

# Commodities trend link prediction on heterogeneous information networks

P. do Carmo<sup>\*,1</sup>, I. J. Reis Filho<sup>2,1</sup>, R. Marcacini<sup>1</sup>

<sup>1</sup> University of São Paulo - São Carlos, Brazil  
paulo.carmo@usp.br, ricardo.marcacini@icmc.usp.br

<sup>2</sup> State University of Minas Gerais - Frutal, Brazil  
ivan.filho@uemg.br

**Abstract.** Events can be defined as an action or a series of actions that have a determined theme, time, and place. Event analysis tasks for knowledge extraction from news and social media have been explored in recent years. However, there are still few studies that aim to enrich predictive models using event data. In particular, agribusiness events have multiple components to be considered for a successful prediction model. For example, price trend predictions for commodities can be performed through time series analysis of prices, but we can also consider events that represent knowledge about external factors during the training step of predictive models. In this paper, we present a method for integrating events into trend prediction tasks. First, we propose to model events and time-series information through heterogeneous information networks (HIN) that allow multiple components to be directly modeled through multi-type nodes and edges. Second, we learn features from HIN through network embedding methods, i.e., network nodes are mapped to a dense vector of features. In particular, we propose a network embedding method that propagates the semantic of the pre-trained neural language models to a heterogeneous information network and evaluates its performance in a trend link prediction. We show that the use of our proposed model language-based embedding propagation is competitive with state-of-art network embeddings algorithms. Moreover, our proposal performs network embedding incrementally, thereby allowing new events to be inserted in the same semantic space without rebuilding the entire network embedding.

CCS Concepts: • **Information systems**; • **Computing methodologies** → **Machine learning**; • **Applied computing** → **Document management and text processing**;

Keywords: agribusiness, event analysis, heterogeneous networks, network embedding, text mining

## 1. INTRODUCTION

Events can be defined as an action or a series of actions that occur at a specific time and place [Allan 2012; Cordeiro and Gama 2016; Chen and Li 2020]. Different types of event analysis can be used to allow real-time monitoring for a number of application domains like the economy, agribusiness, epidemics, medicine, sentiment analysis, and many other social behavior studies [Marcacini et al. 2017]. Agribusiness events are challenging to predict since they depend on (i) climate changes, (ii) historical data from the market, (iii) supply and demand, (iv) aggregate demand, and (v) politics. Climate changes are predictable through meteorologic data. Market history, as well as supply and demand data, are obtained from temporal series. However, aggregate demand and politics are also available in text data from news or social networks. Since textual data is made for humans, text mining is a challenging task that requires multiple steps of processing [Venter et al. 2013].

Event analysis is the computational task for automatically identifying related events through their text and other data such as timestamps, places, people, and entities involved [Hamborg et al. 2018; Xue et al. 2019]. The data extracted comes in 5W1H model which is described as [Chen and Li 2020]:

---

\*This work was partially funded by Capes

Copyright©2021. Permission to copy without fee all or part of the material printed in KDMiLe is granted provided that the copies are not made or distributed for commercial advantage, and that notice is given that copying is by permission of the Sociedade Brasileira de Computação.

**what** happened; **when** it happened; **where** it happened; **who** is involved; **why** it happened; and **how** it happened. Thus, these techniques can be used to explore useful agribusiness data for trend prediction [Radinsky and Horvitz 2013; Ning et al. 2019], since they can be modeled as events. Event analysis is a challenging task since different topics can be related at some level. For example, the corn value can drop due to new exporting politics that lower international demand, and if identified earlier, a producer might be able to close a better domestic deal [dos Reis Filho et al. 2020].

Recently, heterogeneous networks have been used successfully for modeling large event datasets [Shi et al. 2016], since they model different components from events as nodes (e.g., when, where, who, why and how), and network links express different relationships between these nodes. Thus, we can create a heterogeneous network, where each event is represented by itself and a set of components that complete the 5W1H. The use of heterogeneous networks also allows link prediction tasks. It represents the fundamental problem of estimating a connection between two nodes based on observed links and node features. Link prediction is a simple binary classification problem [Shi et al. 2016]. Whenever a network node can be used as a label, link prediction can also be modeled as a multi-class classification problem. Link prediction as a classification uses node features as data. There are many techniques for obtaining node features from a network, allowing different implementations of link prediction and other techniques.

Network embeddings methods map each node into a low dimensional vector that represents the network topology and node type information [Chang et al. 2015; Huang and Mamouli 2017; Setty and Hose 2018; Cui et al. 2018; Wu et al. 2020]. The recent literature has explored several methods to solve this task, from random walks within the entire network to deep learning methods [Wu et al. 2020]. However, since events can be partially represented as texts, such as news headlines or social networks posts, we can use neural language models, such as BERT [Devlin et al. 2018], to compute meaningful representations. Neural language models are trained over large textual datasets, including news headlines, so they have valuable general-purpose knowledge for event prediction. However, the use of just text embedding is not recommended since it ignores essential component information [Chen and Li 2020], which may lead to deeper connections between events, like location and actors involved.

This paper presents a framework to enable price trend analysis on heterogeneous information networks through link prediction. We present a method called Trend Prediction on Heterogeneous Information Networks (TPHIN), a language model-based embedding propagation method for heterogeneous event networks. While most existing network embedding methods mainly explore the network’s topology, our method maps both (i) textual information about events and (ii) the complex relationships between events and their components to a low dimensional vector space. Existing methods of network embeddings are offline, which requires repeating the entire process when new nodes are added to the network. Some pipelines may include different techniques for dynamic insertions [Deng et al. 2019; 2020], but they must be modeled apart from the network embedding. On the other hand, TPHIN is naturally incremental, as it takes advantage of a pre-trained feature semantic space from neural language models by propagating from its initial embeddings to the entire network.

We evaluate the proposed method with other information network embeddings methods on an agribusiness news dataset. We consider corn and soybean prices for trend prediction. We also demonstrate how enrolling the general knowledge of a pre-trained neural language model into a heterogeneous event information network for link prediction scenarios performs through quantitative metrics. We show that the use of TPHIN is competitive with state-of-art network embeddings algorithms. Moreover, our proposal performs network embedding incrementally, thereby allowing new events to be inserted in the same semantic space without rebuilding the entire network embedding.

## 2. RELATED WORK

Commodities trend prediction is usually based on time series analysis techniques like Integrated Au-

autoregressive Moving Average (ARIMA, [Darekar and Reddy 2017]) and Integrated Seasonal Autoregressive Moving Average (SARIMA [Adanacioglu et al. 2012]). With the advance of text mining techniques, some works began to combine these text features with time series data for stock and commodity prediction [Wang et al. 2019; Chen et al. 2016]. In dos Reis Filho et al. [2020] the authors combine text and time-series data. The text was obtained from a bag-of-words model and concatenated with a decision tree model based on time series data. They evaluate the models by comparing the results from time series and the proposed model in a Support Vector Regression model for commodity price prediction.

Network embedding methods allow for node feature extraction on local and global levels. From Perozzi et al. [2014], the DeepWalk is a Word2Vec based information network embedding method that executes multiple random walks in order to sample data for skip-gram model training. Node2Vec from Grover and Leskovec [2016] is an extension of the DeepWalk method that allows the user to choose between two parameters  $p$  and  $q$  that control how much of the random walk will follow breadth or depth search, respectively. Metapath2Vec was proposed by Dong et al. [2017] and is another extension of the DeepWalk that was designed for heterogeneous networks. It transforms the random walks into metapath-biased walks, i.e., the user defines a set of node types the algorithm should follow in order to explore the heterogeneous network. Struc2Vec [Ribeiro et al. 2017] aims to generate a representation that is aware of the network structure by constructing a multi-layer graph to encode the structural context through the hierarchy of paths and then applies random walks like DeepWalk. LINE was proposed by Tang et al. [2015], a network embedding method to very large information networks by edge sampling walks for local and global neighbors separately and unifying them after through negative sampling. The authors Kipf and Welling [2017] proposed the Graph Convolutional Network (GCN) method based on the assumption that the 1-hop connections between nodes might not always indicate they have the same label, so they modeled a neural network model to learn the graph correlations within some labeled nodes and then propagate these relations throughout the network.

In Deng et al. [2019] the authors developed a dynamic vocabulary graph and modeled a GCN to identify key events and understand their evolution. Their proposal was evaluated within protest events datasets. Models like these can also be used for a number of forecasting problems, such as the multi-event and multi-actor forecasting tackled by Deng et al. [2020] with dynamic knowledge graphs and graph completion techniques, which are derived from link prediction. Even though these methods successfully introduced structures for multi-typed events, they lack general information that we believe can be obtained through the propagation of pre-trained neural embeddings on top of an event HIN architecture.

### 3. TREND PREDICTION ON HETEROGENEOUS INFORMATION NETWORKS

Events extracted from news data act as digital sensors, as they are published around the time it happened and contain information surrounding it. This information can be extracted directly from the text and its metadata [Hamborg et al. 2018]. We chose to extract a 4W1H variation without the *when* characteristics since it is not as good data as the publication date. We also consider metadata from commodities prices to generate the trend indicators of the temporal series. The trend labels are calculated according to the  $[week/month]/year$  period and commodity they represent. With the news features extracted and trend labels calculated, we can model different heterogeneous networks to each combination (Figure 1).

A heterogeneous event network is formally defined as a triple  $N = (O, R, W)$ , where  $O$  is the set of object nodes,  $R$  is the set of connections between objects, and  $W$  is the weight of those connections [dos Santos et al. 2020]. Heterogeneous event networks are information networks that have the set  $O > 1$ , meaning that they have more than one object node type. This allows the representation to adapt itself to various real-world data and event analysis, in particular events represented by multiple components, such as places, time, name of people, and organizations. In this case, the set of objects is

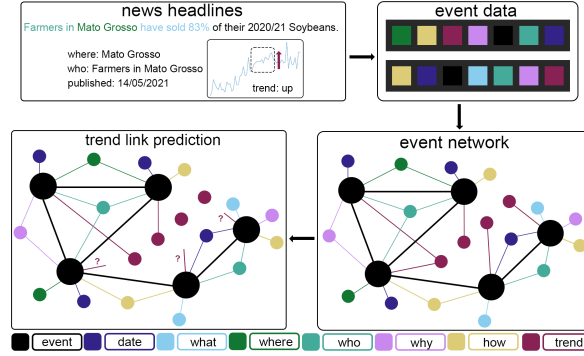


Fig. 1. Visual representation of the proposal. The event components are extracted from news headlines and metadata, while the trend symbolizes the trend of a commodity at the end of the period the news was published on.

defined as  $O = \{O_e \cup O_d \cup O_w \cup O_l \cup O_a \cup O_y \cup O_h \cup O_t\}$ , where the subset  $O_e$  represents event nodes,  $O_d$  represents date nodes,  $O_w$  represents what nodes,  $O_l$  represents location nodes,  $O_a$  represents actors nodes (e.g. people or organizations),  $O_y$  represents why nodes,  $O_h$  represents how nodes and  $O_t$  represents trend nodes (our desired labels).

Network embeddings methods map each node in a low dimensional vector that represents network topology and node type information [Chang et al. 2015; Huang and Mamoulis 2017; Setty and Hose 2018; Cui et al. 2018; Wu et al. 2020]. Formally, network embedding generates a function  $g : N \rightarrow \mathbb{R}^d$ , where the nodes  $N$  are represented in a  $d$ -dimensional space. Each node  $n \in N$  has an embedding  $\mathbf{f}_n \in \mathbb{R}^d$ .

Our paper investigates network embedding through embedding propagation [Yang et al. 2019], where we can take advantage of initial event embeddings using a BERT based pre-trained neural language model. The general problem for embedding propagation proposed in this paper is presented in Equation 1, inspired by a general graph regularization framework [Ji et al. 2010]. It can be solved via minimization with quadratic programming or through iterative methods based on label propagation [Belkin et al. 2006; Zhu et al. 2003]. The first term determines that neighboring nodes  $n_i$  and  $n_j$  in the network have similar embedding vectors  $\mathbf{f}_{n_i}$  and  $\mathbf{f}_{n_j}$ , where  $\Omega$  is a distance function, and  $w_{n_i, n_j}$  indicates the weight of the connection between the nodes  $n_i$  and  $n_j$ . In the second term, the objective is that the initial embeddings  $\mathbf{g}_{n_i}$  of the event nodes  $O_e$  are preserved according to a  $\mu$  factor (with  $\mu > 0$ ). In this case,  $\mathbf{g}_{n_i} \in \mathbb{R}^d$  are  $d$ -dimensional embeddings obtained by pre-trained language models from the texts of the event node  $n_i$ , where  $n_i \in O_e$ .

$$Q(\mathbf{F}) = \frac{1}{2} \sum_{n_i, n_j \in N} w_{n_i, n_j} \Omega(\mathbf{f}_{n_i}, \mathbf{f}_{n_j}) + \mu \sum_{O(n_i) \in O_e} \Omega(\mathbf{f}_{n_i}, \mathbf{g}_{n_i}) \quad (1)$$

We argue that using the initial BERT embeddings strategy for events is potentially promising for network event analysis. In particular, trend prediction is a challenging repeatable task. The use of a pre-trained embedding allows new nodes to be added and used in the network immediately after an incremental embedding propagation. The embedding propagation maintains the same original semantic space of the BERT model. In contrast, other implementations of network embedding may require recreating the embeddings for the whole network.

With the network embeddings calculated, we must apply the data collected to a machine learning algorithm. In this paper, we've chosen a long-short term memory (LSTM) for supervised multi-class classification with the network embedding data. The LSTM we used accommodates memory units. The memory units are composed of three smaller units ( $u, c_{in}, c_{out}$ ). The units  $u$  (Equation 2 (i)) apply a weighted sum, where  $w$  represents the weight, for each value  $y$ , and its result is outputted

through an activation function. These units are limited by control units  $c_{in}$  (Equation 2 (ii)) and  $c_{out}$  (Equation 2 (iii)) that control when the iteration  $t$  will go forward or keep recurring. The initial value  $y$  inserted to the LSTM comes from the propagated embeddings  $f$ , and changes throughout training.

$$\begin{aligned} (i) \quad u_u^{(t)} &= \sum_u w_u y^{u(t-1)} \\ (ii) \quad c_{in}^{(t)} &= \sum_u w_{in_j u} y^{u(t-1)} \\ (iii) \quad c_{out}^{(t)} &= \sum_u w_{out_j u} y^{u(t-1)} \end{aligned} \quad (2)$$

These units must learn the correct weights to predict trends from an embedding vector. Thus the LSTM is trained by a process called back-propagation. A back-propagation optimizer uses a gradient (e.g. Equation 3) that calculates the error  $E$  between the actual label from a training to the time step  $T'$ ,  $T$  and generates a  $\Delta$  to update each weight  $w_{ij}$  from the LSTM network.

$$\Delta w^{E^{total}}(T', T) = \sum_{t=T'+1}^T \Delta w^{E(t)} \rightarrow \Delta w_{ij} = -\alpha \frac{\partial E^{total}(T', T)}{\partial w_{ij}} \quad (3)$$

The LSTM used in this proposal has 256 memory units and uses the Adam optimizer. This number was chosen to accommodate a 512 dimension embedding. The last layer for the LSTM calculates the probabilities of each trend in a SoftMax [Goodfellow et al. 2016] function. Thus our proposal can be summarized as a network embedding propagation from BERT-based embeddings that ingests them in the LSTM multi-class classification pipeline for trend link prediction.

## 4. EXPERIMENTAL EVALUATION

### 4.1 Datasets

To evaluate the trend prediction within an agribusiness real-world-scenario we use a dataset with news related to the corn and soybean commodities, extracted from Soybean & Corn Advisor<sup>1</sup> as well as the historical prices from CEPEA (*Centro de Estudos Avançados em Economia Aplicada*<sup>2</sup>). This dataset allows us to extract the events and components from the news text, and the trends are calculated in four different ways, which formulates four networks as shown in Table I. Each network has the same events and components but generates trends' labels nodes by different commodities and time windows. In addition, we use the DistilBERT-Multilingual<sup>3</sup> model to generate the initial embedding for each event headline text.

Table I. Overview of heterogeneous event networks used in the experimental evaluation.

Network	#Nodes	#Events	#Dates	#Whats	#Wheres	#Whos	#Whys	#Hows	Time window	Commodity
#1	4348	2322	380	48	115	9	244	1226	weekly	corn
#2	4056	2322	89	48	115	9	244	1226	monthly	corn
#3	4348	2322	380	48	115	9	244	1226	weekly	soybean
#4	4056	2322	89	48	115	9	244	1226	monthly	soybean

### 4.2 Evaluation criteria and experiment setup

We use classic metrics to evaluate the multi-class classification scenario: accuracy, macro precision, macro recall and macro  $F1$  [Powers 2008]. They were chosen to provide a standardized framework for

<sup>1</sup>Available at <http://soybeansandcorn.com>

<sup>2</sup>Available at <https://www.cepea.org.br>

<sup>3</sup>Available at <https://github.com/UKPLab/sentence-transformers>

future comparisons and demonstrate the differences between algorithms and test cases from different points of view, accounting for imbalances between classes.

We compare TPHIN with state-of-art network embeddings methods: DeepWalk, Node2Vec, Metapath2Vec, Struc2Vec, LINE, and GCN. We use the parameters recommended by the authors in the respective original papers for each baseline method. Regarding the number of dimensions of the embeddings, we used 512 for all methods since it is the dimension used by the DistilBERT language model. All experimental data, source code, and networks are available at the GitHub repository <https://github.com/PauloRVdC/tphin-experiments>.

### 4.3 Results and Discussion

We predicted links on two different time distances for all four networks, comparing it to six baseline methods to position the TPHIN method against the state-of-art unsupervised (and the GCN, which is semi-supervised) network embeddings methods on the literature for a real-world scenario. Table II breaks down the results for a time window of three months or weeks with three different metrics.

Table II. Trend link prediction performance on four metrics (accuracy (acc), precision (pre), recall (rcl) and  $F1$ ) for a time window of three weeks or months. Bold values are the highest.

	3 weeks/months															
	Network #1				Network #2				Network #3				Network #4			
	acc	pre	rcl	$F1$	acc	pre	rcl	$F1$	acc	pre	rcl	$F1$	acc	pre	rcl	$F1$
DeepWalk	0.48	0.48	0.47	0.45	0.61	<b>0.50</b>	0.30	0.37	0.55	0.57	0.57	0.52	0.58	<b>0.50</b>	<b>0.29</b>	<b>0.36</b>
Node2Vec	0.65	0.71	0.64	0.62	0.63	<b>0.50</b>	0.31	0.32	0.48	0.57	0.56	0.46	0.48	0.48	0.23	0.31
Metapath2Vec	0.43	0.32	0.27	0.28	<b>0.69</b>	<b>0.50</b>	<b>0.35</b>	<b>0.40</b>	0.45	0.35	0.44	0.33	0.30	0.33	0.10	0.15
Struc2Vec	0.56	0.46	0.43	0.43	0.26	<b>0.50</b>	0.13	0.20	0.32	0.37	0.19	0.23	0.32	0.33	0.11	0.15
LINE	0.56	0.58	0.56	0.55	0.55	<b>0.50</b>	0.27	0.35	0.48	0.47	0.47	0.41	0.52	<b>0.50</b>	0.26	0.34
GCN	0.48	0.56	0.49	0.45	0.61	<b>0.50</b>	0.31	0.37	<b>0.63</b>	<b>0.59</b>	<b>0.58</b>	<b>0.54</b>	0.27	<b>0.50</b>	0.14	0.20
TPHIN	<b>0.75</b>	<b>0.77</b>	<b>0.75</b>	<b>0.75</b>	0.61	<b>0.50</b>	0.31	0.38	0.38	0.39	0.34	0.30	<b>0.60</b>	0.33	0.20	0.25

The results on week trend prediction were more consistent, which can be explained by having more consistent news and market changes in a shorter span of time. We can infer the news was better organized by the good results on Network #1 where the embedding propagation of TPHIN was able to outperform the baseline methods in every metric. On Network #3 the semi-supervised approach from GCN outperformed every method, which infers that soybean’s news might not bring the same level of information in a three-week span as corn’s. This indicates that GCN can learn the events relationships by looking at the existing labels and the topology extracted more consistently from the text than TPHIN. However, TPHIN is a unsupervised method, so it is expected. In Table III our assumptions can once again be observed.

Table III. Trend link prediction performance on four metrics (accuracy (acc), precision (pre), recall (rcl) and  $F1$ ) for a time window of six weeks or months. Bold values are the highest.

	6 weeks/months															
	Network #1				Network #2				Network #3				Network #4			
	acc	pre	rcl	$F1$	acc	pre	rcl	$F1$	acc	pre	rcl	$F1$	acc	pre	rcl	$F1$
DeepWalk	0.51	0.53	0.54	0.50	0.56	0.48	0.48	0.48	0.52	0.53	0.53	0.52	0.37	0.24	0.29	0.26
Node2Vec	0.45	0.45	0.45	0.43	0.57	0.49	0.49	0.49	<b>0.54</b>	0.54	<b>0.54</b>	<b>0.53</b>	0.41	0.26	0.31	0.28
Metapath2Vec	0.48	0.37	0.31	0.33	0.42	0.42	0.42	0.38	0.47	0.44	0.43	0.40	0.35	23	0.30	0.25
Struc2Vec	0.43	0.32	0.30	0.30	0.52	0.55	0.56	<b>0.51</b>	0.21	0.19	0.11	0.13	<b>0.46</b>	<b>0.30</b>	<b>0.37</b>	<b>0.33</b>
LINE	0.49	0.50	0.51	0.48	0.50	0.50	0.50	0.48	0.48	0.36	0.34	0.34	0.42	0.28	0.33	0.30
GCN	0.37	0.44	0.45	0.33	<b>0.69</b>	<b>0.59</b>	<b>0.54</b>	<b>0.51</b>	0.48	<b>0.57</b>	0.50	0.40	0.28	0.18	0.24	0.20
TPHIN	<b>0.56</b>	<b>0.56</b>	<b>0.55</b>	<b>0.54</b>	0.60	0.51	0.51	0.50	0.51	0.37	0.37	0.39	0.44	0.25	0.31	0.27

For Network #1, TPHIN outperforms all baselines as well. This shows that corn’s news are more easily processed by the DistilBERT method. With Table III we can also see TPHIN has a good performance on network #2 as well. Even though it does not achieve the best results on a monthly

trend, it gets closest to the best on some metrics. We can infer that networks constructed focusing on the corn's price trend are more suitable than soybeans. Soybean's networks #3 and #4 achieve the best results with the topological network embeddings methods, indicating that most of the valuable information was encoded on network topology.

These results show that TPHIN is a competitive method for network embedding in a trend prediction application, especially if the text database is well-curated. It also works best on shorter spans of time, which encourages its use in a real-world scenario. For example, every time a news article from a reputable source comes out, an agribusiness stakeholder can ingest it and generate a single week or month price trend prediction. Maintaining the same semantical space as a pre-trained language model allows it to insert the new nodes and iterate upon the existing network to update embeddings. TPHIN also enables natural languages queries on the network since it's based on language models, which enhances the experience for real-world use.

## 5. CONCLUSION

This paper introduces a network embedding method for link prediction to trend price nodes. It explores event heterogeneous network modeling, embedding propagation, and link prediction through LSTM neural networks for multi-class classification. Using an embedding propagation technique from a BERT-based model allows network embedding without recalculating the entire network. We also show that the use of text information, combined with simple network topology, is competitive against state-of-art topology network embeddings algorithms when the text database is well-curated.

We plan to incorporate weights on the different types of relations for future work, thereby allowing the embedding propagation to consider more topology information. We also want to investigate semi-supervised embedding propagation methods to incorporate the existing labels into the resulting embeddings. Attention mechanisms is another scenario we can evolve this work. Finally, we plan to use language models trained over agribusiness textual corpus instead of general-purpose models.

**Acknowledgments:** This work was supported by National Council for Scientific and Technological Development (CNPq) [process number 426663/2018-7], The São Paulo Research Foundation (FAPESP) [process number 2019/25010-5 and 2019/07665-4] and The Coordination for the Improvement of Higher Education Personnel (CAPES) [process number 88887.513429/2020-00].

## REFERENCES

- ADANACIOGLU, H., YERCAN, M., ET AL. An analysis of tomato prices at wholesale level in turkey: an application of sarima model. *Custos e Agronegócio Online* 8 (4): 52–75, 2012.
- ALLAN, J. *Topic detection and tracking: event-based information organization*. Vol. 12. Springer Science & Business Media, 2012.
- BELKIN, M., NIYOGI, P., AND SINDHWANI, V. Manifold regularization: A geometric framework for learning from labeled and unlabeled examples. *Journal of Machine Learning Research* vol. 7, pp. 2399–2434, 2006.
- CHANG, S., HAN, W., TANG, J., QI, G.-J., AGGARWAL, C. C., AND HUANG, T. S. Heterogeneous network embedding via deep architectures. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. pp. 119–128, 2015.
- CHEN, H.-H., CHEN, M., AND CHIU, C.-C. The integration of artificial neural networks and text mining to forecast gold futures prices. *Communications in Statistics - Simulation and Computation* 45 (4): 1213–1225, 2016.
- CHEN, X. AND LI, Q. Event modeling and mining: a long journey toward explainable events. *The VLDB Journal* 29 (1): 459–482, 2020.
- CORDEIRO, M. AND GAMA, J. Online social networks event detection: a survey. In *Solving Large Scale Learning Tasks. Challenges and Algorithms*. Springer, pp. 1–41, 2016.
- CUI, P., WANG, X., PEI, J., AND ZHU, W. A survey on network embedding. *IEEE Transactions on Knowledge and Data Engineering* 31 (5): 833–852, 2018.
- DAREKAR, A. AND REDDY, A. Predicting market price of soybean in major india studies through arima model. *Journal of Food Legumes* 30 (2): 73–76, 2017.

- DENG, S., RANGWALA, H., AND NING, Y. Learning dynamic context graphs for predicting social events. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. pp. 1007–1016, 2019.
- DENG, S., RANGWALA, H., AND NING, Y. Dynamic knowledge graph based multi-event forecasting. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. pp. 1585–1595, 2020.
- DEVLIN, J., CHANG, M.-W., LEE, K., AND TOUTANOVA, K. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018.
- DONG, Y., CHAWLA, N. V., AND SWAMI, A. metapath2vec: Scalable representation learning for heterogeneous networks. In *ACM SIGKDD international conference on knowledge discovery and data mining*. pp. 135–144, 2017.
- DOS REIS FILHO, I. J., CORREA, G. B., FREIRE, G. M., AND REZENDE, S. O. Forecasting future corn and soybean prices: an analysis of the use of textual information to enrich time-series. In *Anais do VIII Symposium on Knowledge Discovery, Mining and Learning*. SBC, pp. 113–120, 2020.
- DOS SANTOS, B. N., ROSSI, R. G., REZENDE, S. O., AND MARCACINI, R. M. A two-stage regularization framework for heterogeneous event networks. *Pattern Recognition Letters* vol. 138, pp. 490–496, 2020.
- GOODFELLOW, I., BENGIO, Y., AND COURVILLE, A. 6.2. 2.3 softmax units for multinoulli output distributions. *Deep learning* (1): 180, 2016.
- GROVER, A. AND LESKOVEC, J. node2vec: Scalable feature learning for networks. In *Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining*. pp. 855–864, 2016.
- HAMBORG, F., LACHNIT, S., SCHUBOTZ, M., HEPP, T., AND GIPP, B. Giveme5w: main event retrieval from news articles by extraction of the five journalistic w questions. In *International Conference on Information*. Springer, pp. 356–366, 2018.
- HUANG, Z. AND MAMOULIS, N. Heterogeneous information network embedding for meta path based proximity. *arXiv preprint arXiv:1701.05291*, 2017.
- JI, M., SUN, Y., DANILEVSKY, M., HAN, J., AND GAO, J. Graph regularized transductive classification on heterogeneous information networks. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. Springer, pp. 570–586, 2010.
- KIPF, T. N. AND WELING, M. Semi-supervised classification with graph convolutional networks. In *International Conference on Learning Representations (ICLR)*, 2017.
- MARCACINI, R. M., ROSSI, R. G., NOGUEIRA, B. M., MARTINS, L. V., CHERMAN, E. A., AND REZENDE, S. O. Websensors analytics: Learning to sense the real world using web news events. In *Simp. Brasileiro de Sistemas Multimídia e Web*. pp. 169–173, 2017.
- NING, Y., ZHAO, L., CHEN, F., LU, C.-T., AND RANGWALA, H. Spatio-temporal event forecasting and precursor identification. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. pp. 3237–3238, 2019.
- PEROZZI, B., AL-RFOU, R., AND SKIENA, S. Deepwalk: Online learning of social representations. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*. pp. 701–710, 2014.
- POWERS, D. Evaluation: From precision, recall and f-factor to roc, informedness, markedness & correlation. *Mach. Learn. Technol.* vol. 2, 01, 2008.
- RADINSKY, K. AND HORVITZ, E. Mining the web to predict future events. In *Proceedings of the sixth ACM international conference on Web search and data mining*. pp. 255–264, 2013.
- RIBEIRO, L. F., SAVERESE, P. H., AND FIGUEIREDO, D. R. struc2vec: Learning node representations from structural identity. In *ACM SIGKDD international conference on knowledge discovery and data mining*. pp. 385–394, 2017.
- SETTY, V. AND HOSE, K. Event2vec: Neural embeddings for news events. In *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*. pp. 1013–1016, 2018.
- SHI, C., LI, Y., ZHANG, J., SUN, Y., AND PHILIP, S. Y. A survey of heterogeneous information network analysis. *IEEE Transactions on Knowledge and Data Engineering* 29 (1): 17–37, 2016.
- TANG, J., QU, M., WANG, M., ZHANG, M., YAN, J., AND MEI, Q. Line: Large-scale information network embedding. In *Proceedings of the 24th international conference on world wide web*. pp. 1067–1077, 2015.
- VENTER, M., STRYDOM, D., AND GROVÉ, B. Stochastic efficiency analysis of alternative basic grain marketing strategies. *Agrekon* 52 (sup1): 46–63, 2013.
- WANG, J., WANG, Z., LI, X., AND ZHOU, H. Artificial bee colony-based combination approach to forecasting agricultural commodity prices. *International Journal of Forecasting*, 2019.
- WU, Z., PAN, S., CHEN, F., LONG, G., ZHANG, C., AND PHILIP, S. Y. A comprehensive survey on graph neural networks. *IEEE Transactions on Neural Networks and Learning Systems*, 2020.
- XUE, F., HONG, R., HE, X., WANG, J., QIAN, S., AND XU, C. Knowledge based topic model for multi-modal social event analysis. *IEEE Transactions on Multimedia*, 2019.
- YANG, C., ZHANG, J., AND HAN, J. Neural embedding propagation on heterogeneous networks. In *2019 IEEE International Conference on Data Mining (ICDM)*. IEEE, pp. 698–707, 2019.
- ZHU, X., GHAHRAMANI, Z., AND LAFFERTY, J. D. Semi-supervised learning using gaussian fields and harmonic functions. In *Proceedings of the 20th International conference on Machine learning (ICML-03)*. pp. 912–919, 2003.