Journal of Advanced Artificial Intelligence
Volume 1 – No.6, March 2025

Investigating Knowledge Graphs for Identifying the Scope of Advancements in Knowledge Graph Embedding and Reasoning Techniques

Mosam Patel
Assistant Professor,
GSFC University,
Vadodara, India

ABSTRACT
Enormous amount of data in varied formats is generated daily through various sources. Knowledge Graphs has got the amazing capability to fetch the information and efficiently organize it so as to explore the linkages of the Entities. Hence, ever since the Knowledge Graphs (KG) were introduced by Google in 2012, its popularity has increased in almost all fields owing to its promising performance. Many existing techniques are being replaced by Knowledge Graphs, taking into consideration, its efficient ways to extract the required information by interlinking the nodes. Though the researchers are working on generating a generic KG, so that it can cater more applications, still there are many challenges which the researchers are facing in achieving it. This paper focuses on investigating the Knowledge Graph Embedding, and Reasoning Techniques. The benefits and challenges of various methods for dealing with Embedding and Reasoning is being investigated and presented in the paper. At the end, to summarize the study, the various scopes of advancements in Knowledge Graph are also presented in brief. This investigation will help the researchers in paving different paths of application and research in Knowledge Graph Embedding and Reasoning.

Keywords
Knowledge Graphs

1. INTRODUCTION
Knowledge Graph was first introduced by Google in 2012 with a purpose to improvise its search results. Any Graph-based representation of some knowledge wiz. Entities/Objects/Events/Concepts could be considered as a knowledge graph. An enormous amount of data that is generated daily is in Structured or Unstructured formats. Knowledge graphs helps to bring meaningful insights into the existing Data and also makes it possible to derive new Facts/Relations from the existing Facts/Relations. These features make KG widely acceptable and applicable at various different levels in different domains as shown in a below table. The general way of storing knowledge in a knowledge graph is using RDF triples i.e. subject-predicate-object. Representing data through Embedding have gained much attention due to its promising results. Also tensors are used by some researchers to store the data.

Table 1: Summarization of uses of KG in various Domains

Domain	KG Use
Medical	For diagnosing the various diseases from the symptoms, etc
Cyber Security	For identifying a security breach.
News	Fake news detection, etc
Education	To organize scholarly data to enhance the research activities, etc
Social Network	For Friend suggestion, fetching Friend circle, etc
Food	For fetching food related information, From ingredients suggesting food item, etc

Knowledge Graph: Phases
Knowledge Extraction & Construction
Different techniques like NER, Text Mining/ Analytics, Machine Learning, etc. can be used for extracting the Entity, Relation & Attribute. Extracted information is mainly stored in machine readable format like RDF, JSON-LD format, etc. (Zhao et al., 2018) have bifurcated Knowledge Extraction techniques into 2 types, namely: Rule based / Dictionary based approach and Machine Learning approach. Both approaches need the manually annotated corpus. To avoid the manual annotation of corpus, some semi-supervised and unsupervised algorithms were proposed. As per (Zhao et al., 2018), Machine learning classification types like Hidden Markov Models, Conditional Random Fields, K-Nearest Neighbors, Maximum Entropy Models and Support Vector Machines have been applied so far for extracting knowledge. Table shows the available tools for different types of knowledge extraction as listed by (Zhao et al., 2018). (Lee et al., 2018) have proposed a multi-label knowledge learning model that learns multi-labels from structured knowledge graphs applying Zero-shot learning technique. The technique is applied to image from which at a time multiple labels are identified. Deep neural network is applied to seen and unseen data to learn the multiple labels from an image.

Table 2: Available Tools for Knowledge Extraction

Type of Knowledge Extracted	Available Tools
Entity Extraction	Stanford NER, OpenNLP, AIDA, CiceroLite, FOX, Open Calais.
Sense Tagging	CiceroLite, FOX, Wikimeta
Relation Extraction	CiceroLite, FOX, Open Calais, ReVerb
Multilingual Entity Recognition	Wikimeta

KG construction can be done from Ontology or using entities extracted from any of the NER techniques, etc.Knowledge Graphs can be constructed from the different sources as listed in below Table with the examples of KGs:

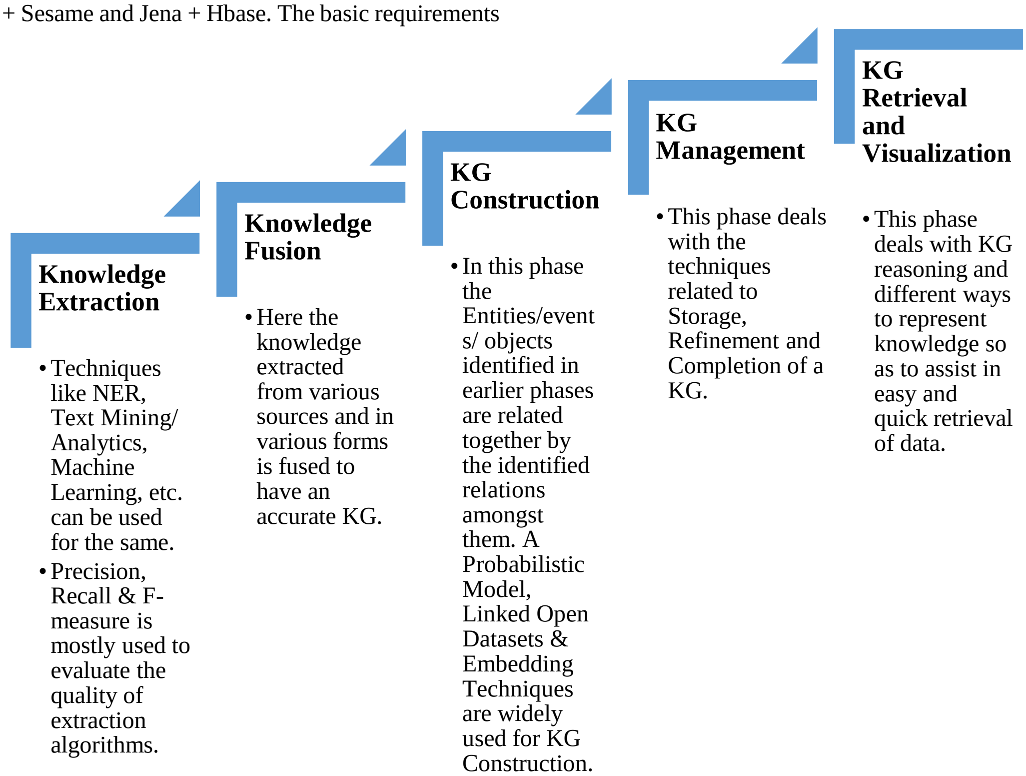
Table 3: Source & examples of KG

Source	Examples of KG
Structured Data- Tables, databases, social nets	Dbpedia, YAGO
Unstructured Data- WWW, news, social media, Reference Articles	Nell, Prospera, Knowledge Vault
Edited by Crowd	Freebase, Wikidata
Images, and Videos- YouTube, video feeds	

Knowledge Graph Storage

NoSQL databases are used for the storage of KG along with RDF (Resource Description Framework) and Graph database as the main storage types. (Zhao et al., 2018) have listed few of the hybrid storage approaches used like Hive + HBase, Cassandra + Sesame and Jena + Hbase. The basic requirements of storage techniques for KGs are: Scalability, Data Segmentation and timely caching and indexing.

The various phases of a KG is shown in the flow chart below.



Having obtained the overview of Knowledge Graphs like its construction and storage, now the investigation of Knowledge Graph Embedding & Reasoning is presented in upcoming sections.

2. KNOWLEDGE GRAPH EMBEDDING

Knowledge Graph Embedding (KGE) embeds the Entities and its Relations into a low-dimensional representation while preserving the semantics. The resultant embedding model has the ability of Knowledge inference and fusion. KGE overcomes most of the limitations of KG like issues of full coverage, Computational efficiency, Data Sparsity, coping with needs of real-time computation. (Dai et al., 2020) have shown the advantages of using a distributed representation:

- The data sparsity problem is diminished by embedding the entities and its relations into a low dimensional space.
- Rather than One hot representation, KGE uses a distributed representation approach. This makes semantic computation effective and efficient.

Various KGE Models are summarized in a table given below.

Table : Summary of Deterministic KGE model

Representation learning models based on Triplet Fact	Translational models	Word Embedding algorithm (Word2Vec). TransH, TransR, TransD, TranSparse, StransE, KG2E, TransG are few of the embedding models that were defined.
	Tensor factorization based model	It is an effective method for KG representation. Hole, RESCAL, DisMult are few of the embedding models that used tensor factorization.
	NN based model	Deep Neural Network is used as a base in the models.
Description based Models	Textual Description based model	These models focus on fetching entity type and entity description information, which assists Triplet fact based models to enhance their performance by fetching and attaching more detailed information to the entities and relations.
	Relation path based model	Along with one step relation, it also helps to find multi step semantic relation amongst entities.
KGE including temporal aspects of triplets		The observed triplets and the information related to temporal order of the triplet is embedded to obtain correct time bound relations amongst the entities.

All these KGE methods are distinguished based on a few criteria: Way Entities and Relationships are represented (vector representation, complex vectors or matrix), Scoring function (to find out the likelihood/correctness of a triplet by using relational feature, textual description etc) and Loss function. Changes in any of the criteria, makes a model better than the other models.

Pros and Cons of KGE-

Information stored in KGE is non-ambiguous. Hence it can be used effectively in reasoning systems and explainable systems.

i) Applications of Knowledge Graph Embedding

KGE can be used for various tasks as given below.

- Link Prediction- Used to predict the missing relation between the entities, or missing entity between an entity and a relation. To evaluate the prediction, each and every entity from the entity set is used to form a triplet. All such triplets are ranked by using a scoring function. The triplet with higher rank score is considered as the correct one.
- Triplet Classification- The correctness of an identified triplet is found, by considering it as a binary classification problem. To evaluate the triplet classification method, a score function is used. The threshold value is set to evaluate the score, to classify a triplet as a positive or negative triplet.
- Question-Answering system- KGE is used to understand the natural language question to give a precise and correct answer. Here the information and the question both are embedded to get the close embedding vector of question and answer.
- Recommendation System- The same technique of a Question-Answer system can also be used for recommendation systems.
- Knowledge Learning – The learning process requires a combination of positive and negative data as input. KG contains only positive data. KGE uses the corrupted triples to generate the negative triples.
- Named Entity Disambiguation (NED)- The concept of canonicalization and clustering of embedding is used by (Vashishth et al., 2018) to link the same entities together.
- Relation Extraction- While most relation extraction models focus on extracting local level relations, (Kim et al., 2020) have proposed a two staged (First stage- Extracting local relations & second stage- Constructing KG from o/p of first stage) relation extraction model to extract the relations from the documents.

ii) Scope of improvement

There is the possibility of refining/upgrading the performance of KGE models by adding additional information to entities and relations like- information extracted from other KG, multivariate information,etc. The ways of upgrading the KGE is shown by (Dai et al., 2020) in detail.

3. KNOWLEDGE GRAPH REASONING

Reasoning over Knowledge graphs has the ability to generate the set of triples that are logically derived from the KG and the set of rules. It can identify errors, obtain new knowledge, derive conclusions and bring new insights from the existing knowledge/ data. It can also enrich the existing knowledge graphs by finding the new relations among entities and feeding it back to the graph. (Chen et al., 2020) have categorized the knowledge reasoning methods of a KG into three categories as is summarized in table below. The similarity between these three methods is that all of them abstracts the KG into topology. Then they learn the parameters and model the features by using the topological relations between the entities.

Table: Dissimilarity between Reasoning Methods

Rule Based	Embedding Based	Neural Network Based
Abstract/concrete Horn Clause used for reasoning model.	Entities & its Relations are represented into a low vector space to find meaningful insights.	CNN or RNN used to extract features through its memorized self-learning ability. It outperforms other conventional methods.
Advantage Improves the accuracy by capturing the semantic information in graphs. Incorporates human's prior knowledge & logical reasoning to assist in reasoning. **Disadvantage** Dependency on Domain expert for generation of the rules. Noisy rules can lead to wrong reasoning.	**Advantage** Structural info in KG can be fully utilized Easy to transfer to large-scale KGs. **Disadvantage** Prior knowledge is of no use Reasoning ability is limited because it only uses facts but not compositional data.	**Advantage** It directly models the triples. It has a strong reasoning ability. **Disadvantage** It has poor Interpretability and high complexity.

A model of Knowledge aware path recurrent network was proposed by (Wang et al., 2019). It exploits KG, by composing the meaning of entities and relations, for deriving inferences. The model has the capability to reason the path selection, for deriving inferences.

i) Applications of KG Reasoning

- KG Completion- Inferring unknown facts from the known ones is one of the ways to complete the missing information in the KG.
- KG cleaning- The extracted patterns are not always reliable leading to extraction of false instances. KG Reasoning methods help to clean the noisy KB automatically.
- Entity classification- The IsA relation contained in the KG can be used for entity classification.
- Question-Answer System-

  a. Using Attention to Neighbor and Query dependent Embedding –The reasoning technique is required to extract the answer from a KB using existing facts and a multi-hopping feature. (Bansal et al., 2019) have proposed an A2N (Attention to Neighbor) model, which generates Query-specific embedding using a bi-linear attention on the graph neighbors of an entity. The model attends to each of the neighbors, assigns it a probability for finding its relevance in answering the query. It generates the query dependent embedding by aggregating the neighbor embedding (each assigned a scalar attention score and normalized over all neighbors to obtain the probability). Neighbor embedding is weighted by its relevance to an entity.

  b. Using KGE and multi hopping- (Saxena et al., 2020) have used KGE, Question embedding and Answer embedding to improvise the Q-A system.

ii) Scope of improvement

The scopes of improvement in KG reasoning methods identified by (Chen et al., 2020) is given below:

- Dynamical Knowledge reasoning- Existing KG reasoning methods focuses on Static data. However, KG is an ever evolving graph. Also a fact can be true for a specific time slot. Only a little preliminary work is done to address this problem.
- The existing Q-A systems perform well with single fact questions. For complex questions it sometimes performs poorly.
- Zero-shot reasoning- The reasoning from unseen class/data can help overcome the limitations of the existing reasoning methods like requirement a huge training set.
- Multi-Source information reasoning- Data is available in various different formats like video, audio, images and in textual format. To effectively and efficiently utilize these sources for information is a challenging task.
- Multi-lingual KG reasoning- Applications like machine translation, plagiarism detection (cross-lingual), information extraction from multi-lingual docs, etc can be worked upon.

4. CONCLUSION

The existing KGE models are inefficient in solving the multiple complex relation paths problems for the longer relation paths between entities. This issue can be solved by using deep learning technology by uniformly representing the relevant semantic information. Again, the concept of transfer learning can be applied to the knowledge representation, so as to make the KGE models adapt to new domains/applications.

Also not much work is done to explore the area for the n-ary facts. The open challenge is reasoning the n-ary graph data along with its efficient storage and retrieval of the Knowledge. Also no proper ways is yet found out related to validating the n-ary fact.

5. REFERENCES

[1] Abbas, N., Alghamdi, K., Alinam, M., Alloatti, F., Amaral, G., d'Amato, C., Asprino, L., Beno, M., Bensmann, F., Biswas, R., Cai, L., Capshaw, R., Carriero, V. A., Celino, I., Dadoun, A., De Giorgis, S., Delva, H., Domingue, J., Dumontier, M., … Xu, W. (2020). Knowledge Graphs Evolution and Preservation—A Technical Report from ISWS 2019. *ArXiv:2012.11936 [Cs]*. http://arxiv.org/abs/2012.11936

[2] Akrami, F., Guo, L., Hu, W., & Li, C. (2018). Re-evaluating Embedding-Based Knowledge Graph Completion Methods. *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*, 1779–1782. https://doi.org/10.1145/3269206.3269266

[3] Al-Moslmi, T., Gallofre Ocana, M., L. Opdahl, A., & Veres, C. (2020). Named Entity Extraction for Knowledge Graphs: A Literature Overview. *IEEE Access*, *8*, 32862–32881. https://doi.org/10.1109/ACCESS.2020.2973928

[4] Bansal, T., Juan, D.-C., Ravi, S., & McCallum, A. (2019). A2N: Attending to Neighbors for Knowledge Graph Inference. *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, 4387–4392. https://doi.org/10.18653/v1/P19-1431

[5] Bianchi, F., Rossiello, G., Costabello, L., Palmonari, M., & Minervini, P. (2020). Knowledge Graph Embeddings and Explainable AI. *ArXiv:2004.14843 [Cs]*. https://doi.org/10.3233/SSW200011

[6] Chen, X., Jia, S., & Xiang, Y. (2020). A review: Knowledge reasoning over knowledge graph. *Expert Systems with Applications*, *141*. https://doi.org/10.1016/j.eswa.2019.112948

[7] Cui, Z., Pan, L., Liu, S., & Cui, L. (2019). Infer Latent Privacy for Attribute Network in Knowledge Graph. *2019 IEEE International Conference on Big Data (Big Data)*, 2542–2551. https://doi.org/10.1109/BigData47090.2019.9006509

[8] Dai, Y., Wang, S., Xiong, N. N., & Guo, W. (2020). A Survey on Knowledge Graph Embedding: Approaches, Applications and Benchmarks. *Electronics*, *9*(5), 750. https://doi.org/10.3390/electronics9050750

[9] Fatemi, B., Taslakian, P., Vazquez, D., & Poole, D. (2020). Knowledge Hypergraphs: Prediction Beyond Binary Relations. *ArXiv:1906.00137 [Cs, Stat]*. http://arxiv.org/abs/1906.00137

[10] Galassi, A., Lippi, M., & Torroni, P. (2020). Attention in Natural Language Processing. *IEEE Transactions on Neural Networks and Learning Systems*, 1–18. https://doi.org/10.1109/TNNLS.2020.3019893

[11] Guan, S., Jin, X., Guo, J., Wang, Y., & Cheng, X. (2020). NeuInfer: Knowledge Inference on N-ary Facts. *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 6141–6151. https://doi.org/10.18653/v1/2020.acl-main.546

[12] Ji, S., Pan, S., Cambria, E., Marttinen, P., & Yu, P. S. (2021). A Survey on Knowledge Graphs: Representation, Acquisition and Applications. *IEEE Transactions on Neural Networks and Learning Systems*, 1–21. https://doi.org/10.1109/TNNLS.2021.3070843

[13] Kim, K., Hur, Y., Kim, G., & Lim, H. (2020). GREG: A Global Level Relation Extraction with Knowledge Graph Embedding. *Applied Sciences*, *10*(3), 1181. https://doi.org/10.3390/app10031181

[14] Koncel-Kedziorski, R., Bekal, D., Luan, Y., Lapata, M., & Hajishirzi, H. (2019). Text Generation from Knowledge Graphs with Graph Transformers. *ArXiv:1904.02342 [Cs]*. http://arxiv.org/abs/1904.02342

[15] Lee, C.-W., Fang, W., Yeh, C.-K., & Wang, Y.-C. F. (2018). Multi-label Zero-Shot Learning with Structured Knowledge Graphs. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 1576–1585. https://doi.org/10.1109/CVPR.2018.00170

[16] Lin, J., Zhao, Y., Huang, W., Liu, C., & Pu, H. (2021). Domain knowledge graph-based research progress of knowledge representation. *Neural Computing and Applications*, *33*(2), 681–690. https://doi.org/10.1007/s00521-020-05057-5

[17] Ma, Y., Tresp, V., & Daxberger, E. (2018). Embedding Models for Episodic Knowledge Graphs. *ArXiv:1807.00228 [Cs]*. http://arxiv.org/abs/1807.00228

[18] Popovski, G., Seljak, B. K., & Eftimov, T. (2020). A Survey of Named-Entity Recognition Methods for Food Information Extraction. *IEEE Access*, *8*, 31586–31594. https://doi.org/10.1109/ACCESS.2020.2973502

[19] Saxena, A., Tripathi, A., & Talukdar, P. (2020). Improving Multi-hop Question Answering over Knowledge Graphs using Knowledge Base Embeddings. *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 4498–4507. https://doi.org/10.18653/v1/2020.acl-main.412

[20] Shen, Y., Ding, N., Zheng, H.-T., Li, Y., & Yang, M. (2020). Modeling Relation Paths for Knowledge Graph Completion. *IEEE Transactions on Knowledge and Data Engineering*, 1–1. https://doi.org/10.1109/TKDE.2020.2970044

[21] Wang, Q., Mao, Z., Wang, B., & Guo, L. (2017). Knowledge Graph Embedding: A Survey of Approaches and Applications. *IEEE Transactions on Knowledge and Data Engineering*, *29*(12), 2724–2743. https://doi.org/10.1109/TKDE.2017.2754499

[22] Wang, X., Wang, D., Xu, C., He, X., Cao, Y., & Chua, T.-S. (2019). Explainable Reasoning over Knowledge Graphs for Recommendation. *Proceedings of the AAAI Conference on Artificial Intelligence*, *33*, 5329–5336. https://doi.org/10.1609/aaai.v33i01.33015329

[23] Zou, X. (2020). A Survey on Application of Knowledge Graph. *Journal of Physics: Conference Series*, *1487*, 012016. https://doi.org/10.1088/1742-6596/1487/1/012016