

# Blockchains as Knowledge Graphs – Blockchains for Knowledge Graphs (Vision Paper)

Luigi Bellomarini<sup>1</sup>, Giuseppe Galano<sup>1</sup>, Markus Nissl<sup>2</sup>, and Emanuel Sallinger<sup>2,3</sup>

<sup>1</sup> Central Bank of Italy

<sup>2</sup> TU Wien

<sup>3</sup> University of Oxford

**Abstract.** A body of recent work introduced the modelling of blockchain data as graph-based structures. Nevertheless, advanced tools for processing such data are mostly developed on top of the graph structure and are tailored to a specific analytical task, while the use of knowledge graph management systems that provide state-of-the-art reasoning algorithms is still in its infancy. In this paper, we discuss our vision for the FinTech field on the connection of the blockchain and knowledge graph domain, and provide various possible research topics by discussing, among others, the challenges in the field of blockchain analytics and the generation of legally compliant, unmodifiable and verifiable RegTech applications running on blockchain infrastructure by using knowledge graphs.

**Keywords:** Blockchain · Knowledge Graphs.

## 1 Introduction

Knowledge Graphs (KGs) have become a major topic in AI, in academic research and industrial applications. In the FinTech space, KGs are employed for many purposes, including advanced reasoning services to gain insight from that data.

In central bank settings, KGs are currently used for manifold settings such as checking regulatory compliance, anti-money laundering, or hybrid data science pipelines that combine a multitude of AI approaches [5]. Recent developments, e.g., on the regulation of cryptocurrency at the EU level [17], emphasize the need to offer such services also over blockchains.

A **knowledge graph** can be described as a semi-structured data model characterized by three components: (i) a ground extensional component having relational constructs for schema and data, e.g., a graph-like structure, (ii) an intensional component of inference rules over the constructs, and (iii) a derived extensional component produced by activating the inference rules over the ground extensional component in a so-called *reasoning* process [5].

Recent work has suggested solutions to various reasoning and data extraction tasks, such as *entity resolution* [11] for solving the question of whether two nodes refer to the same entity, *link prediction* [29] for predicting edges in the graph,

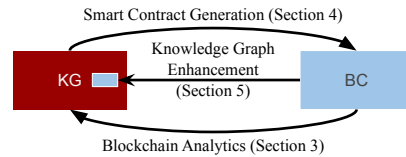
*knowledge fusion* [13] for predicting whether a node-edge-node triple is true, and computation of *KG embeddings* [31] with advanced deep learning algorithms.

Many researchers [19, 26] in the area of **blockchain analytics** have recognized that blockchains share common features with graph-like data structures and started implementing algorithms on top of the graph structure, such as *blockchain identity clustering* [28], *financial fraud detection* [22], *price prediction* [1] or *ransomware payment tracking* [21].

All of these tasks are crucial, yet they are **typically considered in isolation**. At the same time, what these tasks have in common is the goal of inferring new nodes and edges of a graph. That is, they suggest a precise mapping of the blockchain transaction graph into the KG extensional component and the newly generated nodes and edges into the intensional component. Such a *KG-oriented view* allows to see the mentioned blockchain tasks as reasoning, enabling the exploitation of knowledge shared among tasks and domain experience. For example, *share reasoning tasks*, such as the clustering of blockchain identities which is comparable to entity resolution, and *shared inference rules* such as operationalized domain knowledge, e.g., representing known money laundering or fraud schemes, or blockchain contained knowledge representing the functionality of smart contracts, e.g., rules for a lottery application.

With the option to model smart contracts as KGs, we also see the potential to generate smart contracts via KGs. This creates a **bidirectional connection**, allowing to monitor blockchain activity *at the KG* and creating legally compliant, verifiable RegTech applications *for the blockchain* by using the operationalized domain knowledge in the KG. I.e., in one direction the KG is used to create parts of the blockchain, and in the other direction the blockchain is used to create parts of the KG. RegTech stands for *Regulatory Technology* and describes the usage of information technology for enhancing the regulatory process, with its main application in the financial space. We suggest extending KGs with blockchain technology, such as verification, to be able to generate trustworthy smart contracts.

In this paper, we present our vision on the connection of blockchains and KGs. Figure 1 shows the overall vision presented here, and the sections of the paper we will discuss the parts in. Our ultimate goal is analyzing and monitoring blockchain data as well enabling the construction of RegTech applications for blockchain platforms, exploiting the capabilities of KG systems. In particular, our **main contributions** are:



**Fig. 1.** Overview of our vision.

- A novel view on **blockchain analytics** based on KGs by defining analytics as the derived extensional component of a blockchain KG, i.e., produced as a

result of reasoning. This will provide fully explainable analytics, allowing for deeper insights into the complex relations among the involved transactions.

- A new way of **smart contract generation** by using data and inference rules stored in KGs. This will result in legally compliant, verifiable smart contracts perfectly fitting for RegTech applications.
- An **extension of KGs with blockchain technology** by integrating digital signatures and consensus finding mechanisms. This will provide a form of “explainable trust” in KGs, a key feature for FinTech AI.

The remainder of this paper is **organized** as follows: In Section 2 we give background information on blockchains. In Section 3 we present our first vision of using KGs as data structure to reason over blockchain data, in Section 4 we discuss our vision of generating smart contracts by using KGs, and in Section 5 we present our vision to enhance KGs by blockchain technology. We provide additional related work in Section 6 and conclude this paper in Section 7.

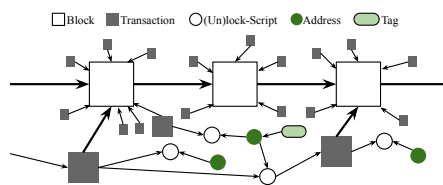
## 2 Background

A *blockchain* is a distributed ledger, where the blocks create a single-linked list (“chain”) by a hash reference from a block to the previous one. Each block contains a list of transactions. A transaction defines the information required for transferring data and coins between different accounts that are normally created by a private and public key pair, where the private key is used to spent the coins and (the hash of) the public key to receive the coins, thus also named the address of the account.

Depending on the blockchain state model, the transactions are handled differently. Bitcoin uses the model of *Unspent Transaction Output (UTXO)*, where the outputs are used as inputs for the next transactions, which creates an acyclic graph of transactions. In detail, the outputs and inputs are scripts, where the input unlocks the lock script of the output, offering support for more advanced operations such as requiring multiple parties to unlock the output.

In comparison, Ethereum simplifies the transaction management by decoupling the data from the transaction in a separate state per account, which is modified during a transaction. Scripts are handled within smart contracts. A smart contract is, intuitively speaking, a type of account that has to be actively invoked by other accounts. Smart contracts can contain and execute arbitrary code. For a detailed introduction into blockchains see [16].

Figure 2 gives a visualization of the Bitcoin structure as a graph, summarizing the discussed concepts. White square represents blocks, dark squares stand for transactions, white circles are inputs and outputs, and dark circles represent



**Fig. 2.** Graph view on bitcoin structure.

addresses. In addition, the figure contains tags of addresses, which is an external information required for blockchain analytics, described in Section 3.

Note that Figure 2 represents just one possible graph-based modeling. Others are possible and may be even favorable in certain reasoning tasks. Note that while in this section we focused on the graph aspects, the *knowledge* graph aspects discussed later will precisely allow translations between different graph-based models.

### 3 Knowledge Graph Reasoning for Blockchain Analysis

Throughout the last years, blockchain analysis has been a focus of research in a number of fields related to the financial, security, and societal domains [2]. Yet, there are challenges which highlight the necessity to develop new approaches and new reasoning techniques for blockchain data, which we lay out in the following.

**Motivation.** Most challenges arise in the context of large data volumes, with the consequent need for high scalability.

*Variability in fundamental operation.* Blockchains provide different ledger, block and transaction structures, smart contract languages (cf. [16, 6]) and privacy enhancing features [3]. While differences between data structures can be addressed using traditional data integration techniques, knowledge of the rules that govern the inner features of blockchains is essential for reasoning over the data.

*Multi-layered blockchain data.* Blockchains present a rich, multi-layered dataset consisting of the transaction graph at the uppermost level but going down to information present – sometimes in compiled form – in smart contracts.

*Specific domain knowledge.* Analytics queries typically need to reason on domain data and knowledge in addition to data contained in the blockchain. For example, tracking revenues of illicit activity [15] requires specific knowledge of laundering patterns. Moreover, information from the outside, such as tags [24], are required to link some sort of real-world entity with blockchain identities.

*Uncertainty about available information.* Probabilistic reasoning is required, for example, to assess the probability of a link between inputs and outputs of Bitcoin transactions or to apply heuristics that cluster addresses which are likely “controlled” by the same entity.

**Solutions.** A promising solution are KGs, which are designed for complex data and knowledge integration tasks as well as reasoning tasks. KGs naturally allow to represent differences in fundamental concepts using knowledge that represents the differences in operation, thus not requiring hard-coding such knowledge into reasoning algorithms. By building a unitary network of transactions involving multiple assets and rules, the data of smart contracts can be encoded in a transparent way for the reasoning process.

Having a system capable of dealing with the challenges mentioned, allows to use enriched blockchain data to get a deep understanding of the cryptoassets phenomenon [10], including the estimation of the daily transferred value or the tracking of illicit activities to identifiable points such as exchanges.

**Looking ahead.** Finally, it is worth noting that, although KGs enrich the information context of blockchains and enable sophisticated reasoning tasks, technology alone cannot overcome intrinsic limitations of blockchain analysis, which depend on the potential lack of the required information. For example, off-chain transactions, where only the final position is settled on the blockchain and mixing techniques, where the transaction target is hidden, increase the privacy of the user, but affects the analyses of the blockchain data. However, KGs can help to reduce the intrinsic limitations, for example by integrating different data sources in the reasoning process.

## 4 Knowledge Graph Generated Smart Contracts

Smart contracts are used to specify human-readable contracts and associated obligations by code. While smart contracts are written nowadays mostly in object-oriented style, a mismatch has been detected between the style and the intended purpose of enforcing conditions in a contract while the environment (blockchain) changes [12].

**Motivation.** Recent work identified numerous challenges to correctly codify smart contracts such as unmodifiability or invulnerability [18]. They suggested to use different smart contract formats based on Prolog [12] to close the mismatch mentioned above as well as to generate smart contracts via a grammar of institution [18] to protect against insecure smart contracts. While both are good approaches to simplify smart contract creation, they both have some limitations. The former requires writing logical programs not compatible with current blockchains, while the latter requires to structure sentences into logical parts.

**Solutions.** We suggest *generating smart contracts based on the knowledge stored in the KG*. We see three major motivations for doing so: *(i)* KGs already provide inference rules and a knowledge base, *(ii)* established blockchain platforms are supported, and *(iii)* external domain knowledge can be integrated.

We demonstrate the benefits through an example where we use the *bidirectional communication* to generate trustworthy initial coin offerings (ICOs) for funding new projects. ICOs have been suspect of exit scams over the last years [25]. For simplicity, let us assume that ICOs have to follow legal regulation to be considered as safe and there is a publicly certificated KG provider that ensures the validity of the ICO. By using KGs, we are able to **integrate domain knowledge** such as legal text or news announcements to check the validity of the provided ICO information. In case of a valid ICO, the KG provider **generates** the smart contract and **publishes** it with its own signature (trusted). Since RegTech applications are exposed to frequently changing rules, the KG provider has the possibility to create a **bidirectional communication channel** by monitoring the smart contract and updating the rules according to legal changes. For example, assume that the KG provider monitors the composition of the team stored in the smart contract. When the KG gets updated with a new team composition stemming, e.g. from trusted news sources, this updated knowledge is inspected. If it is determined that a member has left the team and

that this change is not included in the smart contract, the KG provider updates the rule of the smart contract.

We want to note, that this example uses exhaustive rights of the KG provider, which have to be prohibited. We refer to Section 5, where we present our vision of verified KG events.

**Looking ahead.** To realize this vision, a number of concrete challenges need to be solved. We mention a few of them:

*Interactions with smart contracts.* Today’s infrastructure of major blockchain technologies are not capable to update the code of smart contracts, thus requiring storing changeable rules in the storage of the contract and use them in the reasoning process of smart contracts.

*Generation of smart contracts.* Blockchains have diverse languages. Therefore, a modular algorithm for generating smart contracts has to be defined that includes a mapping between KG content and imperative smart contract languages.

*Verification and integration of heterogeneous domain knowledge.* The KG uses various heterogeneous data sources such as legal texts, user inputs, or web data. Algorithms have to be developed to intelligently verify the claims.

## 5 Enhancing Knowledge Graphs by Blockchain Technology

In this section we briefly present our vision to enhance KGs by fundamental concepts of blockchains, namely *(i) consensus protocols* which define the rules for block generation including conflict management as well as the rules of a valid block and transaction, *(ii) digital signatures* to sign transactions so that the block creators can verify that the transactions are executed on behalf of the owning parties, and *(iii) an unmodifiable data structure* by using hash algorithms to prevent changes of historical data.

**Motivation.** The application of such concepts to the KG allows to improve the reasoning algorithms by enriching the KG with trustworthy and historical knowledge to produce more reliable results. This requires an adaption of state-of-the-art reasoning algorithms to include the trustworthy aspect as well as a sounded analysis of the different integration possibilities of such concepts in the KG. For example, an unmodifiable data structure on the node-layer allows for fast history scans per node but signing the validity of a connected component requires at least the applicability on a cluster of nodes and edges.

**Solutions.** Having a KG capable of these concepts would improve the quality of KGs. Such a system can help to solve the **long term evolution** of real-time KGs, which is still an open problem [7], by integrating the unmodifiable and accessible history concept and can provide **verified knowledge graphs** by integrating the concept of digital signatures, which would allow to contain verified events in the KG, building a trustful connection between KGs and blockchains, solving, among others, the exhaustive rights problem of the KG provider mentioned in Section 4.

**Looking ahead.** Apart from direct improvements to KGs, there are also other disciplines which may profit from integrated blockchain technology. For example, many AI researchers are currently working on **explainable AI systems**. This means, they try to build intelligent systems that are able to answer questions concerning how and why automatic decisions were made in a human-comprehensible way. One way to build explainable AI systems is by using KGs [20]. We think that a verified and trusted state provided by blockchain technology in KGs may help explainable AI systems to decide and argue why they have made a specific decision. One aspect of this is the full, verifiable history of all facts established through a blockchain. For example, if a historic decision should be explained, it is very easy to go back in time via the blockchain and thus give fully verifiable explanations of AI decisions. Another aspect is the that blockchains allow a second level of explanation: For example, assume that a decision based on a KG is explained via a number of facts. If these facts are actually established via blockchain processes, e.g. voting, the explanation actually does not end here, but can be continued by explaining how that fact was established via a blockchain processes.

## 6 Related Work

The discussion of related work regarding blockchain and KGs is very limited. In the following, we discuss related work we see related to this topic and which has not been mentioned in the previous sections.

*Exchange of Assets.* We first focus on the blockchain as a mechanism for exchanging assets. GraphChain [27] uses the blockchain to store serialized versions of an RDF (Resource Description Format)-graph. Similarly, Naim and Klas [23] suggest to embed an RDF-graph in the blockchain. GraphOs [8] announced a network for exchanging knowledge assets, such as data, code, or asset ownership. *Crowdsourcing.* Systems are proposed where blockchains are used for verify KG knowledge established via crowdsourcing. For example, Wang et al. [32] suggests using crowdsourcing on the blockchain platform to update the KG and the AI system with a “trustful” value.

*Analytics.* We now move from surveying blockchains *for* KGs to using KG technology for blockchain analytics. Bartoletti et al. [4] created a general framework for blockchain analytics focusing on common tasks used in recent work to generate a shared view between them. This is an important first step, yet, this is not enough to address the challenges we discussed in Section 3. Similarly, regular path query (RPQ)-based languages in general show insufficient expressive power for our tasks as they do not support full recursion. Vo et al. [30] described research perspectives for the database community for blockchains in the fields of database management (creating indexes, data structures, generation of smart contracts, etc.) and blockchain analytics (missing data, query federation, etc.). In comparison, our vision on blockchain analytics focuses particularly on the intersection with KGs. We provide novel perspectives and highlight tasks specific to KGs.

*Other Sources.* Fluree [14] announced a semantic graph database with support of blockchain functionality to allow history-based queries. Cagle [9] discussed that KGs need blockchains to secure the keys, and that blockchain needs the KG to provide a context and provenance for the keys. By keys they mean some sort of public key to uniquely identify users and thus establish a chain of liability.

## 7 Conclusion

In this paper we have highlighted possible research directions at the intersection of blockchain and KG research by showing a bidirectional connection between these technologies. On the one hand, the monitoring aspect of the blockchain and on the other hand the smart contract generation aspect by the usage of KGs. In future work, we want to focus on these visionary topics in detail and look forward to presenting first solutions in this domain.

**Acknowledgements.** The work on this paper was supported by the WWTF (Vienna Science and Technology Fund) grant VRG18-013, the EPSRC grant EP/M025268/1, and the EU Horizon 2020 grant 809965.

## References

1. Akcora, C.G., Dey, A.K., Gel, Y.R., Kantarcioglu, M.: Forecasting bitcoin price with graph chainlets. In: PAKDD (2018)
2. Akcora, C.G., Kantarcioglu, M., Gel, Y.R.: Blockchain data analytics. In: ICDM (2018)
3. Alonso, K.: Zero to monero: First edition. <https://www.getmonero.org/library/Zero-to-Monero-1-0-0.pdf> (2018), [Online; accessed 2019-02-24]
4. Bartoletti, M., Lande, S., Pompianu, L., Bracciali, A.: A general framework for blockchain analytics. In: SERIAL@Middleware (2017)
5. Bellomarini, L., Fakhoury, D., Gottlob, G., Sallinger, E.: Knowledge graphs and enterprise AI: the promise of an enabling technology. In: ICDE (2019)
6. block.one: Eos.io technical white paper v2. <https://github.com/EOSIO/Documentation/blob/master/TechnicalWhitePaper.md> (2018), [Online; accessed 2019-03-23]
7. Bonatti, P.A., Decker, S., Polleres, A., Presutti, V.: Knowledge graphs: New directions for knowledge representation on the semantic web (dagstuhl seminar 18371). Dagstuhl Reports (2018)
8. Butcher, M.: Graphpath plans to combine knowledge graphs with the blockchain. <https://techcrunch.com/2018/05/14/graphpath-plans-to-combine-knowledge-graphs-with-the-blockchain/> (2018), [Online; accessed 2019-11-18]
9. Cagle, K.: The coming merger of blockchain and knowledge graphs. <https://medium.com/@kurtcagle/685e052c614c> (2019), [Online; accessed 2019-11-18]
10. Chimienti, M.T., Kochanska, U., Pinna, A.: Understanding the crypto-asset phenomenon, its risks and measurement issues. <https://www.ecb.europa.eu/pub/pdf/ecbu/eb201905.en.pdf> (2019), [Online; accessed 2020-01-07]



11. Christen, P.: Data Matching - Concepts and Techniques for Record Linkage, Entity Resolution, and Duplicate Detection (2012)
12. Ciatto, G., Maffi, A., Mariani, S., Omicini, A.: Smart contracts are more than objects: Pro-activeness on the blockchain. In: BLOCKCHAIN (2019)
13. Dong, X., Gabrilovich, E., Heitz, G., Horn, W., Lao, N., Murphy, K., Strohmann, T., Sun, S., Zhang, W.: Knowledge vault: a web-scale approach to probabilistic knowledge fusion. In: KDD (2014)
14. Doubleday, K.: Flureedb is production-ready and live. <https://medium.com/fluree/86bce82665dc> (2018), [Online; accessed 2020-01-16]
15. ErgoBTC: Tracking the plustoken whale: Attempted bitcoin mixing and its impact on wasabi wallet. <https://medium.com/@ErgoBTC/787c0d240192> (2019), [Online; accessed 2020-01-07]
16. Ethereum Foundation: A next-generation smart contract and decentralized application platform. <https://github.com/ethereum/wiki/wiki/White-Paper> (2015), [Online; accessed 2019-02-24]
17. European Parliament: Directive (eu) 2018/843. <https://eur-lex.europa.eu/eli/dir/2018/843/oj> (2018), [Online; accessed 2020-01-12]
18. Frantz, C., Nowostawski, M.: From institutions to code: Towards automated generation of smart contracts. In: FAS\*W@SASO/ICCAC. IEEE (2016)
19. Haslhofer, B., Karl, R., Filtz, E.: O bitcoin where art thou? insight into large-scale transaction graphs. In: SEMANTICS (2016)
20. Lecue, F.: On the role of knowledge graphs in explainable ai. *Semantic Web Journal* (2019)
21. Liao, K., Zhao, Z., Doupé, A., Ahn, G.: Behind closed doors: measurement and analysis of cryptolocker ransoms in bitcoin. In: eCrime (2016)
22. Möser, M., Böhme, R., Breuker, D.: Towards risk scoring of bitcoin transactions. In: Financial Cryptography Workshops (2014)
23. Naim, B.A., Klas, W.: Knowledge graph-enhanced blockchains by integrating a graph-data service-layer. In: 2019 Sixth International Conference on Internet of Things: Systems, Management and Security (IOTSMS) (2019)
24. OXT: Bitcoin addresses annotations. <https://oxt.me/notes> (2019), [Online; accessed 2020-01-07]
25. Patel, D.: 6 red flags of an ico scam. <https://techcrunch.com/2017/12/07/6-red-flags-of-an-ico-scam/> (2017), [Online; accessed 2020-01-16]
26. Ron, D., Shamir, A.: Quantitative analysis of the full bitcoin transaction graph. In: Financial Cryptography (2013)
27. Sopek, M., Gradzki, P., Kosowski, W., Kuziski, D., Trójezak, R., Trypuz, R.: Graphchain: A distributed database with explicit semantics and chained rdf graphs. In: Companion Proceedings of the The Web Conference 2018 (2018)
28. Spagnuolo, M., Maggi, F., Zanero, S.: Bitiodine: Extracting intelligence from the bitcoin network. In: Financial Cryptography (2014)
29. Taskar, B., Wong, M.F., Abbeel, P., Koller, D.: Link prediction in relational data. In: NIPS (2003)
30. Vo, H.T., Kundu, A., Mohania, M.K.: Research directions in blockchain data management and analytics. In: EDBT (2018)
31. Wang, Q., Mao, Z., Wang, B., Guo, L.: Knowledge graph embedding: A survey of approaches and applications. *IEEE TKDE* (2017)
32. Wang, S., Huang, C., Li, J., Yuan, Y., Wang, F.: Decentralized construction of knowledge graphs for deep recommender systems based on blockchain-powered smart contracts. *IEEE Access* (2019)