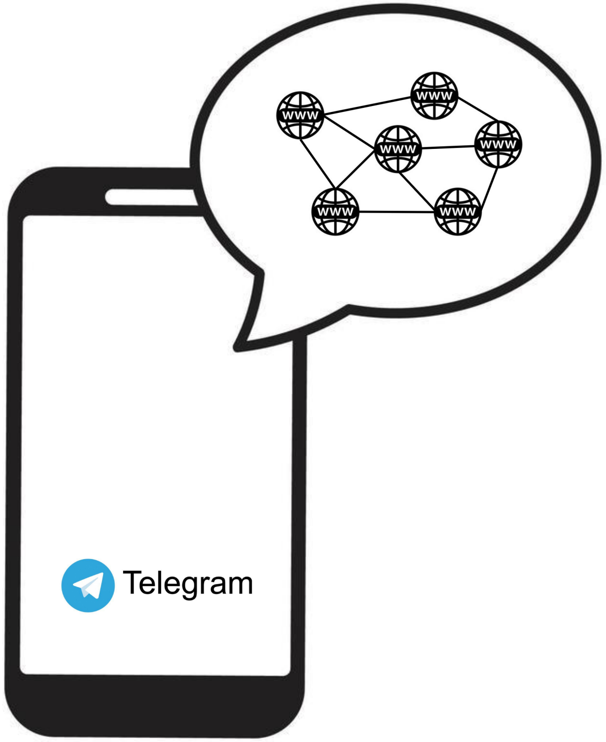
Universität Konstanz

Classifying Domains as Misinformation with a GNN



Raphaela Keßler
Konstanz, 18.02.25

Content

1. Research Question

How to classify domains as misinformation with a GNN?

Hypothesis:
GNN outperform state of the art methods for classifying domains into misinformation

2. Method

Graph Neural Network

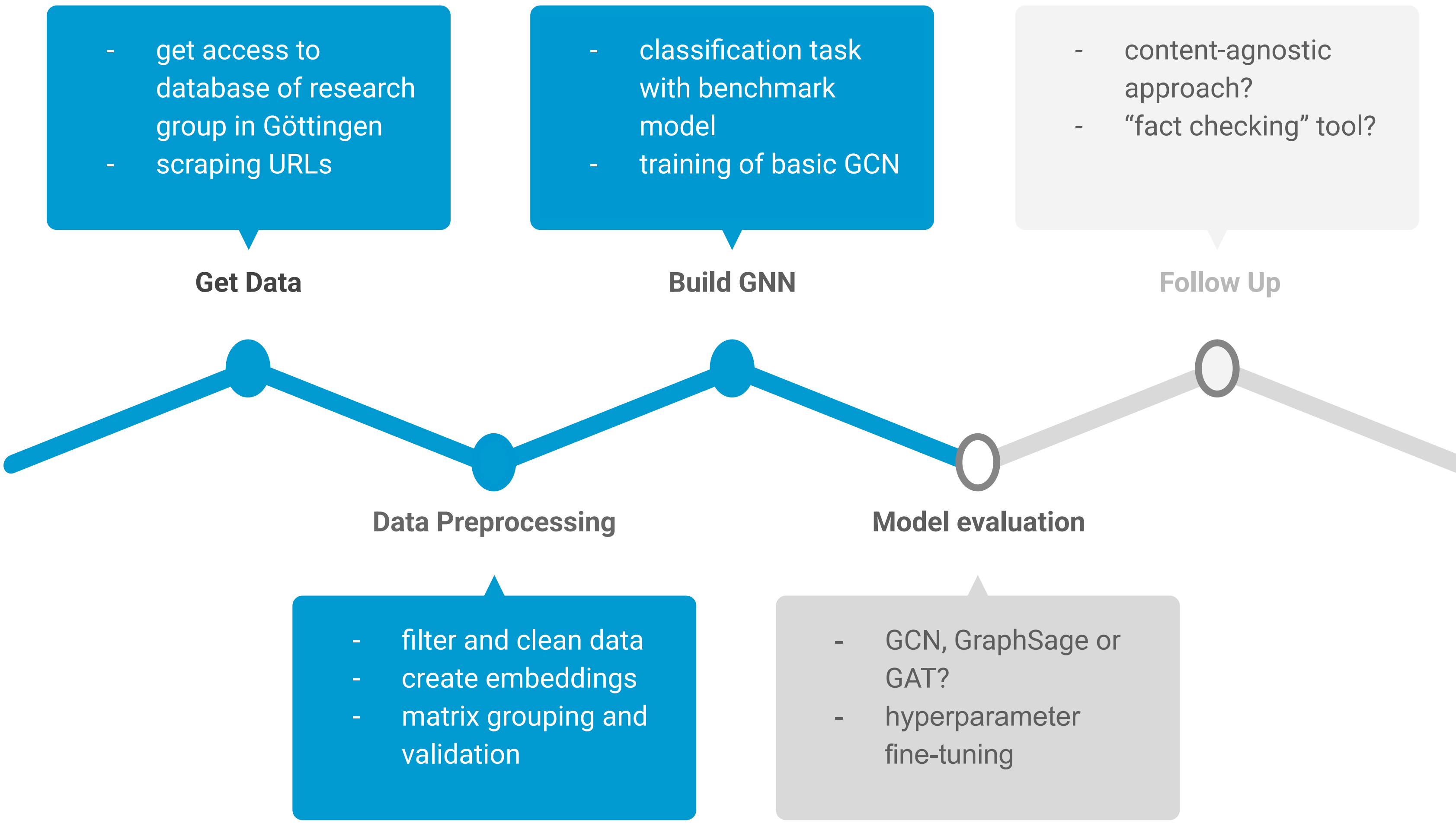
- Co-occurrences of Domains in Telegram chats build the graph
- node classification task for labeling domains as misinformation

Node Features:

- **messages:** how many messages in the data contained this url
- **avalanches**: how many avalanches (i.e. individual bursts of spreading)
- **virality**: how "bursty" the spread of the url was. virality=1 if it spreads in one large continuous burst, virality=0 if it spreads over time in many small independent bursts
- **article embedding**

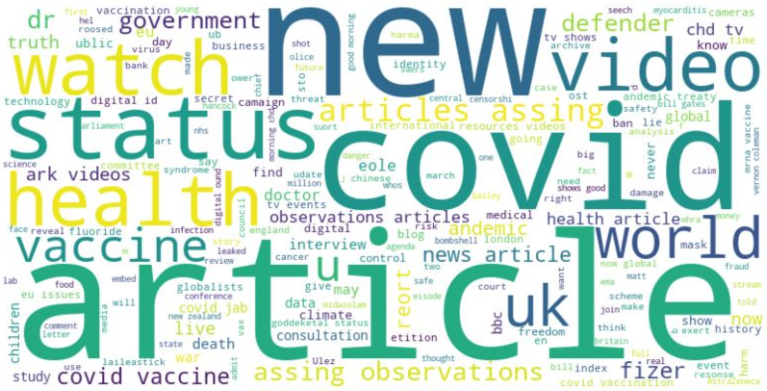
- message sentiment
- chat popularity
- chat topics

3. Work Plan & Challenges



4. Data

- Data from Priesemann Research Group in Göttingen
- Telegram Chat Data from 2020 - 2023
- Data includes
  - chats with names, description
  - URLs with timestamps
  - Chat x Url share matrix
  - 11 520 domain ratings from Lasser et al.

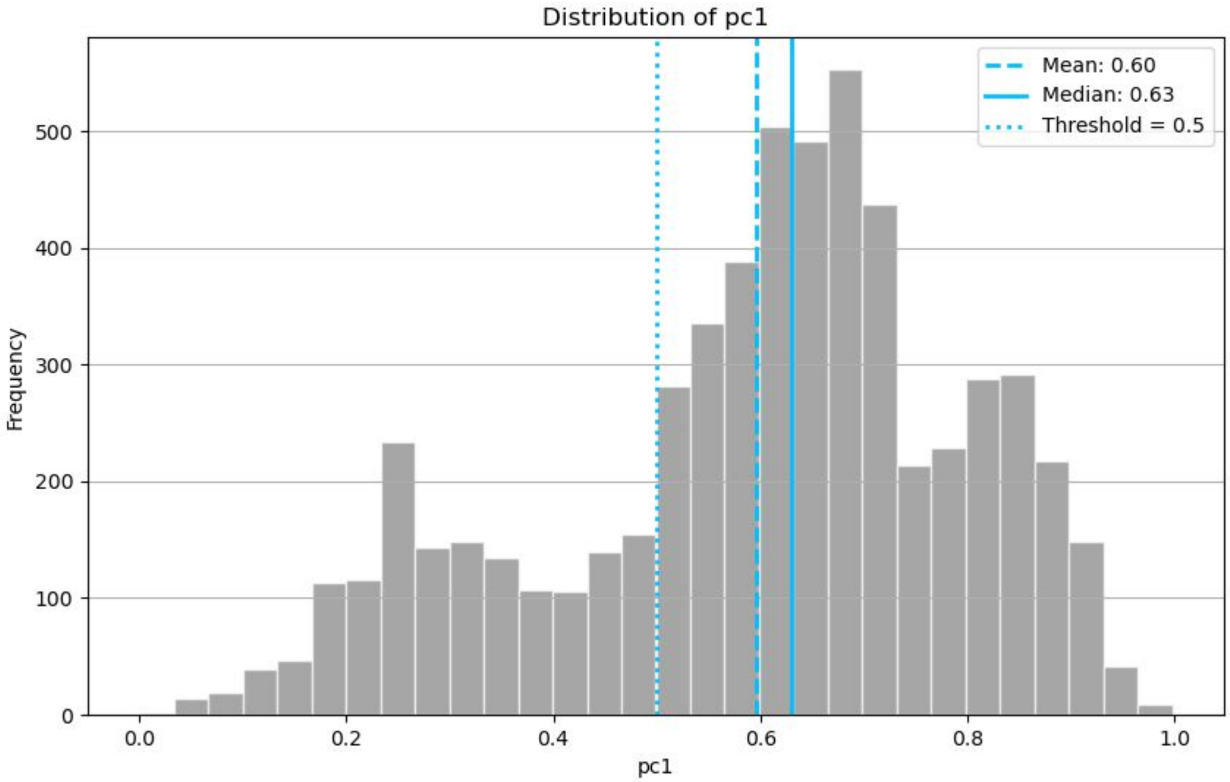


Word Cloud of Domains in URLs

	domain	pc1	afm	afm_bias	afm_min	afn
0	reuters.com	1.000000	0.962600	0.950100	0.950100	0.97
1	apnews.com	0.998049	0.960400	0.933400	0.933400	0.98
2	charitynavigator.org	0.985752	0.929423	0.934419	0.909962	0.92
3	rollcall.com	0.982851	0.916600	0.911500	0.911500	0.92
4	smithsonianmag.com	0.971184	0.891200	0.883200	0.883200	0.89
...	...	...	...	...	...	...

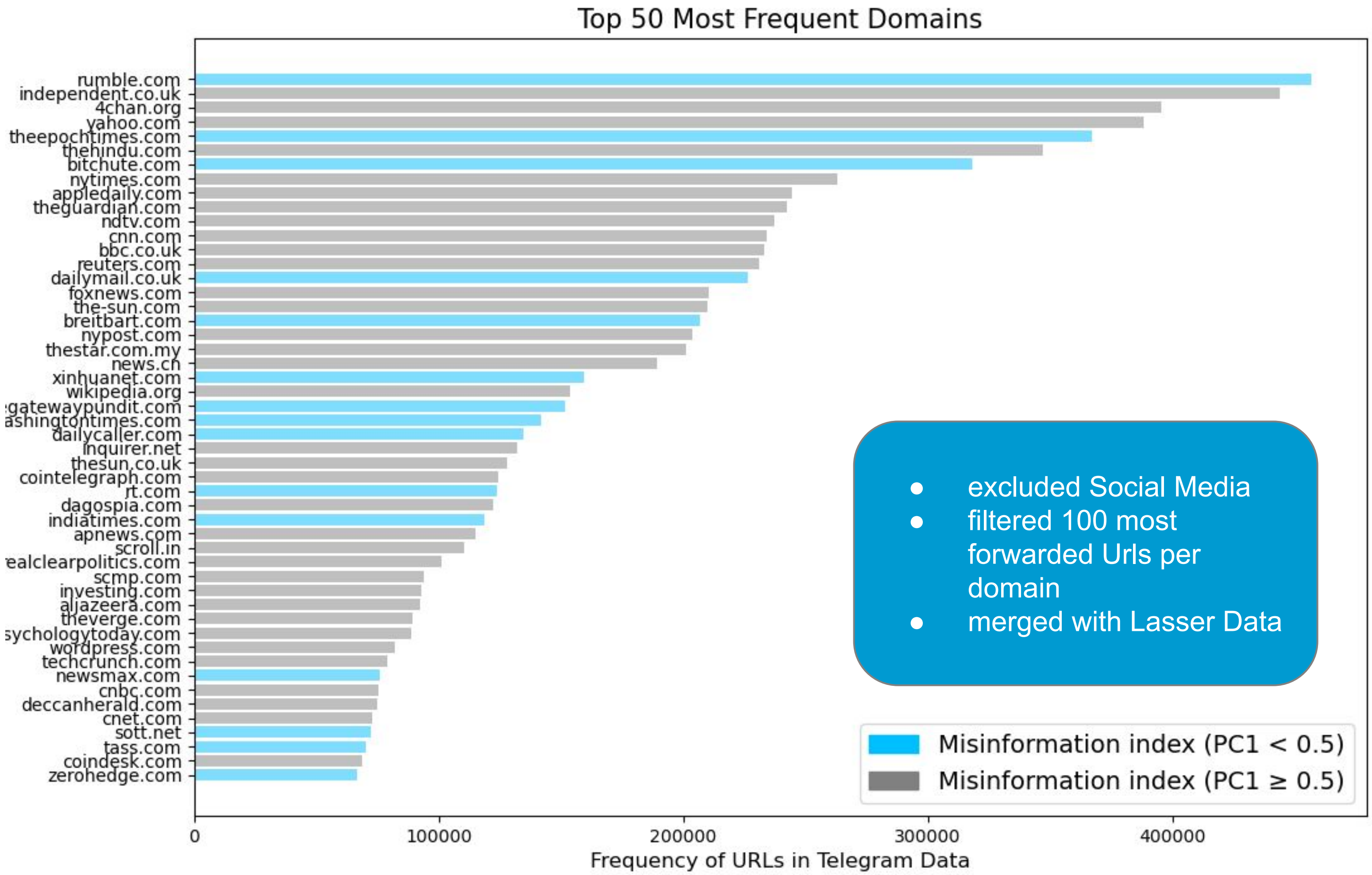
Cut-out of Lasser at al. domain ratings

4. Data - Misinformation Score



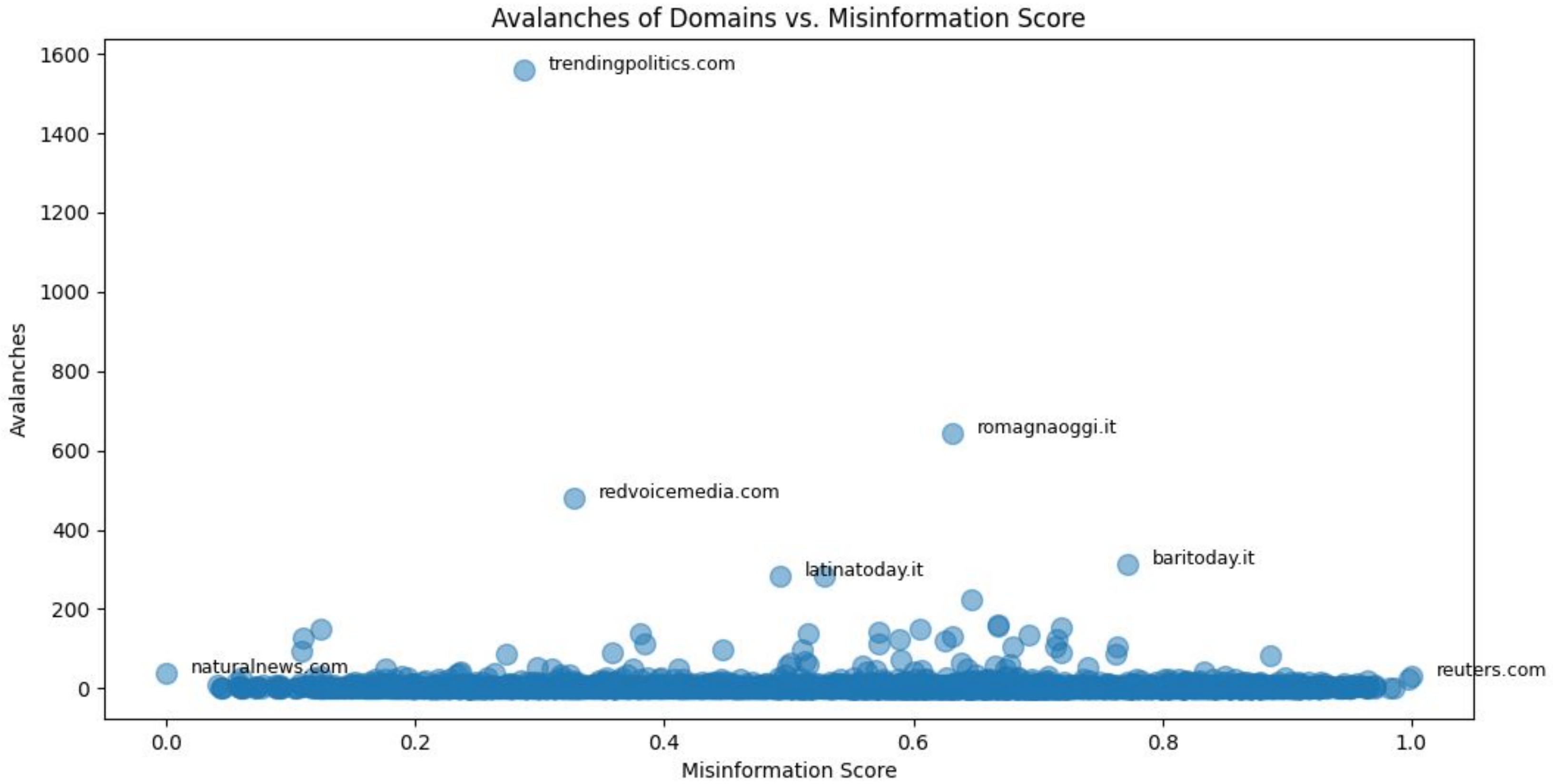
0 = Misinformation; 1 = Information

4. Data - Domains

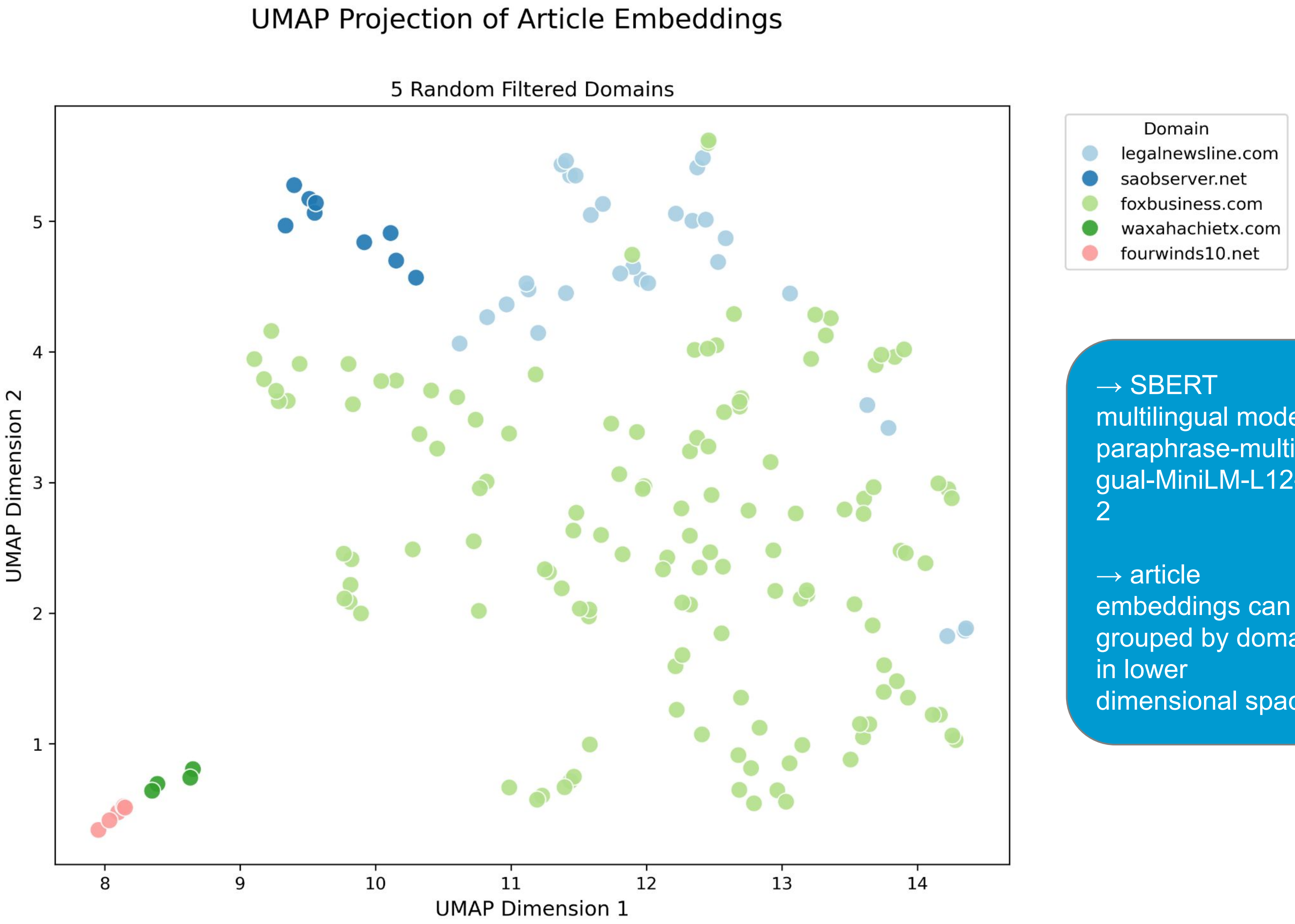


- excluded Social Media
- filtered 100 most forwarded Urls per domain
- merged with Lasser Data

4. Data - Domains



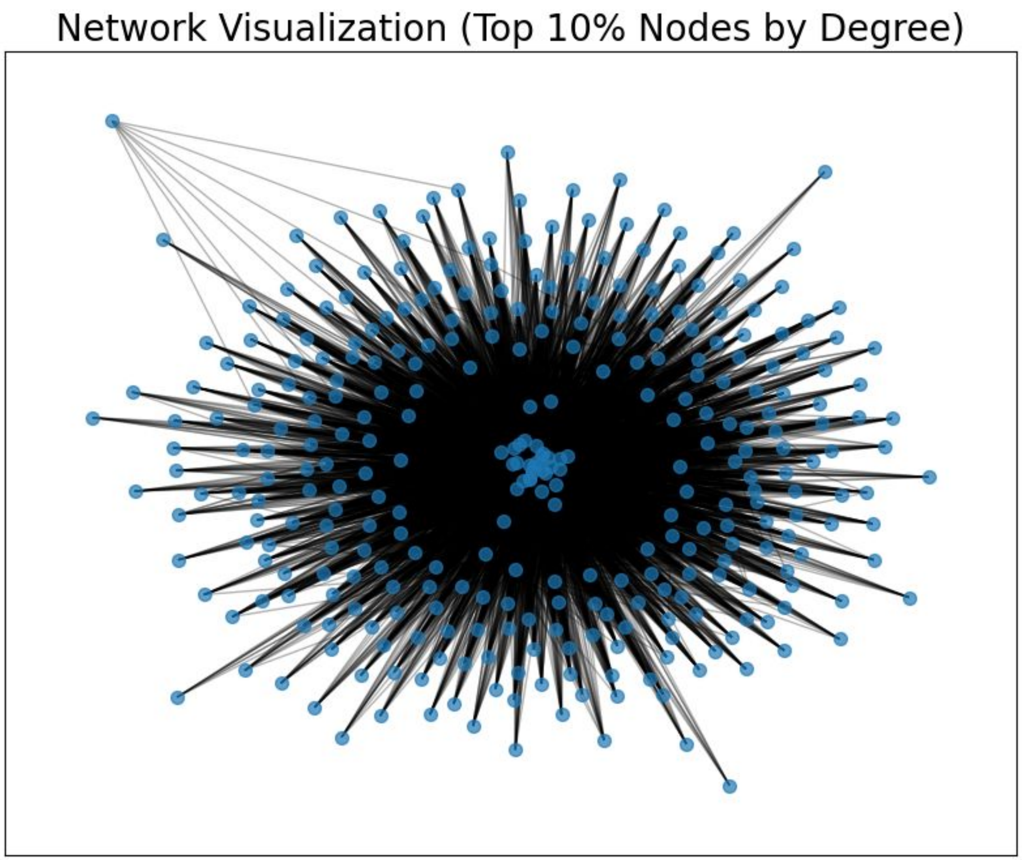
4. Data - Article Embeddings



→ SBERT multilingual model paraphrase-multilingual-MiniLM-L12-v2

→ article embeddings can be grouped by domain in lower dimensional space

4. Data - Network Structure



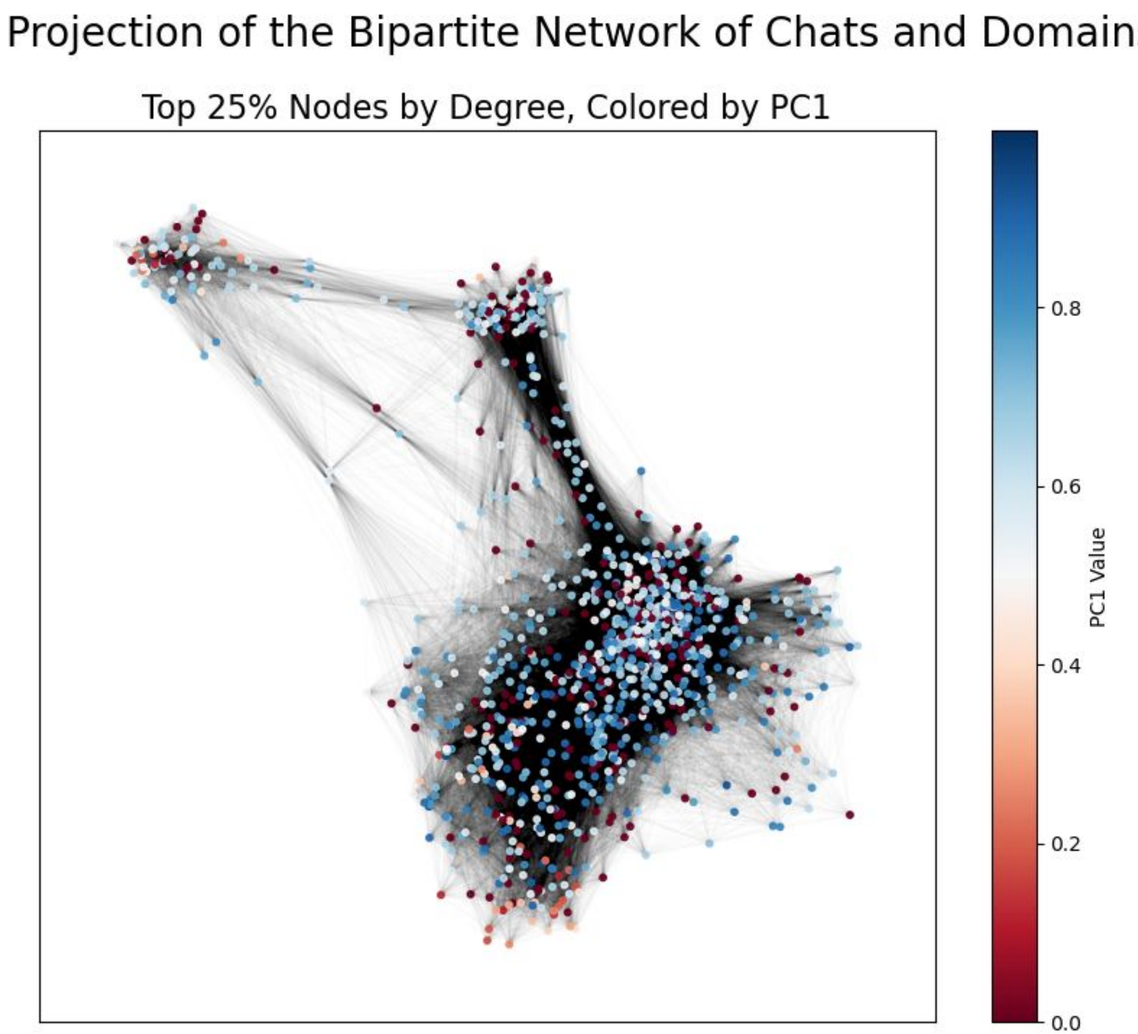
→ 388 chats

→ 4753 train domains

→ 1189 test domains

→ 1519640 edges

4. Data - Network Structure



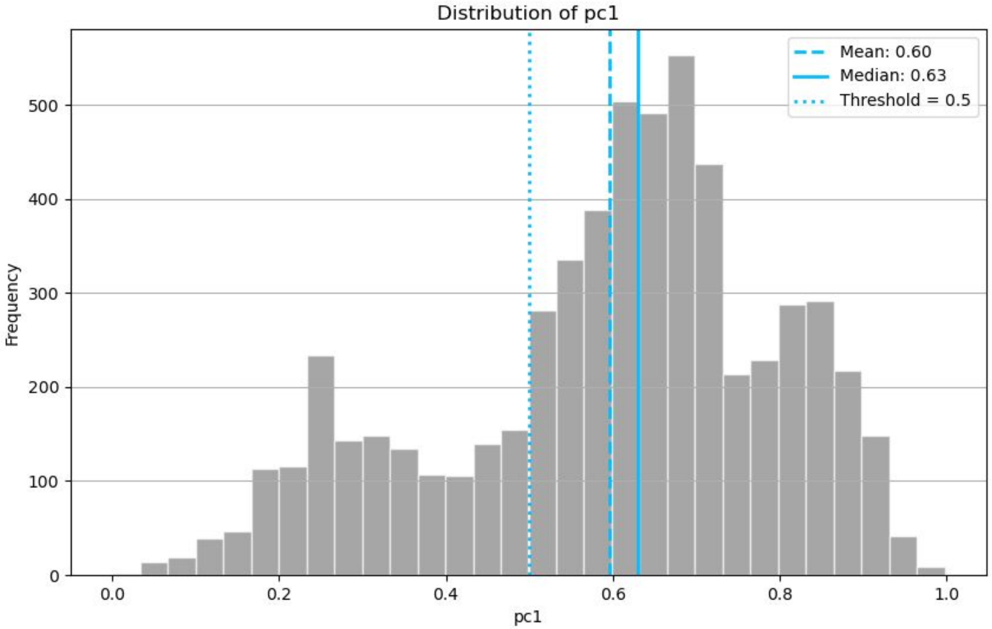
→ Validating Network with Bipartite Configuration model

→ using Poisson Distribution to check for statistically validated node similarities

→ reduced edges from 1519640 to 154254

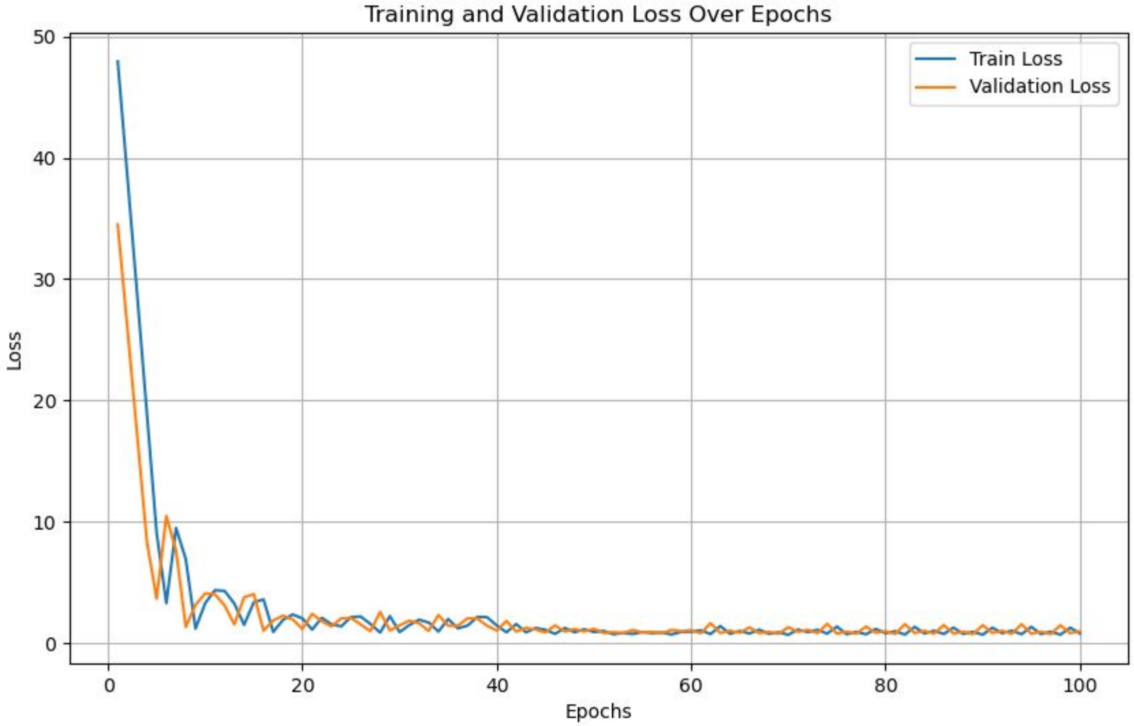
5. Model - Classification

- Classification Task as starting point
- goal is binary classification, with “trustworthy” and not “trustworthy” domains as a result
- Finding Threshold is difficult, because of unbalanced data
- for now: Threshold = 0.5



5. Model - GCN

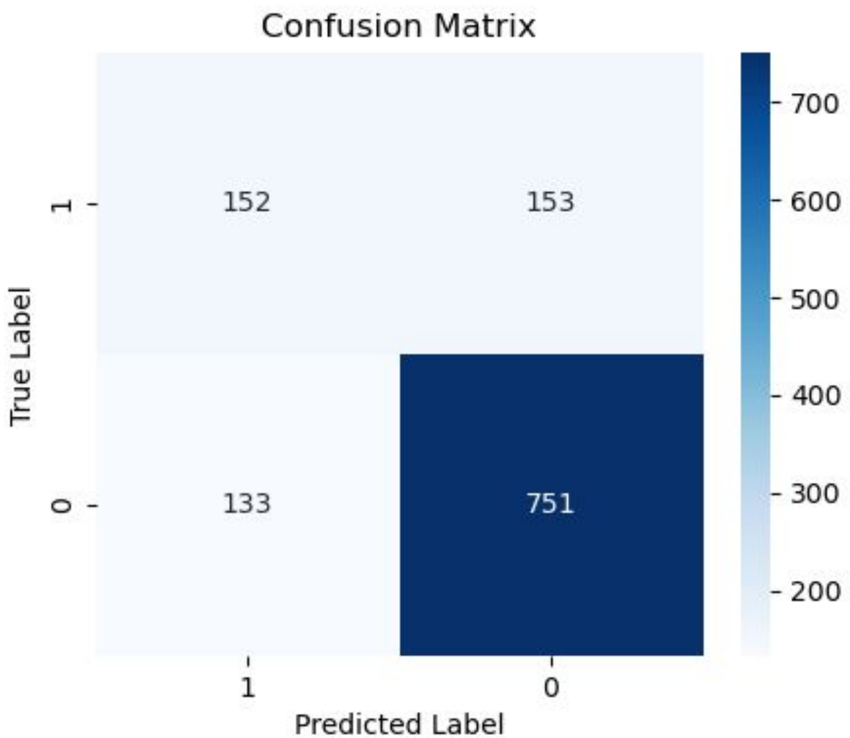
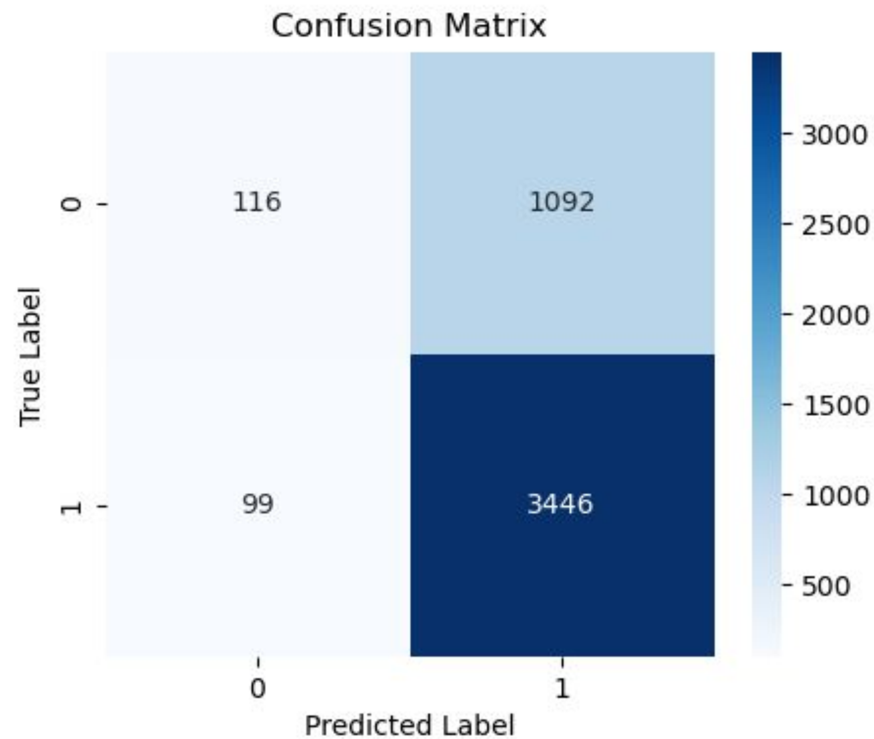
First basic GCN with the features:
'virality', 'year', 'avalanches', 'messages'



GCN(
  (conv1): GCNConv(388, 64)
  (conv2): GCNConv(64, 2)
)

lr = 0.0001
epochs = 100

5. Model - SBERT vs. GCN

SBERT - Classification Task – paraphrase-multilingual-MiniLM-L12-v2	**GCN - Classification Task** – with validated Network
	

5. Model - SBERT vs. GCN

SBERT - Classification Task

Classification Report:

	precision	recall	f1-score	support
0	0.53	0.50	0.52	305
1	0.83	0.85	0.84	884
accuracy			0.76	1189
macro avg	0.68	0.67	0.68	1189
weighted avg	0.75	0.76	0.76	1189

GCN - Classification Task

Classification Report:

	precision	recall	f1-score	support
Class 0	0.54	0.10	0.16	1208
Class 1	0.76	0.97	0.85	3545
accuracy			0.75	4753
macro avg	0.65	0.53	0.51	4753
weighted avg	0.70	0.75	0.68	4753

(messed up to test with the test data)

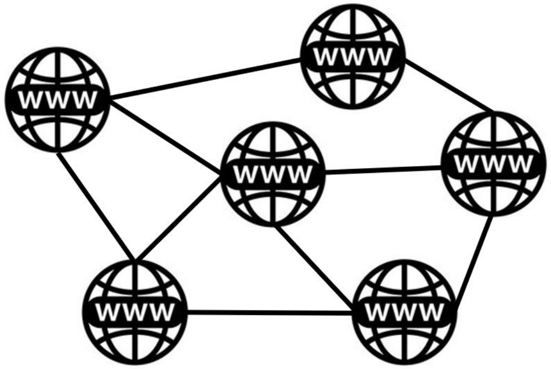
6. Next Steps

- adding additional features: messages and chat topics
- try another BERT model for embeddings
- set different thresholds and balance the data
- improving the models
- experiment with different Graph Neural Networks
- try content agnostic approach (without article embeddings)

Universität Konstanz

Thank You!

Questions or Advices?

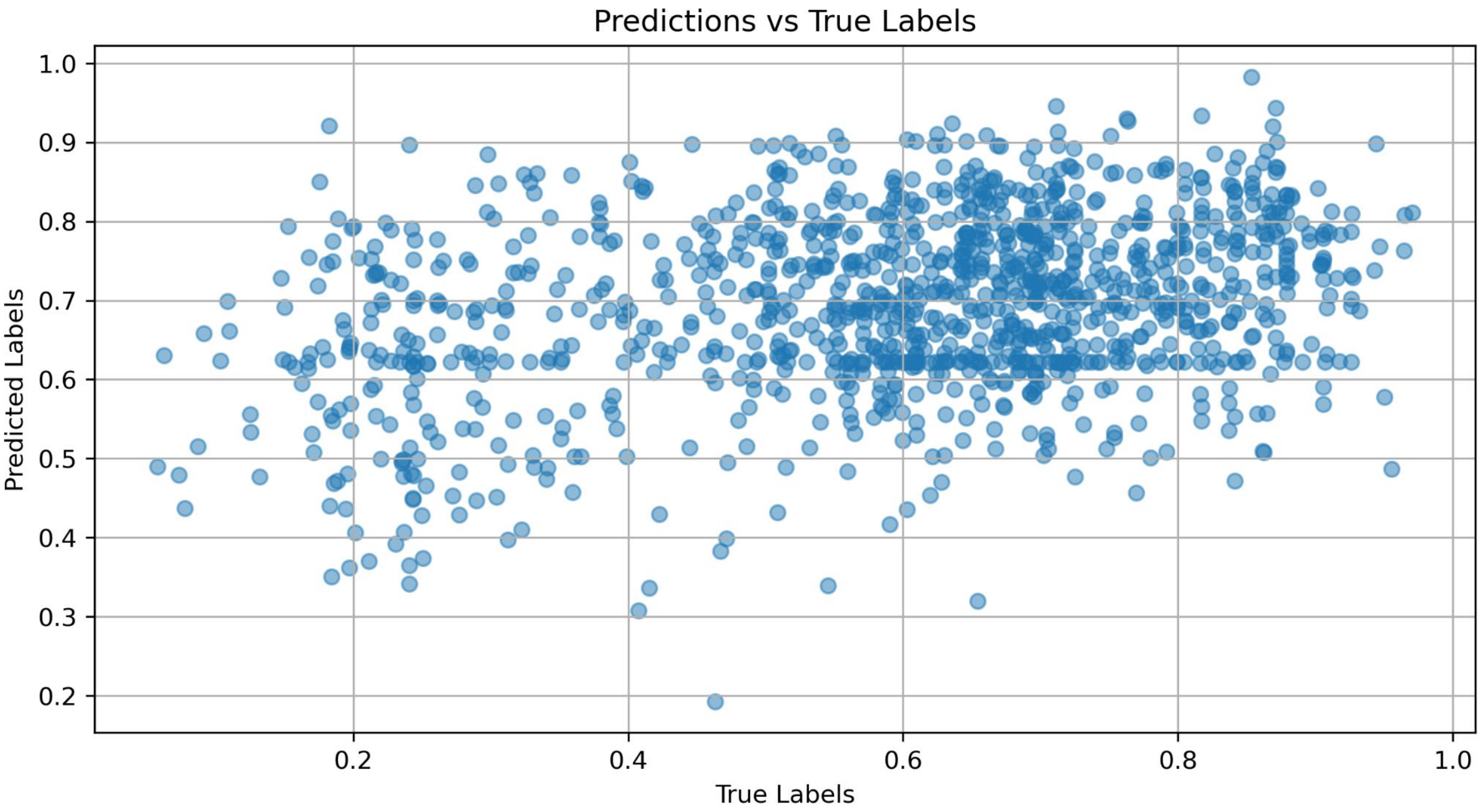


Sources

- https://www.statista.com/statistics/1263360/most-popular-messenger-apps-worldwide-by-monthly-downloads/
- https://www.skillademia.com/statistics/telegram-statistics/
- Carragher et al. 2024, https://doi.org/10.1184/R1/25174193.v1
- A. Maulana and J. Langguth, doi: 10.1109/SNAMS60348.2023.10375407.
- Hause Lin, Jana Lasser, Stephan Lewandowsky, Rocky Cole, Andrew Gully, David G Rand, Gordon Pennycook, https://doi.org/10.1093/pnasnexus/pgad286

5. Model - Bert

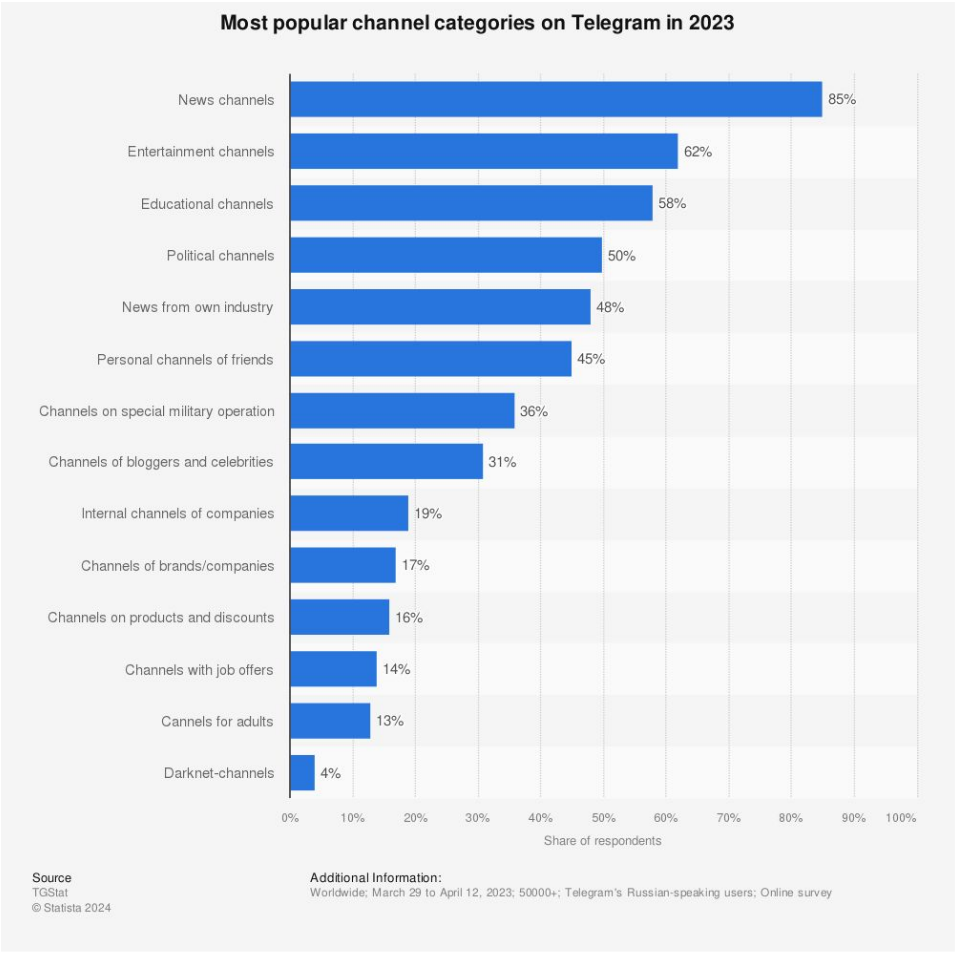
Regression Task with last layer of SBERT model



5. Model - Bert

Regression Task with last layer of SBERT model

3. The Data



Telegram

Telegram:
- free instant messaging service
- third most downloaded messenger worldwide
- 800 million monthly active users
- end-to-end encryption and high user privacy level