A Framework to Construct Financial Causality Knowledge Graph from Text

Author: Ziwei Xu, Hiroya Takamura, Ryutaro Ichise

Presenter: Ziwei XU

Date: Feb 5th 2024

Content

1. Introduction

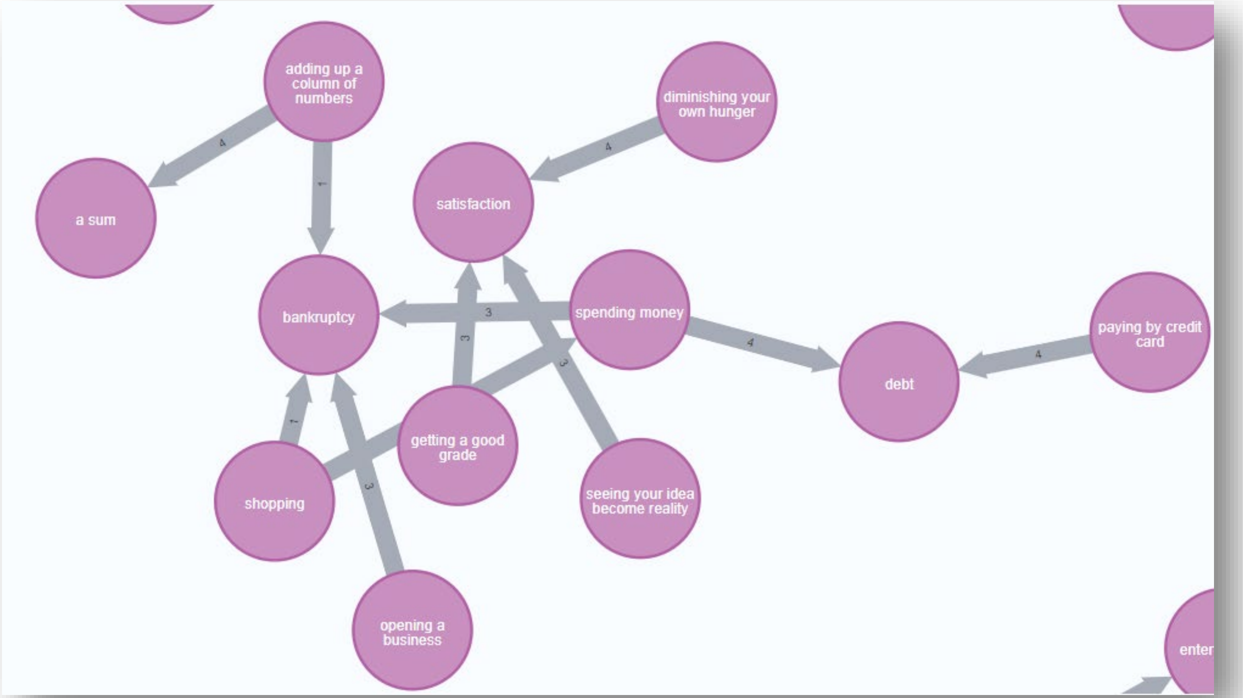
Motivation:

- Causality analysis holds a prominent role in finance
- The presentation of causality could offer valuable insights for risk mitigation, investment decisions, and portfolio optimization.

Gaps:

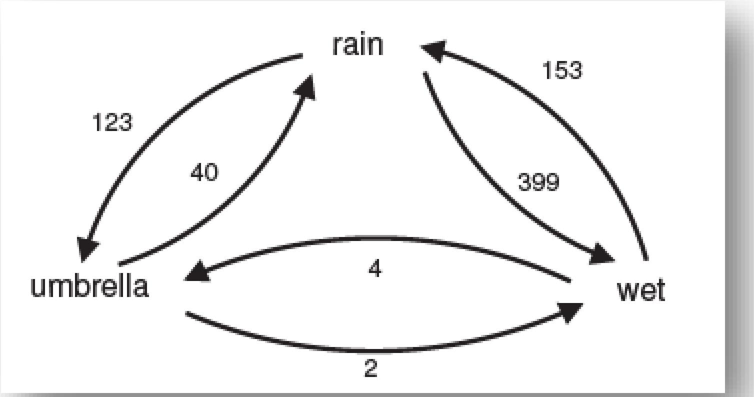
- Recent research has extensively investigated the identification of causality from text
- yet there is still a significant deficiency in providing a comprehensive causality presentation from those textual discoveries

2. Background



ConceptNet-Cause [1]

- Causality here is scattered, common-sense and action-based
- Causality is contextless, thus prone to be doubtable

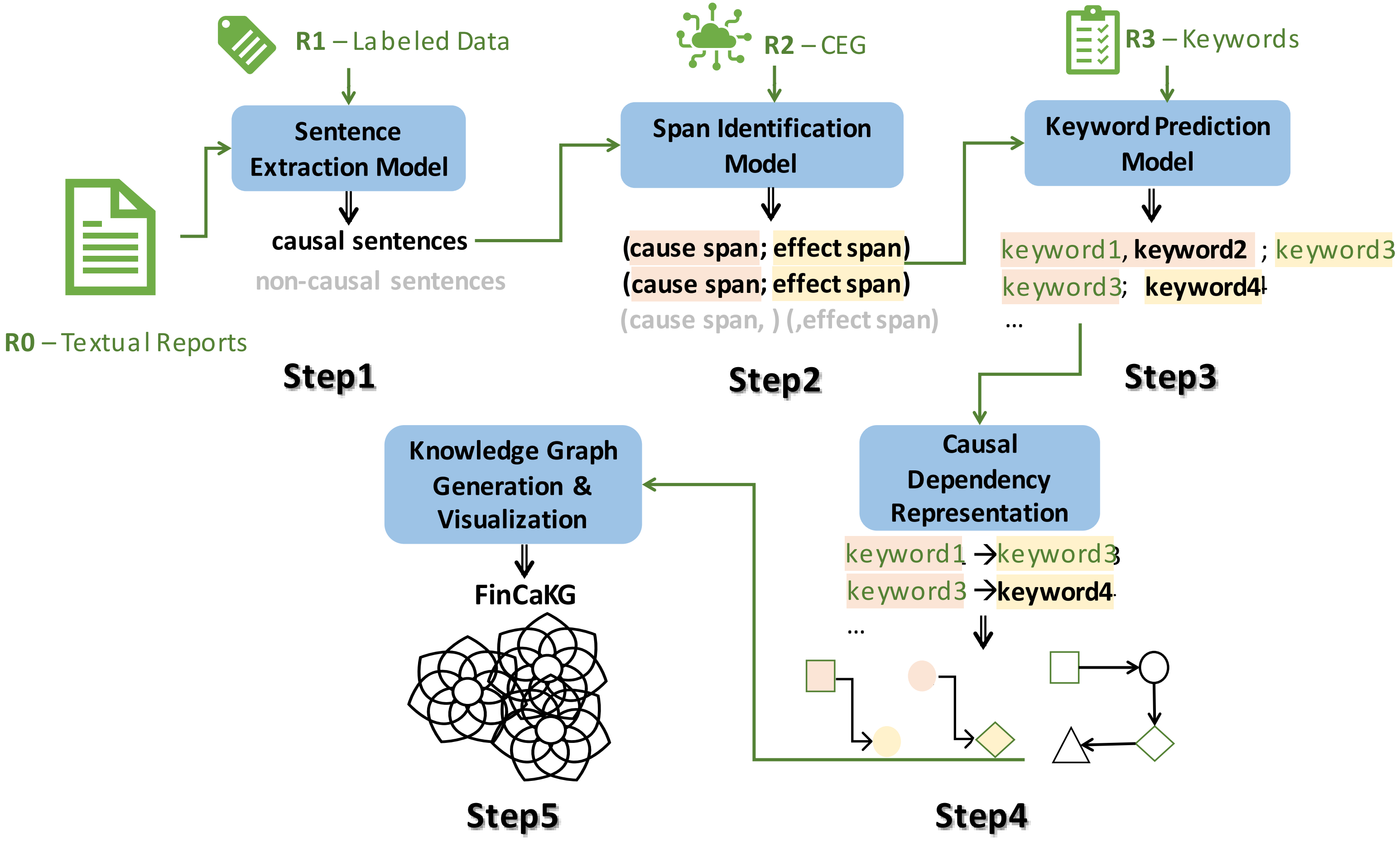


Cause Effect Graph [2]

- Information is redundant (not only nouns, but adj. , adv., verbs)
- hard to detect the semantic meaningful causality from massive pairs(89.1 M), here is the simplest and most meaningful example

[1] R. Speer, et al."Conceptnet 5.5: An open multilingual graph of general knowledge," AAAI 2017
[2] Z. Li, et al."Guided generation of cause and effect," IJCAI 2020

3. FinCaKG Construction Framework - Steps

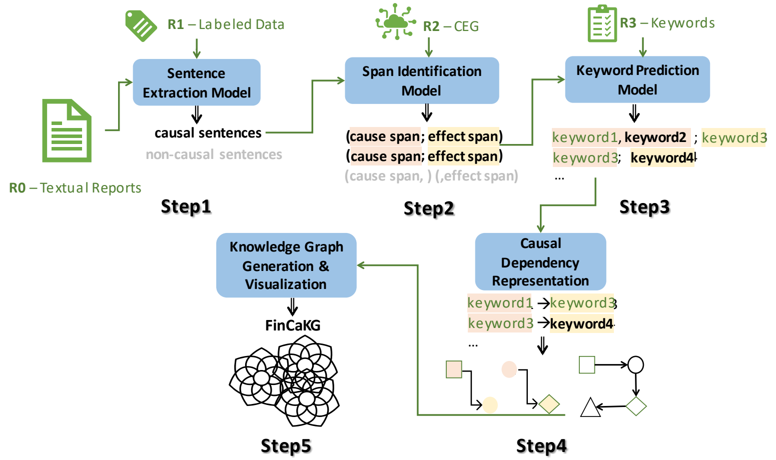


Description of constructive steps:

- **Step1**: extract causal sentences
- **Step2**: identify cause-effect textual spans from causal Sentences
- **Step3**: predict financial keywords of cause-effect spans
- **Step4**: formulate the keyword causality dependency from cause-effect spans
- **Step5**: generate and visualize FinCaKG

Xu. et al., A Framework to Construct Financial Causality Knowledge Graph from Text, IEEE-ICSC 2024

3. FinCaKG Construction Framework - Datasets



Financial 10-k Reports(R0)

The financial report in S.E.C of the top 3000 company in the last 5 years.

Labeled Data (R1)

The causal labels for sentences in financial reports.

Cause Effect Graph - CEG (R2)

Cause Effect Graph contains 89.1 million causal pairs.

Financial Keywords (R3)

We consider Investopdia’s vocabulary are financial Keywords.

Xu. et al., A Framework to Construct Financial Causality Knowledge Graph from Text, IEEE-ICSC 2024

3. FinCaKG Construction Framework - Experiment Settings

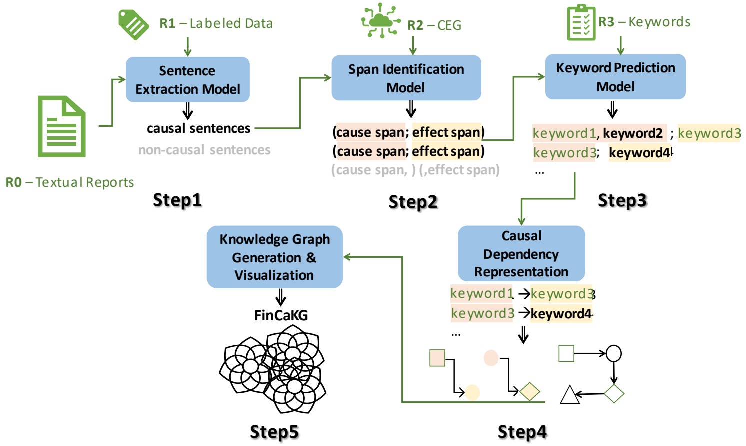


TABLE I: The statistics of resources applied in relative steps.

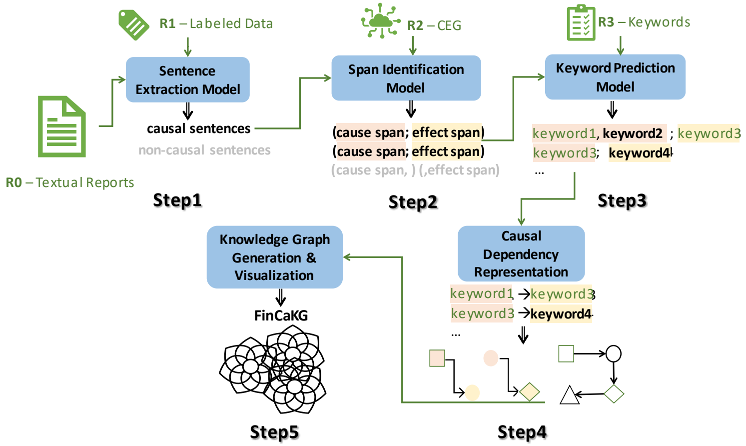
Steps	Units	Train	Prediction
Step1	#sents (#causal sents)	$14k(1k)$	$10m(525k)$
Step2	#sents	$2.7k$	$525k$
	#sents after postprocessing	-	$497k$
Step3	#sents	$497k \times \frac{9}{10}$	$497k \times \frac{1}{10}$
	#unique investopedia NPs	1317	1317
	#unique new financial NPs	-	1623

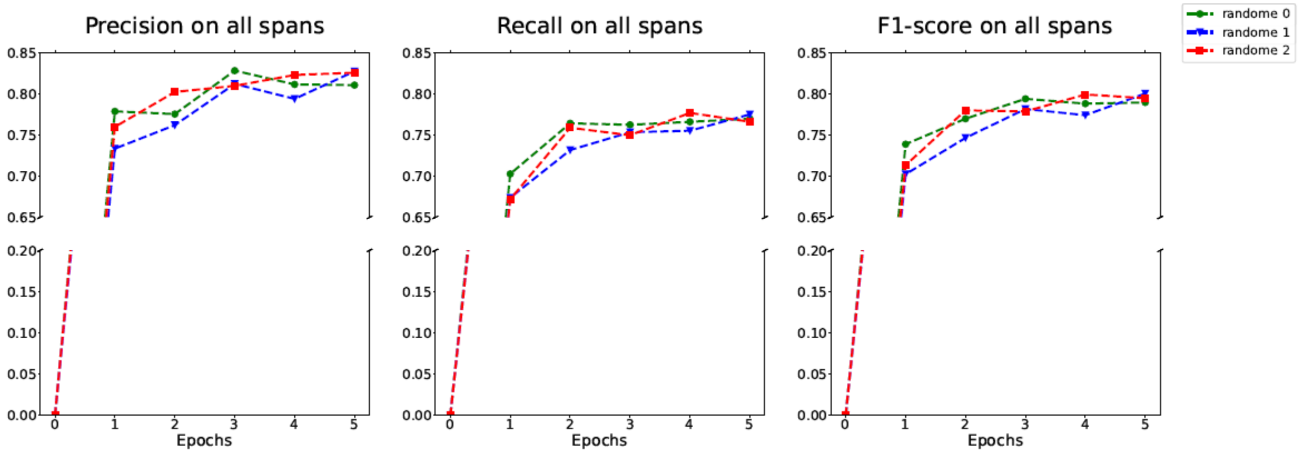
TABLE II: The parameters of models in each step.

Steps	labels	models	ep.	bt.	lr.	opt.
Step1	0, 1	BERT	10	8	2e-5	epsilon =1e-8
Step2	B-C, I-C, B-E, I-E, O	ilab-FinCau [24]	5	4	5e-5	
Step3	B-F, I-F, O	XLM-Roberta	5	8	2e-5	

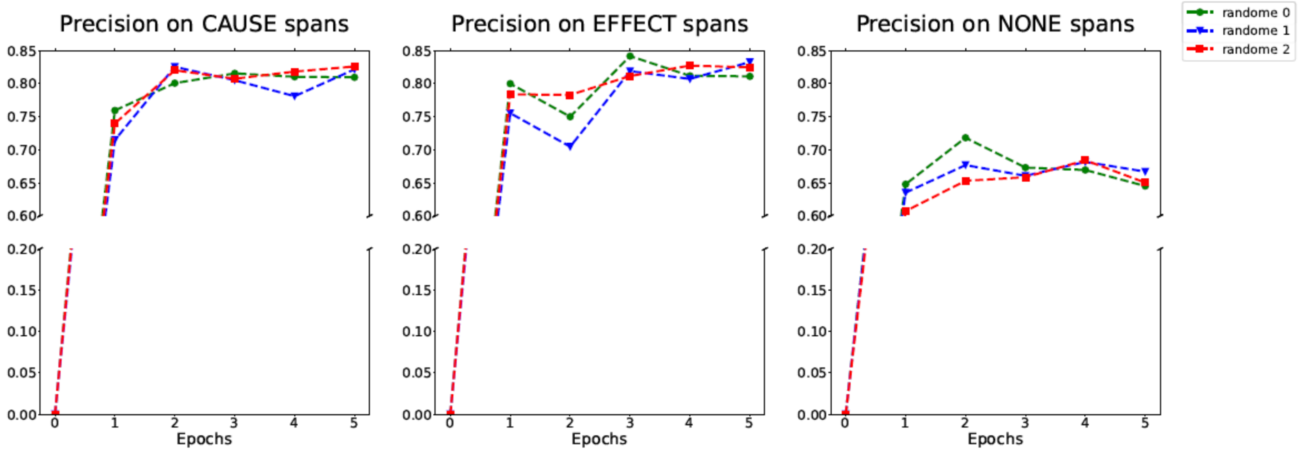
Xu. et al., A Framework to Construct Financial Causality Knowledge Graph from Text, IEEE-ICSC 2024

3. FinCaKG Construction Framework - Experiment Results





(a) The overall evaluation for all identified textual spans.



(b) The independent evaluation of CAUSE, EFFECT and NONE spans.

Fig. 2: The performance of cause-effect span identification model in $\boxed{\text{Step2.}}$

TABLE III: The performance of causal sentences extraction model in $\boxed{\text{Step1}}$

	Dataset	Precision	Recall	F1-score
Validation	Labeld Data (R1)	96.31%	96.07%	96.19%
Prediction	Financial Reports (R0)	95.93%	95.89%	95.91%

TABLE IV: The performance of financial keywords prediction model in $\boxed{\text{Step3}}$

10-folds	Precision	Recall	F1-score
avg.	98.72%	98.68%	98.70%
min.	98.10%	98.14%	98.12%
max.	99.07%	99.05%	99.06%

Xu. et al., A Framework to Construct Financial Causality Knowledge Graph from Text, IEEE-ICSC 2024

4. Presentation of Variant FinCaKGs from different resources

- **FinCaKG-FR**: The **F**inancial **R**eports, the original input = R0
- **FinCaKG-ECT**: The **E**arning **C**all **T**ranscripts are textual records of conference calls, including questions from analysts and answers from company representatives.
- **FinCaKG-AR**: The **A**nalyst **R**eports refer to documents produced by financial analysts who study and analyze publicly traded companies.

TABLE V: The statistical comparison between distinct versions of FinCaKGs.

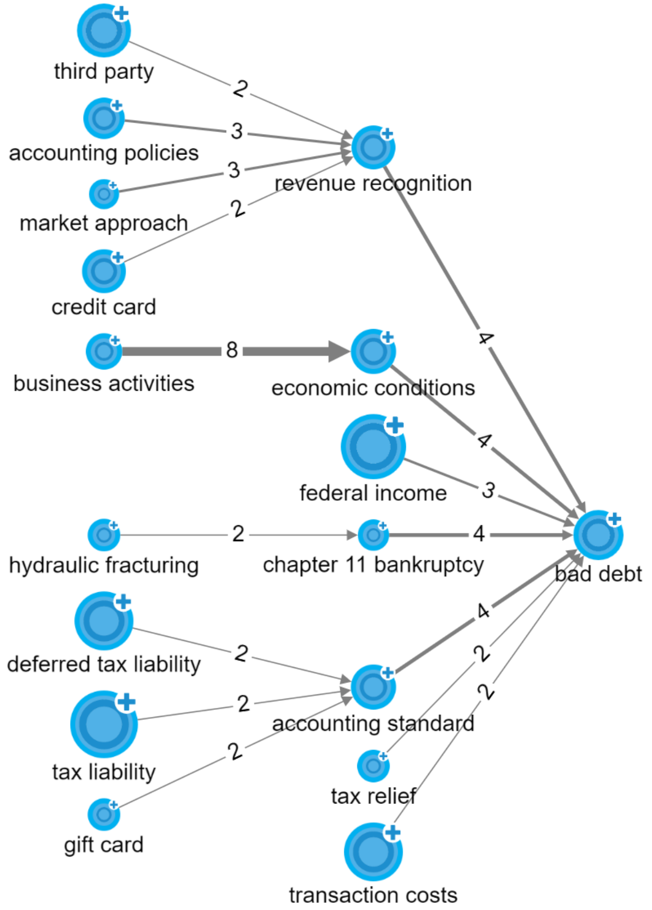
	FinCaKG-FR	FinCaKG-ECT	FinCaKG-AR
# nodes	1,717	546	199
# rels	11,633	1,802	283
# docs	5,093	5,547	10,057

Xu. et al., A Framework to Construct Financial Causality Knowledge Graph from Text, IEEE-ICSC 2024

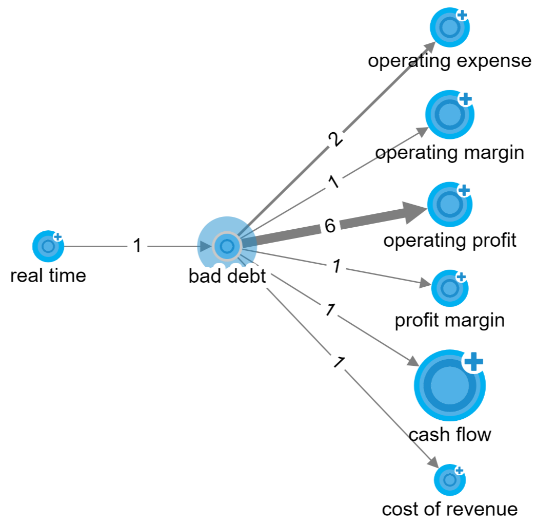
4. Presentation of Variant FinCaKGs from different resources

- Anchor Term: bad debt

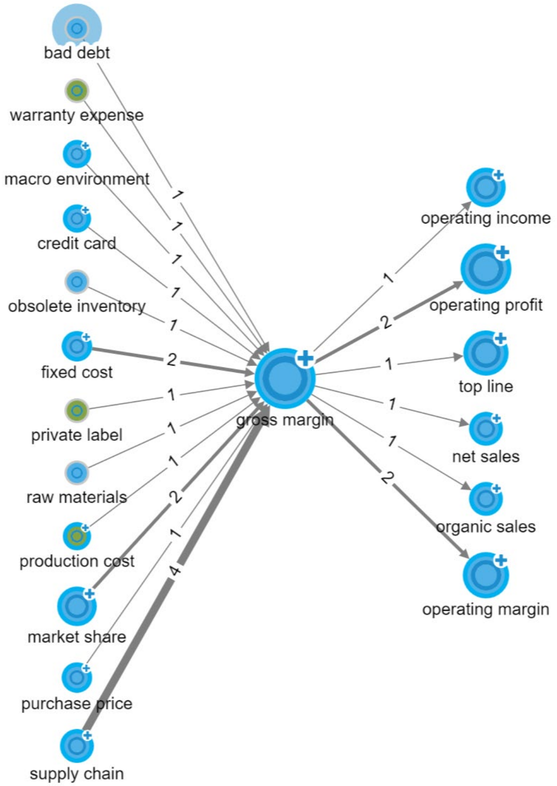
	FinCaKG-FR	FinCaKG-ECT	FinCaKG-AR
# nodes	1,717	546	199
# rels	11,633	1,802	283
# docs	5,093	5,547	10,057



FinCaKG-FR
(Financial Reports)



FinCaKG-ECT
(Earning Call Transcripts)



FinCaKG-AR
(Analyst Reports)

Xu. et al., A Framework to Construct Financial Causality Knowledge Graph from Text, IEEE-ICSC 2024

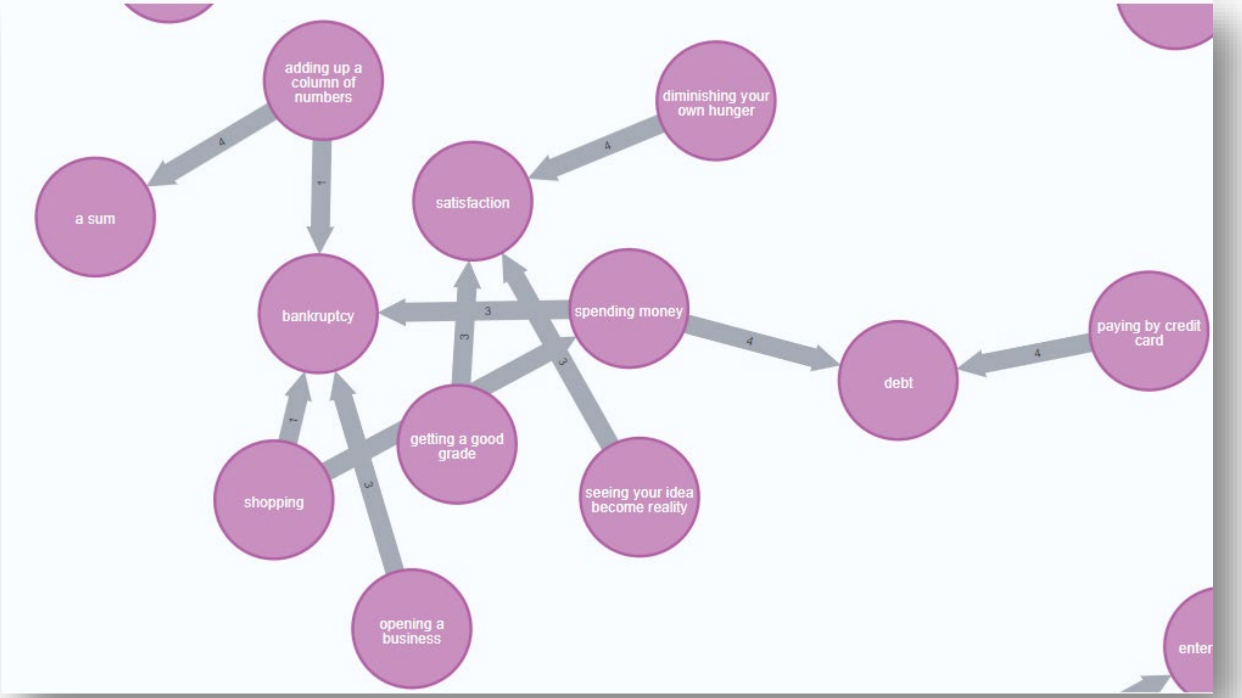
4. Variant FinCaKGs v.s. ConceptNet-Cause

TABLE V: The statistical comparison between distinct versions of FinCaKGs.

	FinCaKG-FR	FinCaKG-ECT	FinCaKG-AR	ConceptNet-Cause[11]
# nodes	1,717	546	199	12,832
# rels	11,633	1,802	283	16,801
# docs	5,093	5,547	10,057	-

TABLE VI: The investigation on connectivity and domain coverage of FinCaKGs and ConceptNet-Cause.

Causal KGs	Density	Node Degree	Domain Coverage
ConceptNet-Cause	0.01%	2.62	0.72%
FinCaKG-FR	0.39%	13.55	60.51%
FinCaKG-ECT	0.61%	6.60	83.88%
FinCaKG-AR	0.72%	2.84	83.42%



ConceptNet-Cause

Xu. et al., A Framework to Construct Financial Causality Knowledge Graph from Text, IEEE-ICSC 2024

5. Future Work

Downstream Tasks of FinCaKGs

- Finance domain research questions
  - Causal analysis generation of financial reports
  - Managers' characteristic depiction according to its causal chains

Temporal FinCaKG Generation

- Technical research questions
  - temporal link prediction
  - temporal node prediction

6. Conclusion

- We presented a framework to **automatically** construct the causality knowledge graph from financial reports. The results showed this framework can capture **in-dense and domain-related causality**.

- This framework is capable to generate distinct versions of FinCaKGs with different resources and **uncover the different inner logics**.

- FinCaKG could support the expertise discovery **directly from the visualization of FinCaKGs**.

Limitations:

- We only focus on the finance-related concepts in FinCaKGs, it could **not** cover some **news events**, i.e. Covid eases, energy shock …

- The causal chains might be **less informative** along with **larger hops** in causality

Thanks for your listening and welcome for any questions!