Global Business Networks

Christian Breitung and Sebastian Müller, Technical University of Munich

Contribution

- We generate global business networks (BNs) by applying LLMs (T5-XXL, GPT-3) to business descriptions of around 80,000 firms

- We leverage LLMs to generate historical business descriptions

- We show how lookahead bias of LLMs may be addressed → masking company-specific info

Main challenges

- International reports lack harmonized report structure → Use AI to streamline business information → business descriptions

- Word-based methods insufficient for relatively short business descriptions → embedding models

Data

	#Stocks	USA	USA_MV	#Stocks	Global	Global_MV
Panel A: AI-Gen (%)						
2001	5793	70.72	84.46	29309	42.93	76.29
2006	4496	90.48	93.73	33187	58.29	84.70
2011	3777	96.13	95.79	37403	65.65	86.74
2016	3557	96.15	97.67	37542	68.78	85.40
2021	2601	97.12	99.32	40939	67.57	87.54
Panel B: AI-Gen + SDC + S&P Global (%)						
2000	6036	88.17	97.17	27944	56.41	87.96
2001	5793	90.33	98.15	29309	62.93	91.24
2006	4496	96.73	99.63	33187	80.00	95.86
2011	3777	99.31	99.87	37403	89.28	97.11
2016	3557	99.58	99.96	37542	97.49	99.40
2021	2601	99.81	99.95	40939	96.59	98.86

- 67 Countries

→ AI generated descriptions available for up to 2/3 of dataset (87% market value)

→ Including available historical descriptions from Refinitiv and S&P Global → Coverage increases up to 97% (99% market value)
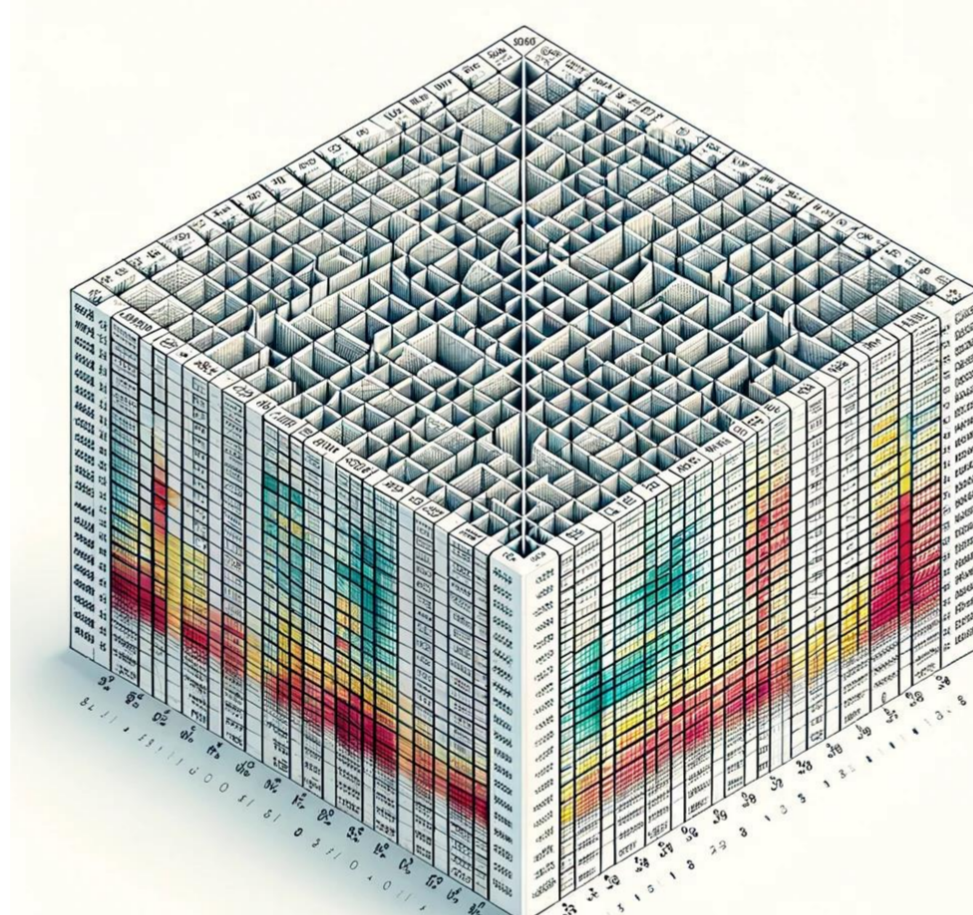
Methodology

Business descriptions

Methods: BOW, T5-XXL, ADA-002

Embeddings

Cosine Similarity matrix

Global Business Network

Competitor Identification

Table 7: **Detection rate of US competition links (full dataset)**

	ORG100	MASK100	ORG50	MASK50	ORG10	MASK10
Panel A: Disclosed 10-K (%)						
TNIC	41.37	-	35.06	-	20.72	-
BOW	27.00	-	19.08	-	8.47	-
T5-XXL	36.22	36.34	27.32	27.14	13.36	12.45
ADA-002	38.46	45.84	29.28	37.41	13.24	19.18
Panel B: Comparable Company Analysis (%)						
TNIC	30.25	-	24.65	-	12.19	-
BOW	20.02	-	15.28	-	6.40	-
T5-XXL	28.28	27.70	21.95	20.95	10.38	9.30
ADA-002	27.12	32.14	20.10	25.58	8.10	12.00
Panel C: Factset Revere (%)						
TNIC	50.07	-	39.56	-	19.14	-
BOW	34.47	-	24.14	-	9.21	-
T5-XXL	47.82	47.63	35.11	35.28	15.08	14.87
ADA-002	45.77	55.77	33.20	43.17	13.21	19.40

→ Context-aware business networks outperform word-based networks

→ Outperformance remains after controlling for look-ahead bias via masking

Lead-lag effect

Table 9: **Lead-lag effect**

Factor	TNIC	BOW	T5-XXL	ADA-002
US	0.64***	0.55***	0.78***	0.81***
	(3.5)	(4.64)	(5.6)	(5.33)
US-Masked	-	-	0.7***	0.83***
	-	-	(4.65)	(6.05)
Global	-	1.31***	1.78***	1.75***
	-	(7.25)	(7.98)	(7.91)
Global-Masked	-	-	1.67***	1.72***
	-	-	(7.85)	(9.22)

- Long-short portfolios based on past month peer performance (lead-lag effect)

US market:

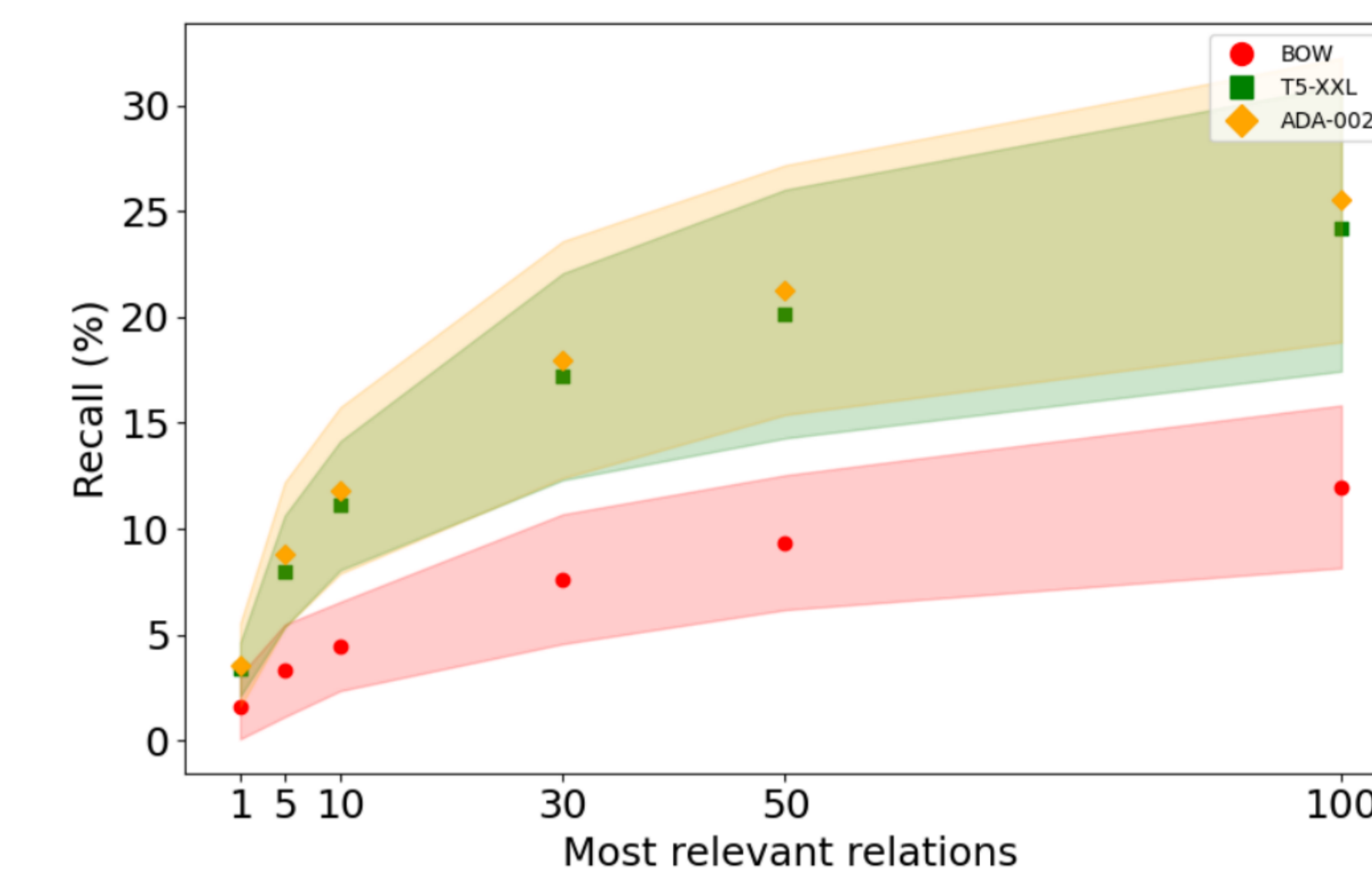→ Context-aware networks outperform TNIC (same firm universe) and word-based approach

→ Masking leads to small decreases in alpha

Global market:

→ Context-aware networks outperform BOW

→ Masking does not significantly reduce alpha

Identifying M&A targets

- Target firm identification using masked networks:

- M&A deals from SDC Platinum (2000-2022)

→ T5-XXL / ADA-003: 25% (yearly average) of M&A targets among most related 100 stocks

→ BOW: Only 10% of targets in *BN*

Relation classifier

val_type	Class 0	Class 1	Accuracy	Precision	Recall	F1
Panel A: Competitor vs. Non-Competitor (%)						
all	6517	3491	88.74	84.97	82.27	83.60
same	6290	3350	89.12	85.59	82.60	84.07
other	227	141	78.80	71.43	74.47	72.92
Panel B: Supplier vs. Customer (%)						
all	4765	2172	91.02	81.17	92.86	86.62
same	4602	2038	91.88	82.24	93.82	87.65
other	163	134	71.72	65.62	78.36	71.43

- We fine-tune RoBERTa models to differentiate between competitors, supplier and customers

→ We achieve f1 scores up to 87%

Contact:

Christian Breitung
Technical University of Munich
TUM School of Management
Center for Digital Transformation
Bildungscampus 9
74076 Heilbronn
Germany

Email: christian.breitung@tum.de