU GUYSS	25.000000	0.0	0.000000	50.000000	25.0	0.000000
1	we love you so much namjoon🥰💜💜💜	0.000000	10.0	0.000000	30.000000	0.0	0.000000
2	haters: 🙄😡😤🦣🎃\nBP: 💗👄🎀💸💋	0.000000	0.0	0.000000	0.000000	0.0	0.000000
3	BLACKPINK BEST GROUP IN THE WORLD💯🔥🔥	0.000000	0.0	0.000000	11.111111	0.0	0.000000
4	I love Changbin's humor 😭😭❤️	12.500000	0.0	0.000000	12.500000	0.0	0.000000
...	...	...	...	...	...	...	...
153	My legs are so big	20.000000	0.0	20.000000	0.000000	0.0	0.000000
154	ne ne	0.000000	0.0	0.000000	0.000000	0.0	0.000000
155	I think not being able to see it live will be ...	9.523810	0.0	0.000000	0.000000	0.0	4.761905
156	MY BABIES	50.000000	0.0	0.000000	50.000000	0.0	0.000000
157	I learned it from the foot steps.	14.285714	0.0	14.285714	0.000000	0.0	0.000000

Agency/Communion (in Python)

at work

	translated	Dic	WC	BigWords	Numbers	AllPunct	Period	Comma	QMark	Exclam	Apostro	Agency	Communion
0	I MISSED YOU GUYSS	25.000000	4.0	0.000000	0.0	0.000000	0.000000	0.0	0.0	0.0	0.000	25.000000	0.000000
1	we love you so much namjoon🥰💜💜💜	30.000000	10.0	10.000000	0.0	0.000000	0.000000	0.0	0.0	0.0	0.000	10.000000	20.000000
2	haters: 🙄😡😤🦣🎃\nBP: 💗🫦🎀💸💋	0.000000	12.0	0.000000	0.0	0.166667	0.000000	0.0	0.0	0.0	0.000	0.000000	0.000000
3	BLACKPINK BEST GROUP IN THE WORLD💯🔥🔥	11.111111	9.0	11.111111	0.0	0.000000	0.000000	0.0	0.0	0.0	0.000	0.000000	11.111111
4	I love Changbin's humor 😭😭❤️	12.500000	8.0	12.500000	0.0	0.125000	0.000000	0.0	0.0	0.0	0.125	0.000000	12.500000
...	...	...	...	...	...	...	...	...	...	...	...	...	...

Socio-psychological linguistic markers on #climatechange

ingroup community orientation within the text

salience of group membership, sense of belonging

a person's striving to be independent to assert, protect and expand one's self

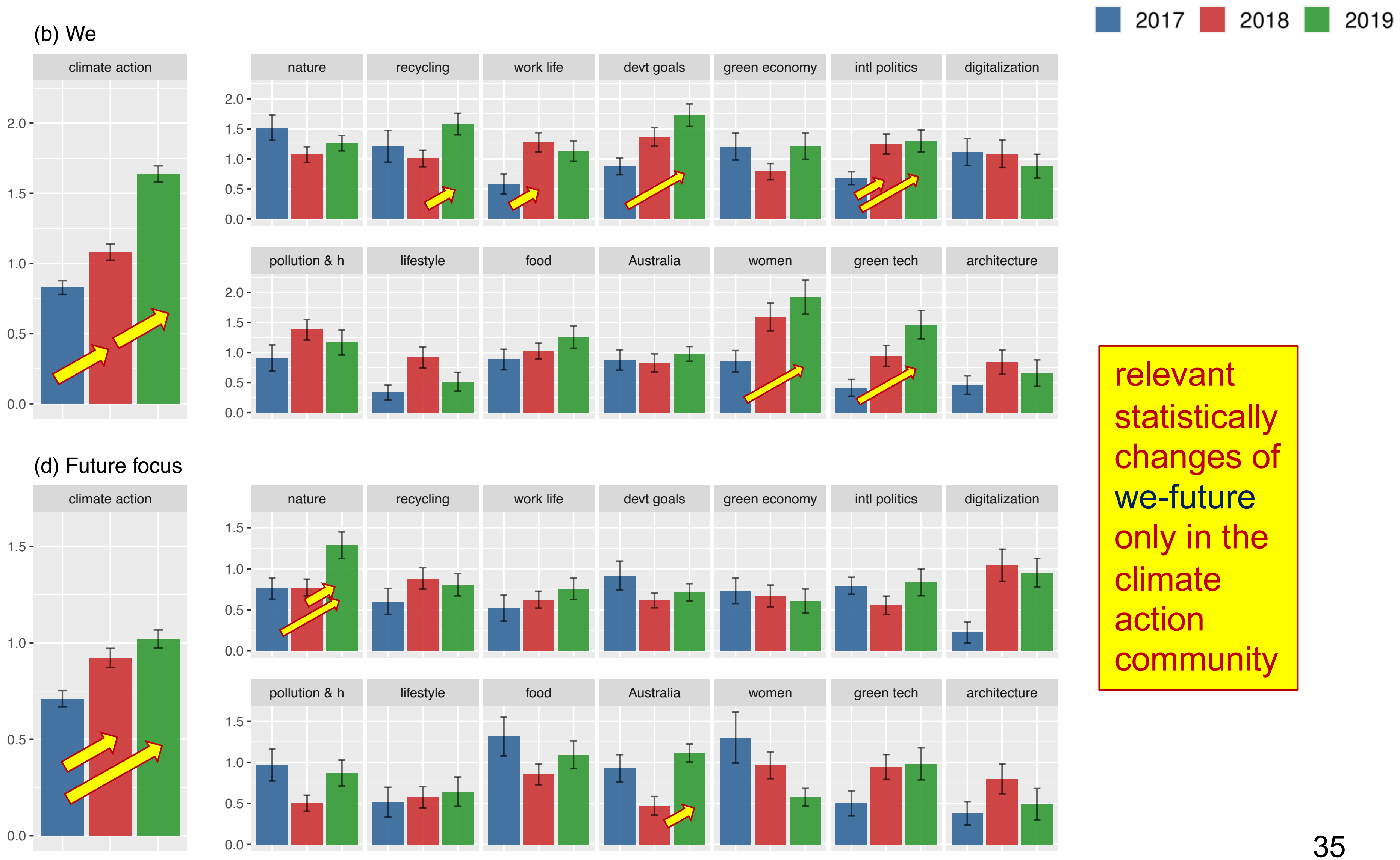
orientation of tweets to the past or future

only a few statistically relevant changes

Università degli Studi di Padova

Socio-psychological linguistic markers

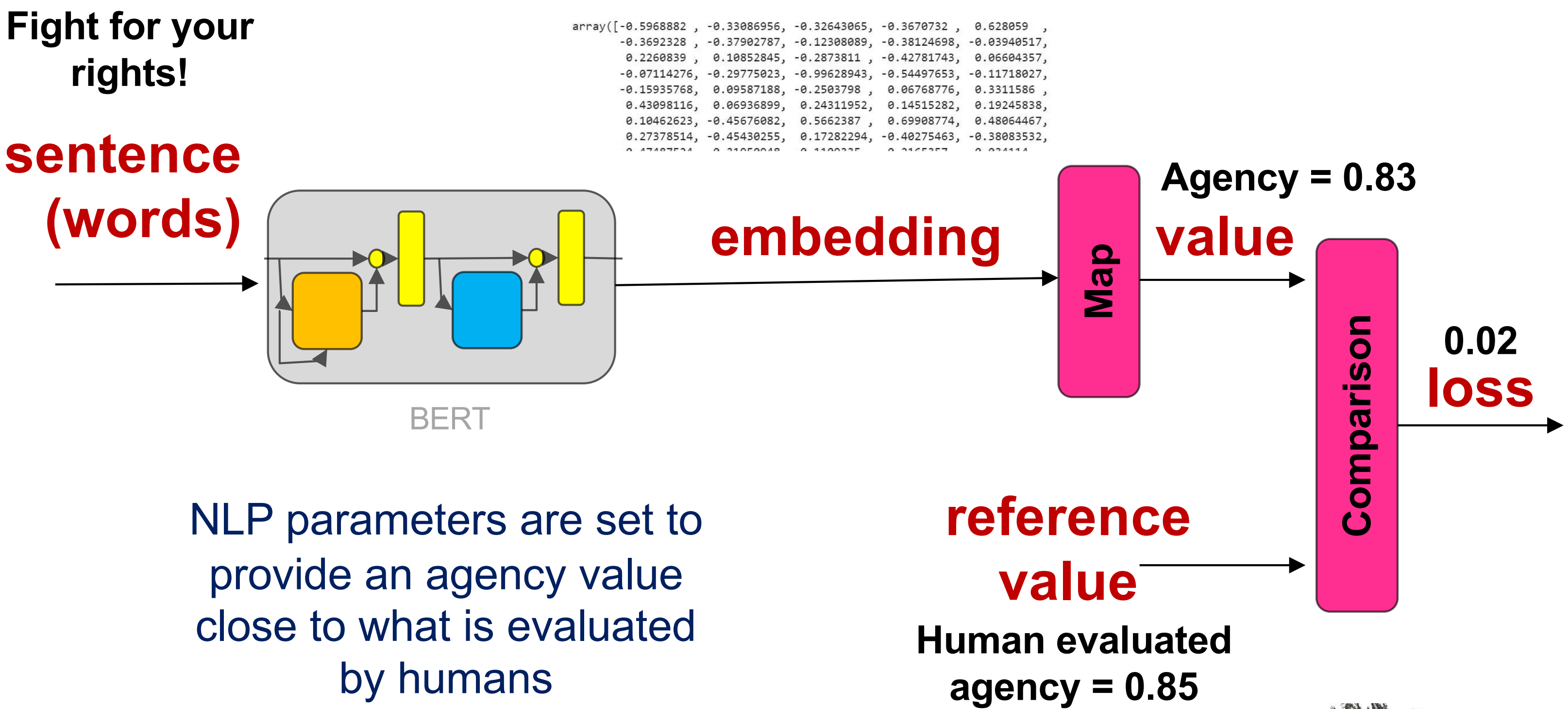
a view inside topics



relevant statistically changes of we-future only in the climate action community

Using AI to predict socio-psychological markers

BERT
Training a NLP tool

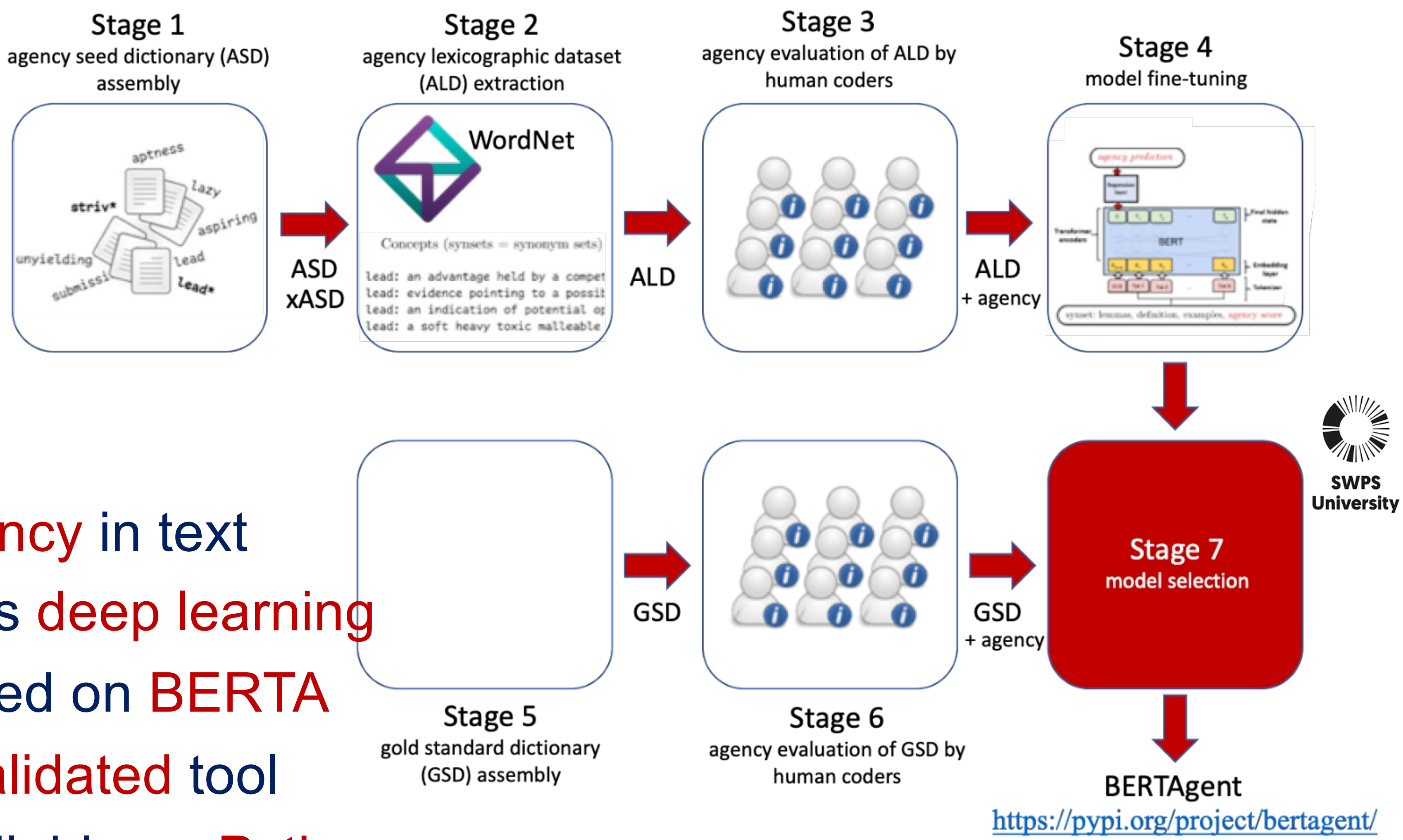


NLP parameters are set to provide an agency value close to what is evaluated by humans

A large and dedicated dataset is needed!

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

UNIVERSITÀ DEGLI STUDI DI PADOVA

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

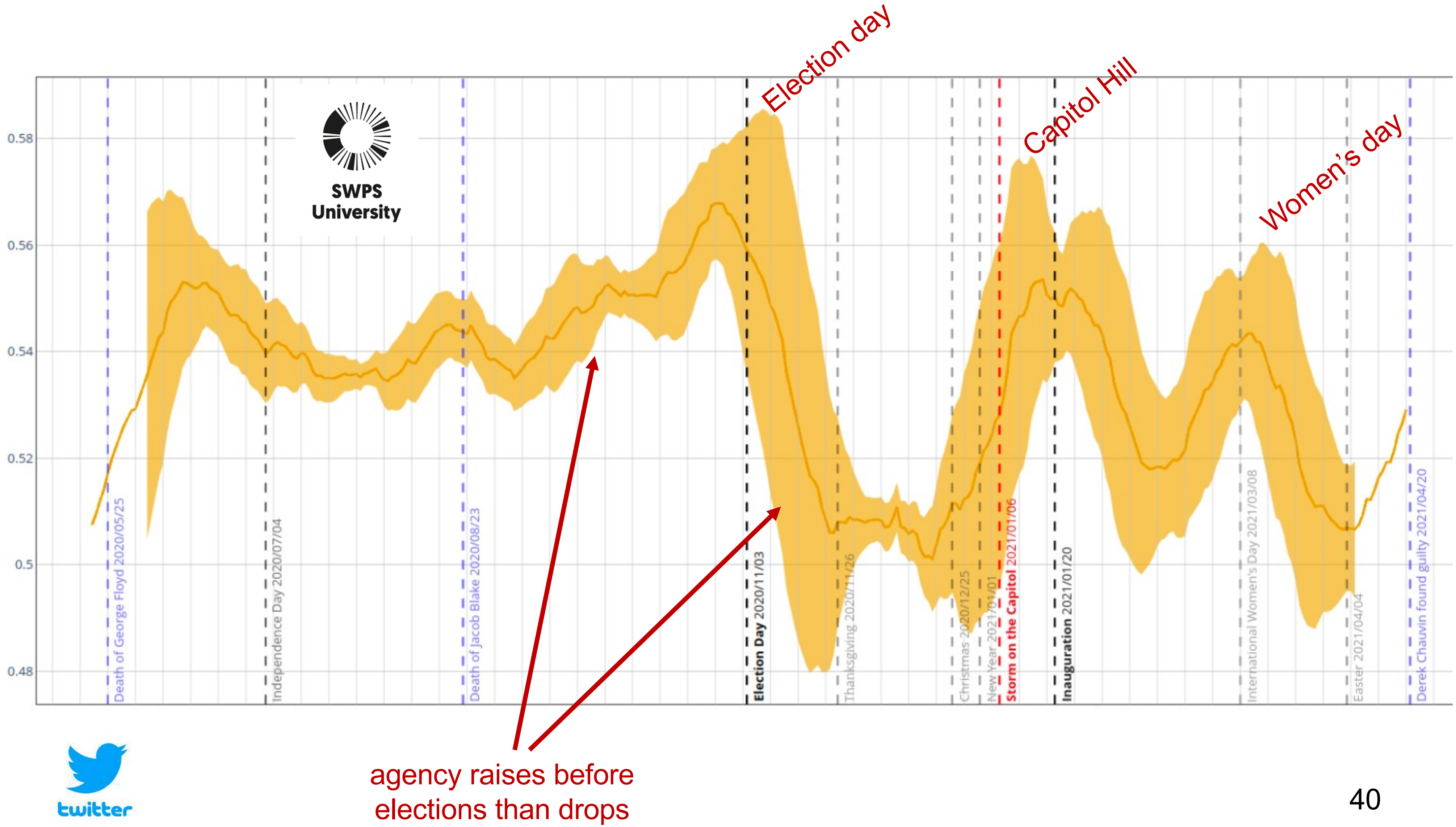
BERTAgent

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



BERTAgent
at work

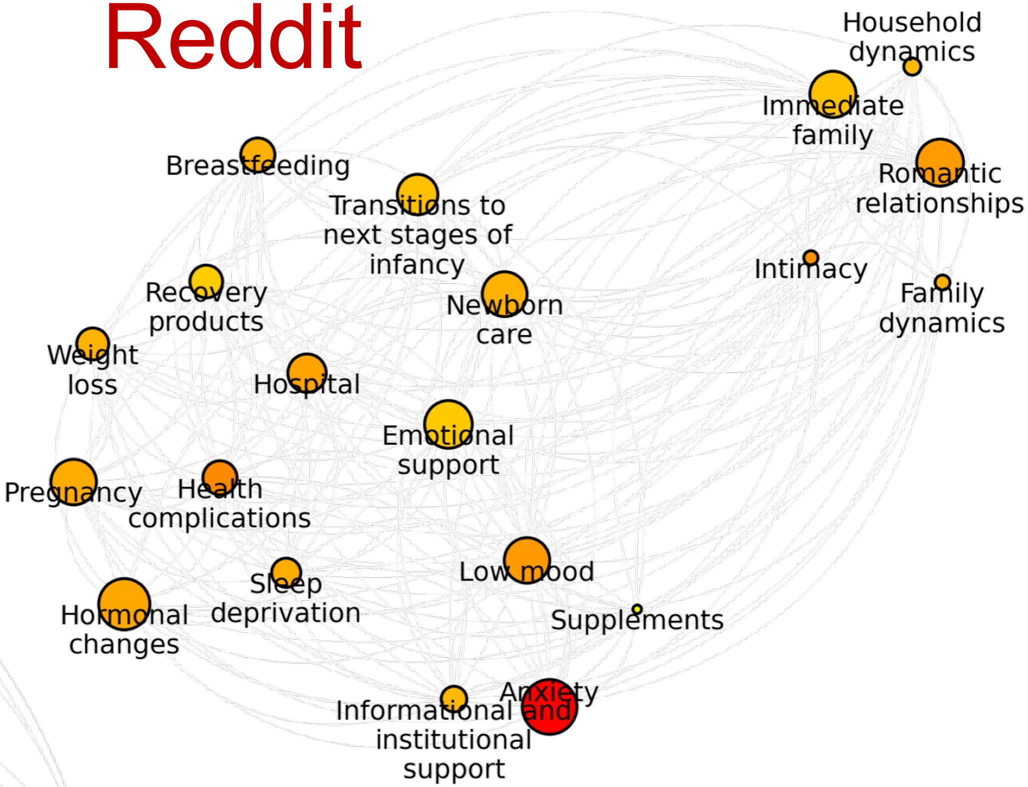
	translated	agency
0	I MISSED YOU GUYSS	-0.331446
1	we love you so much namjoon🥰💜💜💜	0.009748
2	haters: 🙄😡😤🦣🎃\nBP: 💗👄🎀💸💋	-0.261824
3	BLACKPINK BEST GROUP IN THE WORLD💯🔥🔥	0.058124
4	I love Changbin's humor 😭😭❤️	0.077732
...	...	...
153	My legs are so big	-0.065885
154	ne ne	-0.124783
155	I think not being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

UNIVERSITÀ DEGLI STUDI DI PADOVA

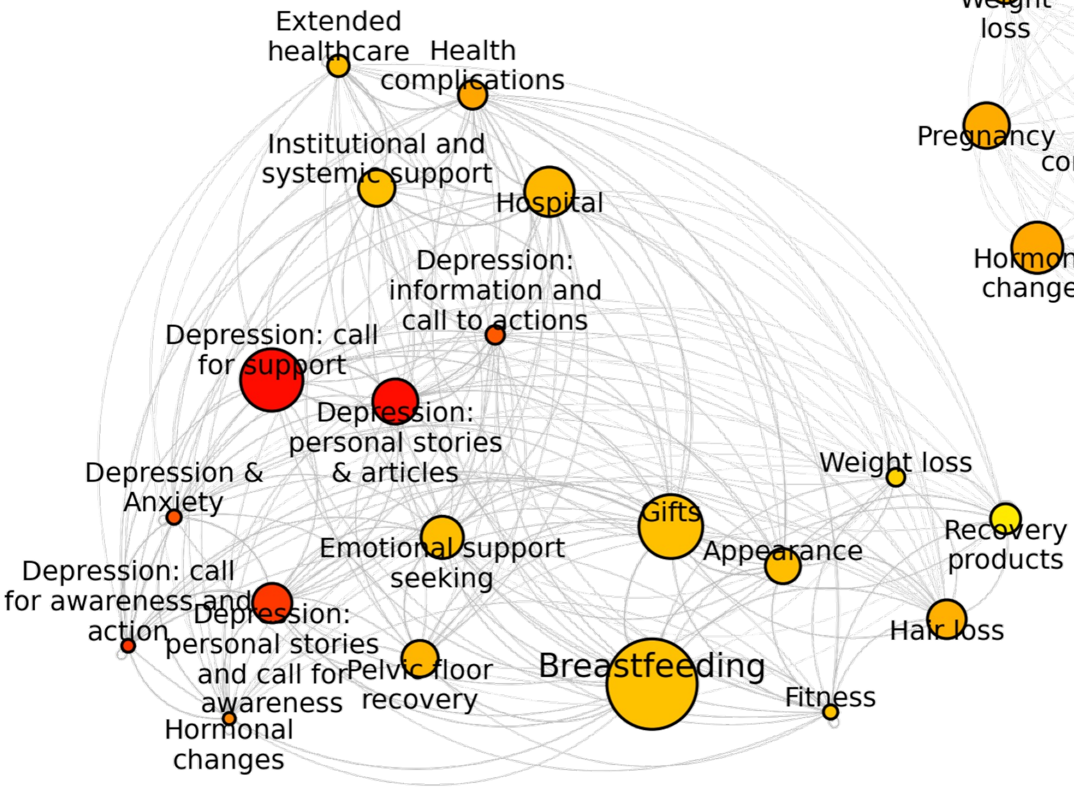
Agency in postpartum depression

Topics view

Reddit



Twitter



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

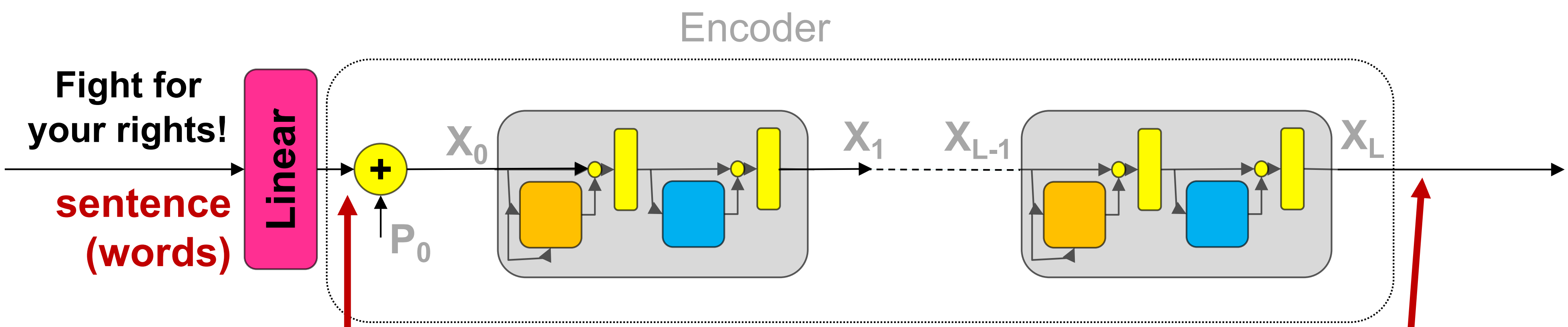
How AI machines work

An intuitive overview

Transformer Architecture

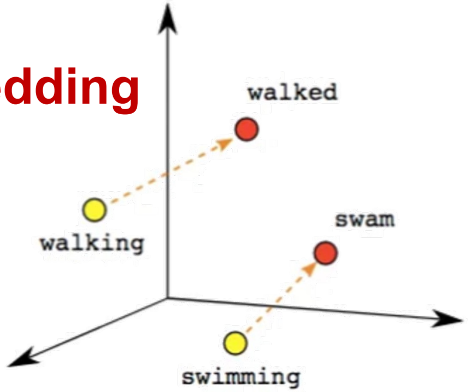
Vaswani, Ashish, et al. "Attention is all you need" (2017)

Google's patent https://patents.google.com/patent/US10452978B2/en



input embedding

each word is associated to a list of numbers = embedding (many, like 756 or 1024)



output embedding

each sentence is mapped to an embedding, words embeddings now capture the context

```
array([-0.5968882 , -0.33086956, -0.32643065, -0.3670732 ,  0.628059  ,
       -0.3692328 , -0.37902787, -0.12308089, -0.38124698, -0.03940517,
        0.2260839 ,  0.10852845, -0.2873811 , -0.42781743,  0.06604357,
       -0.07114276, -0.29775023, -0.99628943, -0.54497653, -0.11718027,
       -0.15935768,  0.09587188, -0.2503798 ,  0.06768776,  0.3311586 ,
        0.43098116,  0.06936899,  0.24311952,  0.14515282,  0.19245838,
        0.10462623, -0.45676082,  0.5662387 ,  0.69908774,  0.48064467,
        0.27378514, -0.45430255,  0.17282294, -0.40275463, -0.38083532,
```

Attention (machine learning)

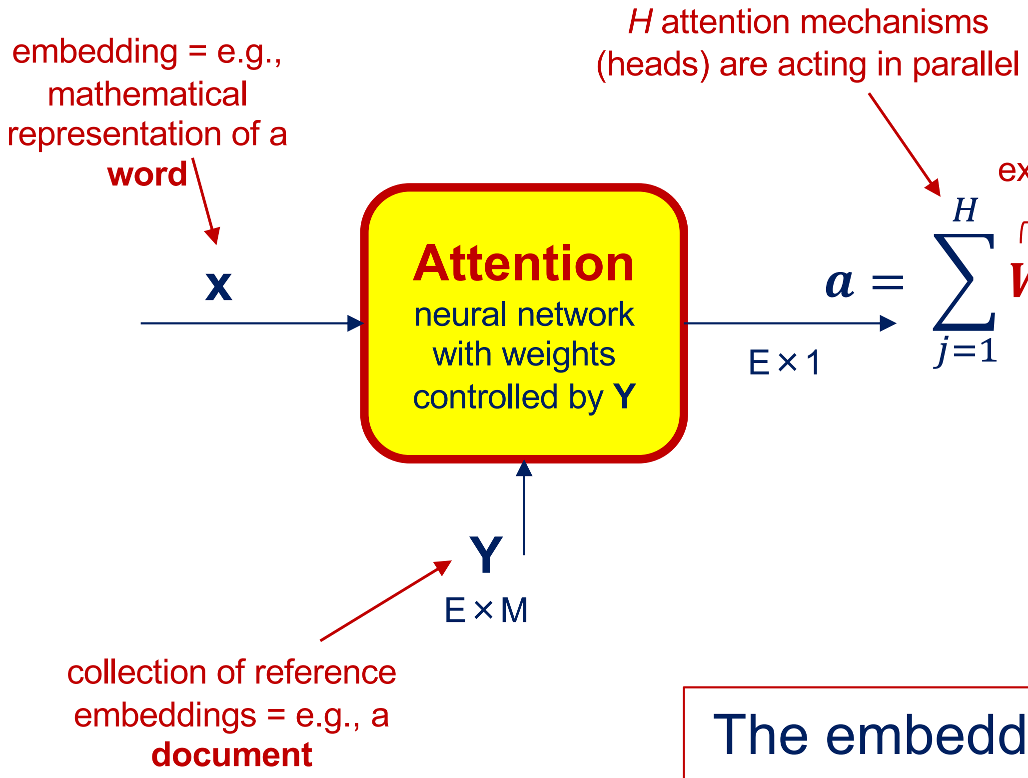
Article Talk

From Wikipedia, the free encyclopedia

In artificial neural networks, **attention** is a technique that is meant to mimic cognitive attention. This effect enhances some parts of the input data while diminishing other parts — the motivation being that the network should devote more focus to the important parts of the data, even though they may be small portion of an image or sentence. Learning which part of the data is more important than another depends on the context, and this is trained by gradient descent.

The Attention Module

Vaswani, Ashish, et al. "Attention is all you need" (2017)



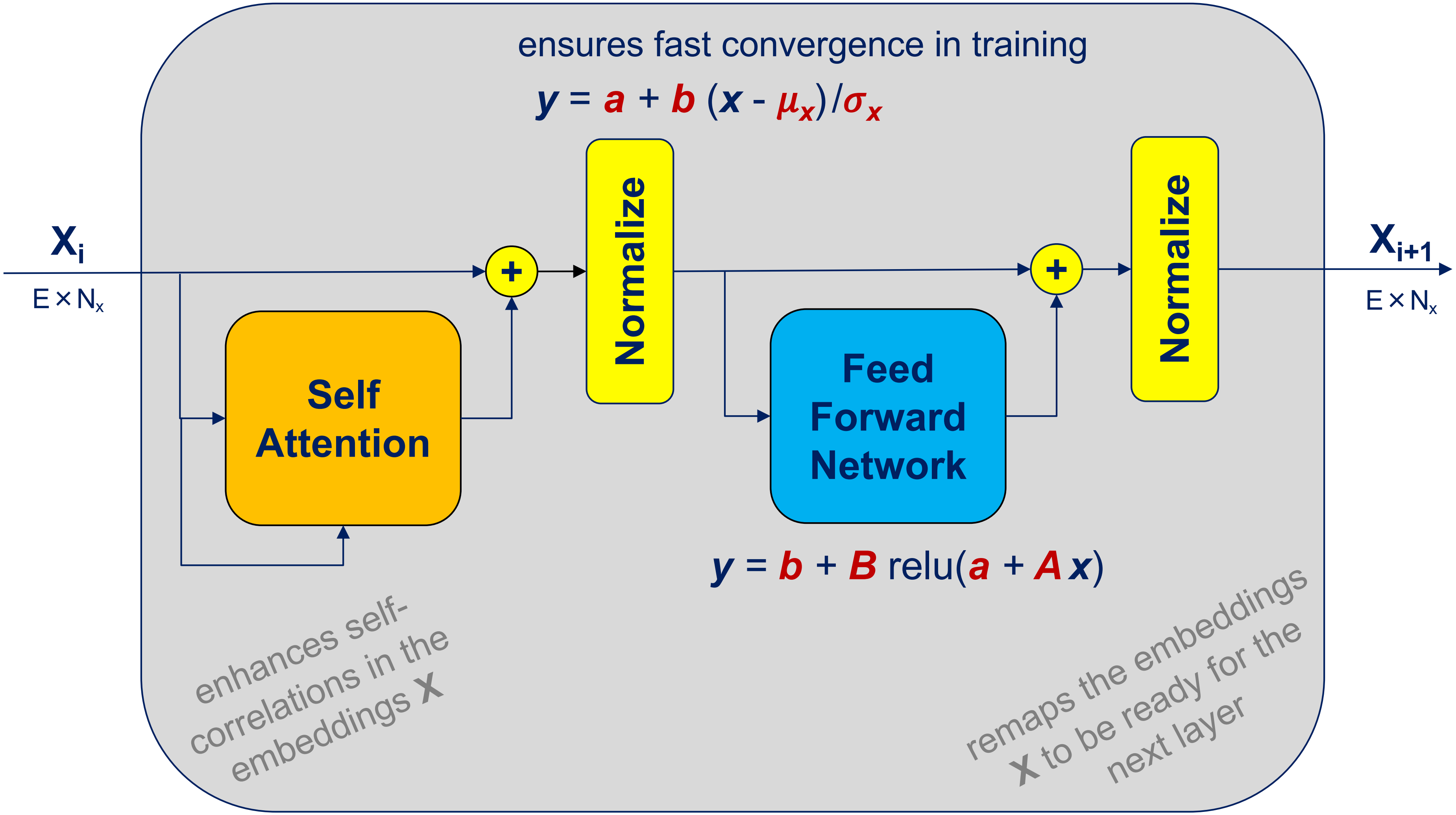
$$\boldsymbol{a} = \sum_{j=1}^{H} \overbrace{\boldsymbol{W}_{2j}\,\boldsymbol{Y}}^{\text{expansion}} \,\mathrm{softmax}(\overbrace{\boldsymbol{Y}^T \boldsymbol{W}_{1j}\,\boldsymbol{x}}^{\text{inner products}})$$

similar in spirit to projecting **x** onto the space spanned by **Y,** but with **softmax** normalization

The embedding of each **word** is built by using only the information (embeddings) of the **other words** in the sentence, in such a way to create **context**, and deal with polysemy and negations

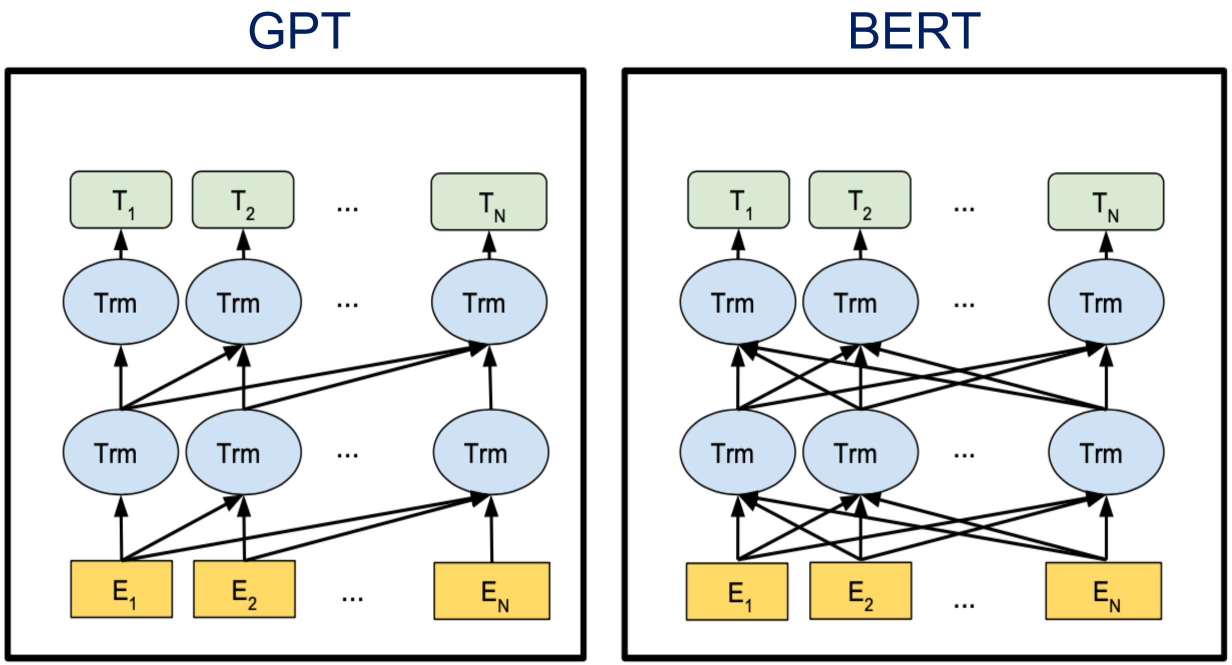
Encoder

a serie of multi-head self-attention modules



Two approaches

To capture content



Only uses preceeding words

Uses all the other words

Context
Causality

Context

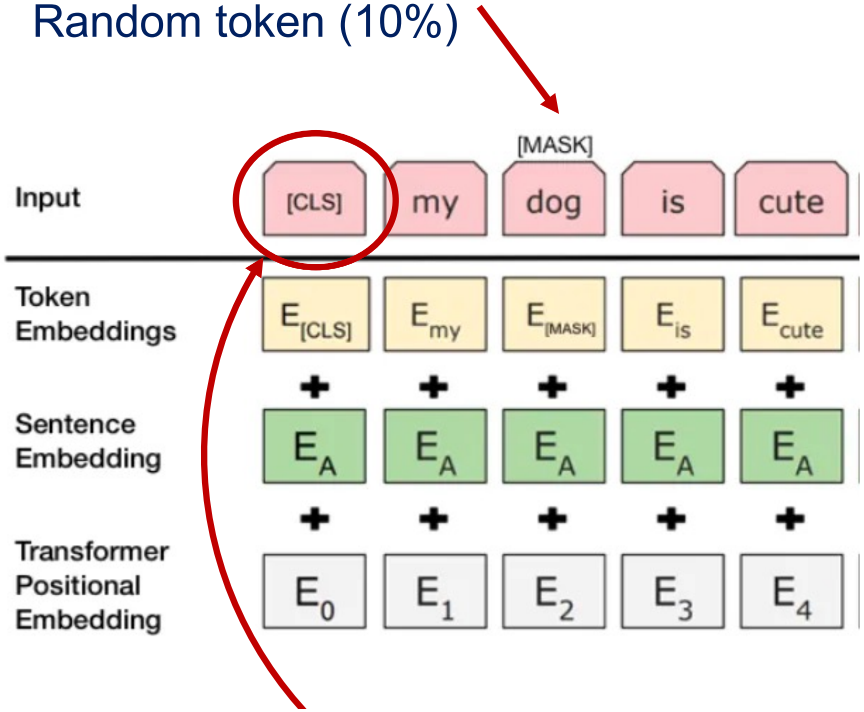
BERT pre-training procedure

BooksCorpus (800M words) + English Wikipedia (2,500M words)

Masked Language Model

15% masked tokens replaced with:

- [MASK] token (80% of the times)
- Original token (10%)
- Random token (10%)



Output [CLS] fed into an additional output layer for softmax classification (of correct/wrong next sequence)

Some words are **masked** and the algorithm is trained in order to correctly recover the original word (by **closeness** in the embedding space)

This creates context, based on **text redudancy** and **statistical occurrencies** (frequently occurrring patterns are memorized)

GPT-2

Radford, Alec, et al. "Language models are unsupervised multitask learners" (2019)

language provides a flexible way to specify tasks, inputs, and outputs all as a sequence of symbols… it is therfore possible to **train a single model** with **sufficient capacity** to infer and perform many **different tasks**

model gets complex!

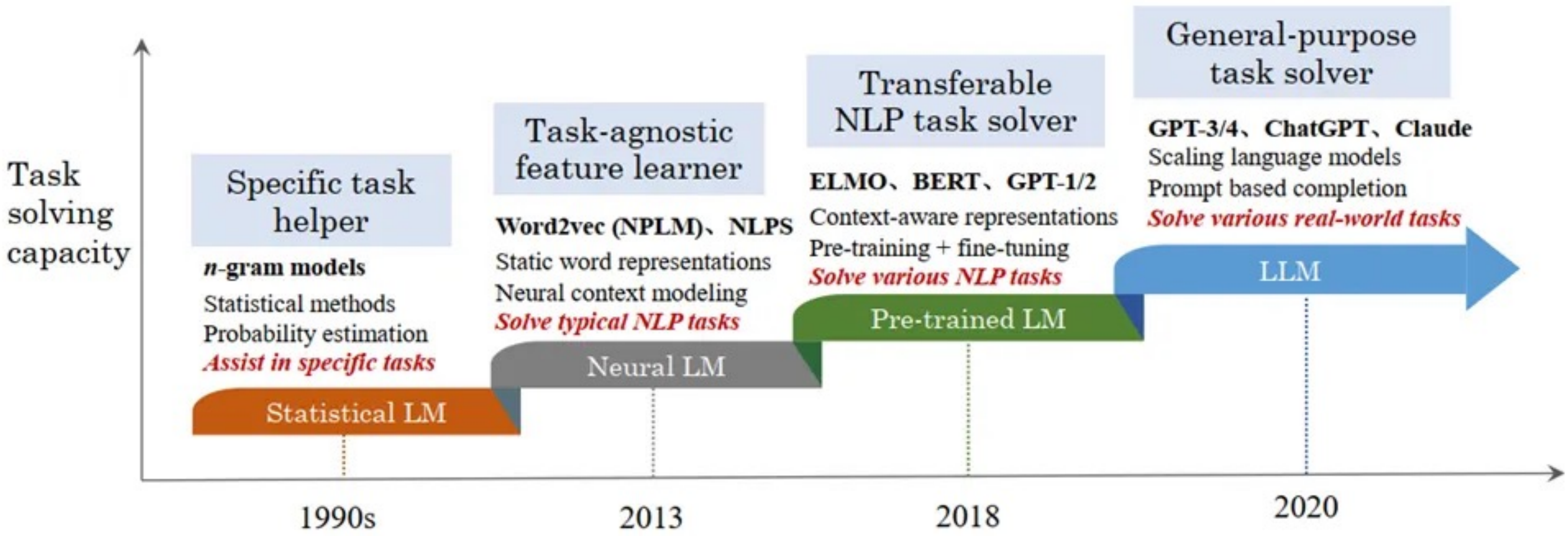
Parameters	Layers	d_{model}	
117M	12	768	GPT, BERT-base
345M	24	1024	BERT-large
762M	36	1280	
1542M	48	1600	GPT-2

data gets larger!

WebText

scraping all outbound links (45M links) from Reddit, a social media platform, which received at least 3 karma – exclude WikiPedia

An evolution map

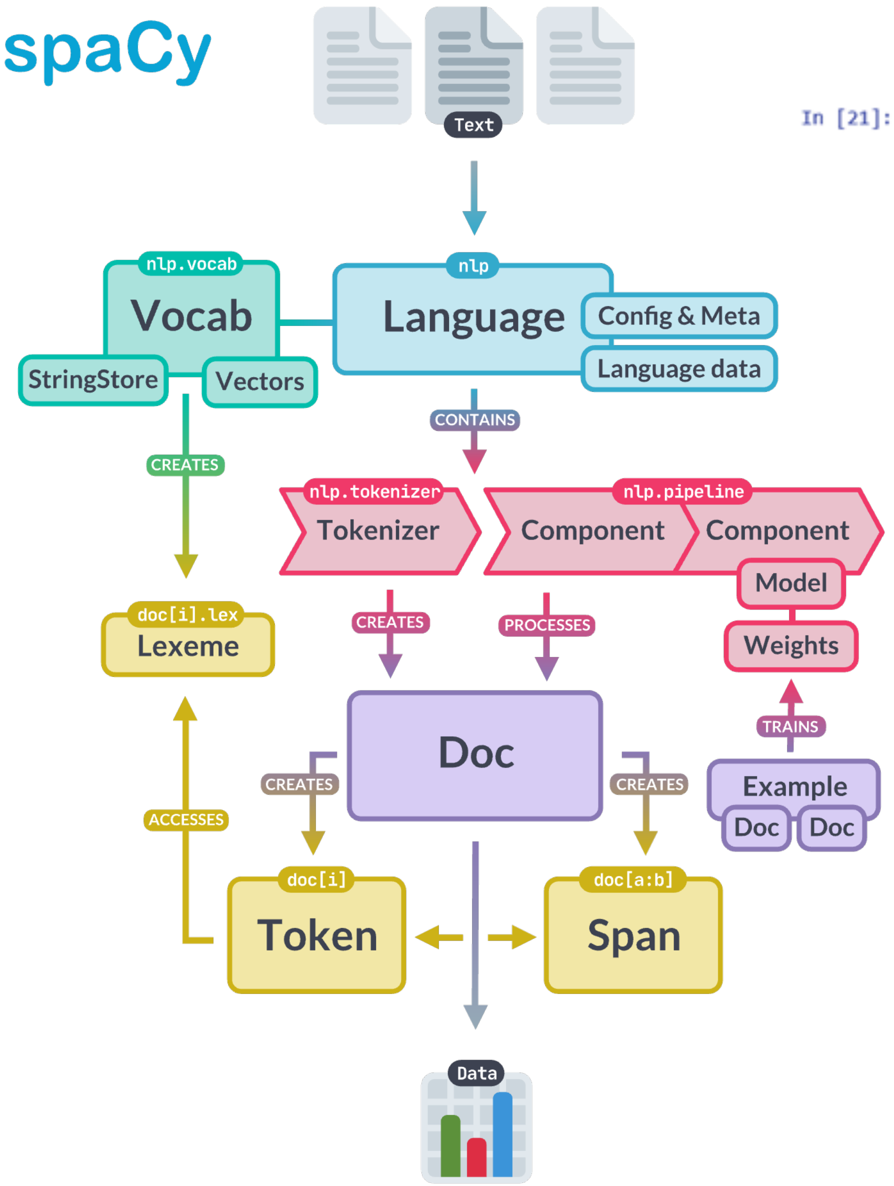


NLP tasks

some fine-tuning possibilities in NLP

Task	Description	Possible approach
Masked language prediction	predict masked words in a text	This is what BERT model is pre-trained for
Text classification or Sentiment analysis	assign a label to a given sequence of text	Apply linear transform+softmax on K classes, and train the model for the specific classification task
Text translation	translate a text	Need to pre-train a full Transfomer Architecture for this task
Summarization	generate a summary of a document	GPT example: context given by a document; then generate 100 tokens by top-2 random sampling (Fan et al., 2018), i.e., take at each step the most likely next word at random among the top-2 candidates; finally select first 3 sentences as abstract
Question answering	answer a question	GPT example: the context of the language model is seeded with example question answer pairs which helps the model infer the short answer style of the dataset
Document question answering	answer a question on a given text	GPT example: context seeded by a text; then as for question answering
Conversational	ChatBot	InstructGPT/ChatGPT: Fine-tuned models using reinforcement learning from human feedback

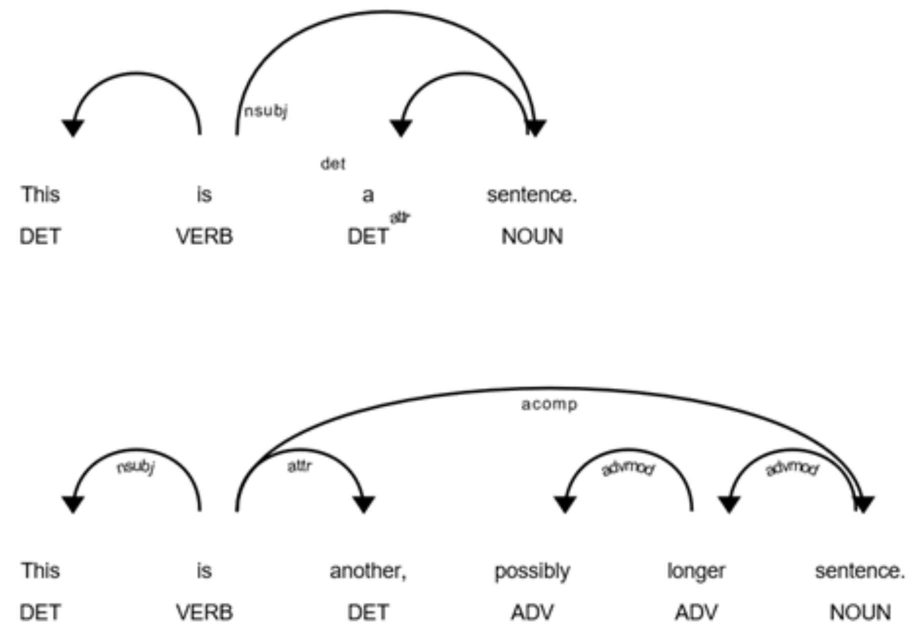
spaCy




```
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})
```



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