XLVIII Symposium on Operational Research
SYM-OP-IS 2021
Banja Koviljača, 20-23. September 2021.
MATΦ University of Belgrade Faculty of Mathematics
Mathematical Institute SANU

COMBINATORIAL OPTIMIZATION FOR SELF CONTAINED BLOCKCHAIN: AN EXAMPLE OF USEFUL SYNERGY

DUŠAN RAMLJAK[1], TATJANA DAVIDOVIĆ, DRAGAN UROŠEVIĆ, TATJANA JAKŠIĆ KRÜGER, LUKA MATIJEVIĆ, MILAN TODOROVIĆ, ĐORĐE JOVANOVIĆ[2]

[1] School of Graduate Professional Studies Penn State University Malvern, PA, 19355
dzr5484@psu.edu

[2] Mathematical Institute, Kneza Mihaila 36, Belgrade, Serbia
{tanjad,draganu,tatjana,luka,mtodorovic}@mi.sanu.ac.rs, giorgaki.jovanovic@gmail.com

Abstract: *With the increasing number of applications and systems that use Blockchain it is important to deal with its downsides. We are considering combinatorial optimization ideas to motivate and address two main issues in maintaining Blockhain. The first issue refers to economic usage of resources, and we propose a new consensus protocol exploring the Proof-of-Useful-Work concept. Deanonymization problem is considered as a second issue. It reduces to community detection in directed graphs which we propose to be consensus protocol's useful work. In order to implement the proposed ideas we provide a short targeted literature review and identify the steps that need to be performed within the implementation.*

Keywords: *Distributed databases, consensus protocols, deanonymization, metaheuristics, community detection*

1. INTRODUCTION

BlockChain (BC) is a specially designed distributed data storage structure, maintained without any third party of absolute trust by the so-called consensus protocol. The first BC implementation was related to the financial domain - performing transactions with Bitcoin cryptocurrency. Nowadays, it can be applied to different domains (digital identity, voting, notary, smart contracts, IoT, insurance, healthcare, etc.) [4, 31]. BC could be regarded as a public ledger where all data (referred to as "committed transactions") are stored in a list of blocks (Fig. 1). In literature and practice structures other than list appear, and the corresponding data storages are referred to by a common name *distributed ledger technology*, DLT [11, 26, 27]. Here, we always consider BC as a list of blocks. Access to BC is realized by exchanging messages between participants (clients, customers, entities, nodes). New data (organized in blocks) are allowed to be added by participants (the so-called miners) simultaneously. The main issues in maintaining BC are security, privacy, consistency, and reducing the consumption of electrical energy. Our goal is to contribute towards efficient resolution of some of these problems by applying combinatorial optimization methodology.

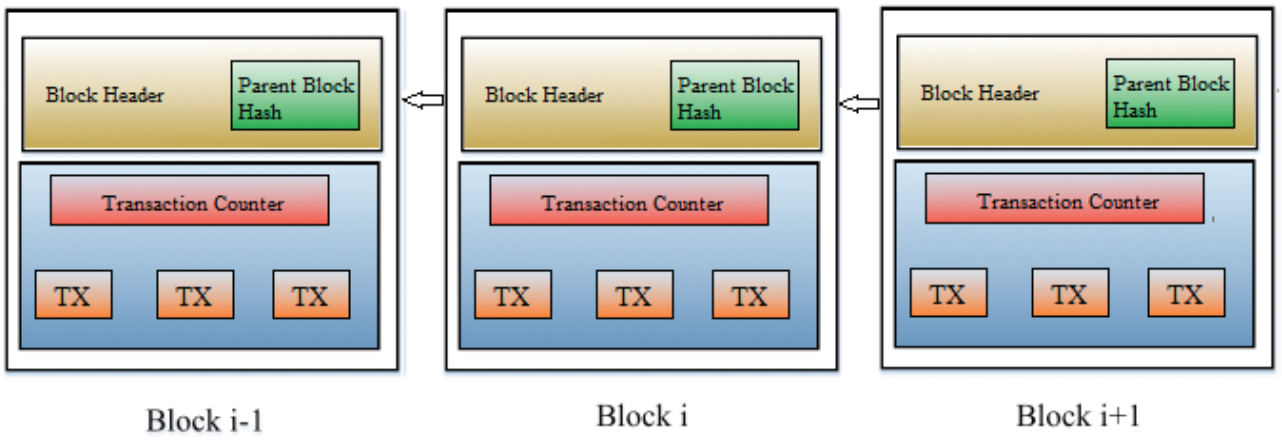

Figure 1 Structure of the BlockChain network.

We focus on the BC consensus protocols and the ledgers with the aims to reduce the power consumption and improve security of this distributed database by applying optimization techniques. Especially, the first aim is to

influence the more economic usage of resources by developing a new consensus protocol based on combinatorial optimization problems. Namely, Proof-of-Useful-Work (PoUW) [2, 7] is already established alternative to just solving the classical cryptographic puzzle (known as Proof-of-Work, PoW). It assumes dealing with some real-life optimization problems and enables miners to be doubly rewarded: for creating blocks and solving submitted problems. We intend to address a real-life optimization problem that hasn't yet been introduced in BC community as a consensus protocol. The second contribution aims to improve security in BC by examining the anonymity level of the participants, having in mind that the corresponding transactions are usually publicly available. The main tool for deanonymization is classification/clustering of transactions according to various criteria [3]. BC can be modelled by a directed graph, therefore in this work, we propose to treat deanonymization as a community detection problem. Having this optimization problem as a PoUW might lead to self contained BC, i.e., the security management in BC is at the same time useful work part of a consensus protocol.

The outline of the paper is as follows: We start with a short review of some relevant concepts of BC (consensus protocol, Proof-of-Useful-Work, deanonymization), as well as the well-known combinatorial optimization problems (clustering and community detection) in section 2. Section 3, then introduces the fundamental drivers for combining the proof of useful work with deanonymization. It then discusses the issues that need to be resolved before the successful implementation. Finally, section 4 concludes the paper.

2. BACKGROUND

2.1. Basic BC Concepts

BC should be regarded as a methodology that keeps the consistency of big data via internet. It can be seen as the chain of transactions' sets that are grouped into blocks. To ensure the consistency and security of BC, each block must contain a hash value of the previous block and must be timestamped [30]. Besides that, block consists of a list of valid transactions and various other data (such as the signature of the block's creator, threshold for the hash value of a valid block, nonce, etc.).

Consensus Protocol: In order to verify the new block to be added to the chain, some consensus protocol must be performed. The widely used consensus protocol, the so-called Proof-of-Work (PoW), requires miners to discover *nonce* value (a kind of secret key) in order to add a block into BC. Each miner tries to find a specific value for the nonce in such a way that the hash value of the whole block (nonce included) is smaller than the given threshold. It is important to note that threshold is not a constant as BC itself can change it in order to ensure that miners find a solution for the puzzle within a specified time window. The miner who finds a suitable nonce that solves the puzzle, broadcasts it to the network and other participants can easily verify if the hash value of the block in question is valid. If the block is valid, it will be appended to the chain. Once verified blocks that are included in the chain cannot be deleted or changed.

Proof-of-Useful-Work: Main drawback of classical PoW is the excessive energy consumption, however, some other issues (such as integrity, security, latency) are also identified. Among others, Proof-of-Useful-Work (PoUW) concept is applied to resolve some of these issues. It involves solving hard optimization problems, such as Travelling Salesman Problem (TSP) [16, 24] and Artificial Neural Network (ANN) training [2], as the work to be done within the PoUW consensus protocol. Two sources of reward are established for miners by PoUW: the basic one related to the insertion of new block into BC and the other one coming from the customers whose problem instances are solved within PoUW. In order to efficiently deal with the instances of selected optimization problems, metaheuristics, such as Variable Neighborhood Search (VNS) [20] and Bee Colony Optimization (BCO) [6, 12] could be used.

Deanonymization: BC offers its participants pseudoanonymity which could be done in the following manner: the transactions are made via e-wallet structure, which makes only its public key available for everyone on the internet. During the creation of e-wallet, synchronization with BC, as well as transaction signing, there should be no traces of the BC participant, and the IP address is not stored in the transaction packet [28]. Although the anonymity is a desirable feature from participants' viewpoint, it has been noticed that BC could be used as means to anonymize transactions that involve illegal activities [9, 22]. It is also known that a special class of viruses, called ransomware, can make transactions via BC. This subclass of viruses was later named crypto-ransomware [10]. The reason why this kind of transactions is hard to trace is the fact that generating new e-wallet addresses is a common occurrence.

The above mentioned issues represent weak points of the BC technology that allow malicious attacks. Therefore, motivations to deanonymize BC data include: fraud detection and law enforcement, identifying illegal and clean owners, system insights, anonymity and traceability, other socio-economical benefits of an open data model [5, 19, 23]. Deanonymization [25] is often performed by applying classification/clustering heuristics.

2.2. Selected Combinatorial Optimization Problems

Clustering is a form of unsupervised machine learning techniques, having the goal of discovering patterns in observed data. Since the consensus on what a cluster precisely is has not yet been reached, there are a certain number of clustering algorithm classes [8]. Some of them being: connectivity models, centroid models, distribution models, density models, subspace models, group models, graph-based models, signed graph models and neural models. The core issue, besides the definition of a cluster, is the choice of a number of clusters, since it heavily depends on the problem and the data. However, there are certain heuristic methods that may aid the process, such as elbow method, average silhouette method and gap statistic method [13]. There are also the algorithmic methods for determining the optimal number of clusters [17, 21].
Community detection in directed graphs helps to reveal the underlying community structure of those graphs. That is possible as real networks possess the clustering or community structure property, under which the graph topology is organized into modules commonly called communities or clusters. Nodes of the same community are highly similar while on the contrary, nodes across communities present low similarity [18]. Edge directions contain useful information that could help go beyond common clustering techniques [14]. In recent years, Nonnegative Matrix Factorization (NMF) has been effectively used to discover cluster structures due to its powerful interpretability property [15, 29]. In order to capture more accurate (positive and negative) relationships among clusters and, thereby, to derive a latent factor that is even proper for clustering and also has much more responsibility in the regularization Semi-Nonnegative Matrix Factorization could be used [1].

3. METHODOLOGY

In the text to follow we introduce the connection between combinatorial optimization and BC, and describe how the deanonymization problem could be used as useful work for PoUW consensus protocol, as illustrated in Fig. 2. We then discuss the issues involved in implementing our ideas.

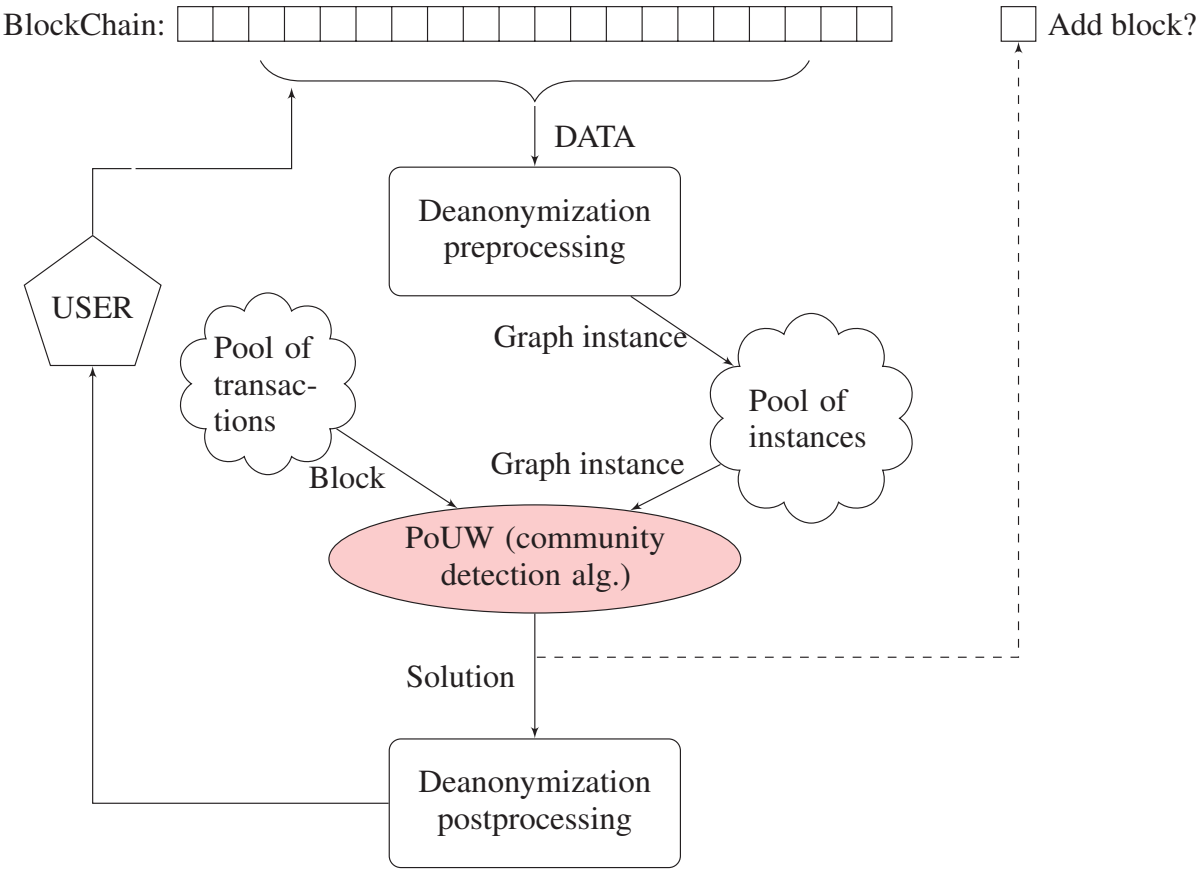

Figure 2 Self contained Blockchain.

3.1. PROPOSED CONSENSUS PROTOCOL

To insure safe, consistent, and efficient operation of BC we need to do both PoUW and deanonymization. As we already mentioned, deanonymization involves classification or clustering, which are well studied combinatorial optimization problems. We chose to treat deanonymization as the clustering problem because classification requires labeled data which we might not have.

Our idea is to consider the instances of clustering problem, generated by deanonymization process, as the useful work to be performed during the execution of PoUW. As BC data could be modelled with directed graphs (DG), our clustering problem boils down to community detection in those graphs.

More precisely, when the user needs to deanonymize the subset of BC data, deanonymization preprocessing phase creates an instance of DG and stores it in the pool of instances (see Fig. 2). PoUW selects transactions from the pool of transactions and forms a block to be added to BC. As in classical PoW, the correspondence between the composed block and an existing DG from the pool of instances needs to be established. The corresponding DG is solved by community detection algorithm as a part of PoUW. The result of PoUW now has additional benefit: besides possibility of adding a new block to BC, it provides user with a solution to the considered instance of the deanonymization community detection problem. If deanonymization is a part of the BC security management, BC becomes self contained framework.

BC may be modeled as a directed graph with vertices that contain entities, consisting of addresses, and the edges representing transactions between the entities. At the beginning of the community detection process, the preprocessing should be performed consisting of merging some vertices, grouping the addresses that are known to belong to same entities. Such a reduced graph becomes the starting point for the heuristic algorithm, which performs community detection based on the chosen similarity definition.

Before any clustering of BC data, there are several steps to be performed. Among the first ones is the selection of proper distance function or other cluster similarity measure. This problem may depend on data itself, and the solutions could be different for various data sets. In addition, to reduce the problem complexity, feature selection should be performed resulting in lowering the dimensionality of BC data to be clustered. This step is a part of data preprocessing phase.

3.2. GENERAL ISSUES

Before we can implement any PoUW type of consensus protocol these issues need to be resolved:

- Format for stating problem instances;
- Correspondence between problem instance and the composed block;
- The efficient exploration of dedicated hardware already owned by miners;
- Increasing the efficiency of the whole system by distributing the work among the available resources of each miner or among group of miners.

In most of the related papers a particular optimization problem is considered (e.g., TSP) and the clients only need to provide the input parameters (i.e., the number of cities and distances between them). However, such an approach is very limited as the number of clients requiring solutions for an instance of TSP may be negligible. On the other hand, the number of optimization problems and their variants is countless.

Once the pool of optimization problem instances is established, besides the set of transactions to compose a block, miners need to select an adequate instance of the optimization problem. The correspondence between problem instance and the composed block needs to be uniquely defined in order to mimic the connection between block, its hash value, and the hash value of the previous block. Only in a few papers this correspondence problem is mentioned, however, no adequate solution is provided.

In addition, dealing with some optimization problems may require special resources, both hardware and software. It may be easy for a miner to get the adequate software because there exists the efficient code for a lot of optimization methods. On the other hand, most of them require hardware that is not usually owned by a typical miner. The main suggestion in the literature how to avoid this problem involves hybrid approach. Namely, the miners are allowed to choose between the PoW and PoUW consensus protocols. This approach is adequate also in the case when the number of optimization problem instances is smaller then the number of blocks to be added in BC, so whenever the pool of optimization problem instances is empty, the miners can perform the classical hash-based PoW. Obviously, in order to increase their rewards, the miners are motivated to perform PoUW (i.e., to select optimization problem instances), as well as to invest in the new hardware suitable for PoUW.

The latency, one of the known PoW drawbacks may also appear in PoUW. Namely, some instances of the considered optimization problem may be very hard preventing each particular miner to solve it efficiently. A

well known solution to this issue in the optimization community is distributed computing. It is very easy to apply it in BC: several miners can join their resources and divide the required computations as well as the corresponding reward.

Moreover, it is important to define how to check the validity of the solution, where the solution will be stored and how to contribute to the BC analysis from the deanonymization standpoint.

4. CONCLUSIONS

Maintaining BlockChain (BC) is a complex and demanding task that involves, among other issues, security, privacy, consistency, and reducing the consumption of electrical energy. It is the well-known fact that BC is not immune to cyberattacks and fraud. Therefore, good security management system is of vital importance in BC. On the other hand, the maintenance of BC requires a huge amount of electrical energy.

In this paper we proposed to consider security and efficiency problems simultaneously and to combine solution approaches into a single procedure. More precisely, we considered deanonymization as the security problem and detected that it can be modeled by the community detection in directed graphs. This is a well-known optimization problem and could be addressed within the Proof-of-Useful-Work (PoUW) consensus protocol. Such a framework enables to achieve a combination of gains. Finding a high-quality solution of selected problem instance opens the possibility to add a new block to BC, while at the same time contributes to the BC security. Moreover, the consumed electrical energy is not wasted just for solving the cryptographic puzzle, rather it supports security management of BC as well. The proposed concept looks very promising, however, it requires several issues to be resolved before the actual implementation, opening fruitful research field for the future work.

Acknowledgement

This work was supported by the Science Fund of the Republic of Serbia, Grant AI4TrustBC: Advanced Artificial Intelligence Techniques for Analysis and Design of System Components Based on Trustworthy BlockChain Technology and by the Serbian Ministry of Education, Science and Technological Development, Agreement No. 451-03-9/2021-14/200029. The authors are grateful to Miloš Simić the PhD student of Univeristy of Belgrade and Nouf Albarakati from Temple University for many stimulating discussions and proofreading efforts. Special thanks to West Texas A&M University, College of Engineering for the help and support.

REFERENCES

[**1**] Abdollahi, R., Seyedi, S. A., and Noorimehr, M. R. (2020). Asymmetric semi-nonnegative matrix factorization for directed graph clustering. In *2020 10th International Conference on Computer and Knowledge Engineering (ICCKE)*, pages 323–328. IEEE.

[**2**] Baldominos, A. and Saez, Y. (2019). Coin. ai: A proof-of-useful-work scheme for blockchain-based distributed deep learning. *Entropy*, 21(8):723.

[**3**] Biryukov, A. and Tikhomirov, S. (2019). Deanonymization and linkability of cryptocurrency transactions based on network analysis. In *2019 IEEE European Symposium on Security and Privacy (EuroS&P)*, pages 172–184. IEEE.

[**4**] Casino, F., Dasaklis, T. K., and Patsakis, C. (2019). A systematic literature review of blockchain-based applications: current status, classification and open issues. *Telematics and informatics*, 36:55–81.

[**5**] Chawathe, S. S. (2019). Clustering blockchain data. In *Clustering Methods for Big Data Analytics*, pages 43–72. Springer.

[**6**] Davidović, T. and Glišović, N. and Rašković, M. (2018). Bee colony optimization for clustering incomplete data. In et al., S. B., editor, *Proceedings of the School-Seminar on Optimization Problems and their Applications (OPTA-SCL 2018)*, volume 2098, pages 94–108, Omsk, Russia. CEUR Workshop Proceedings.

[**7**] Dong, Z., Lee, Y. C., and Zomaya, A. Y. (2019). Proofware: Proof of useful work blockchain consensus protocol for decentralized applications. *arXiv preprint arXiv:1903.09276*.

[**8**] Estivill-Castro, V. (2002). Why so many clustering algorithms: a position paper. *ACM SIGKDD explorations newsletter*, 4(1):65–75.

[**9**] FBI (2012). 'bitcoin virtual currency: Unique features present distinct challenges for deterring illicit activity'. Technical report, Federal Bureau of Investigation, Report from the: Directorate of Intelligence, Cyber Intelligence Section and Criminal Intelligence Section.

[**10**] Gonzalez, D. and Hayajneh, T. (2017). Detection and prevention of crypto-ransomware. In *2017 IEEE 8th Annual Ubiquitous Computing, Electronics and Mobile Communication Conference (UEMCON)*, pages 472–478. IEEE.

[**11**] Hancock, M. and Vaizey, E. (2016). Distributed ledger technology: Beyond block chain. a report by the uk government chief scientific advisor. Technical report, Government Office for Science, London, GB.

[**12**] Jakšić Krüger, T. and Davidović, T. (2016). Analyis of the bee colony optmization method on 3-sat problem. In *Proc. 43rd Symposium on Operations Research, SYM-OP-IS 2016*, pages 297–301, Tara, Serbia.

[**13**] Kassambara, A. (2018). Determining the optimal number of clusters: 3 must know methods. https://www.datanovia.com/en/lessons/determining-the-optimal-number-of-clusters-3-must-know-methods/.

[**14**] Leicht, E. A. and Newman, M. E. (2008). Community structure in directed networks. *Physical review letters*, 100(11):118703.

[**15**] Li, T. and Ding, C.-c. (2018). Nonnegative matrix factorizations for clustering: A survey. In *Data Clustering*, pages 149–176. Chapman and Hall/CRC.

[**16**] Li, W. (2018). Adapting blockchain technology for scientific computing. *arXiv preprint arXiv:1804.08230*.

[**17**] Litvinenko, N., Mamyrbayev, O., Shayakhmetova, A., and Turdalyuly, M. (2019). Clusterization by the k-means method when k is unknown. In *ITM Web of Conferences*, volume 24, page 01013. EDP Sciences.

[**18**] Malliaros, F. D. and Vazirgiannis, M. (2013). Clustering and community detection in directed networks: A survey. *Physics reports*, 533(4):95–142.

[**19**] McGinn, D., McIlwraith, D., and Guo, Y. (2018). Towards open data blockchain analytics: a bitcoin perspective. *Royal Society open science*, 5(8):180298:1–14.

[**20**] Mladenović, N., Todosijević, R., and Urošević, D. (2014). Two level general variable neighborhood search for attractive traveling salesman problem. *Computers & operations research*, 52:341–348.

[**21**] Pelleg, D., Moore, A. W., and and others (2000). X-means: Extending k-means with efficient estimation of the number of clusters. In *Icml*, volume 1, pages 727–734.

[**22**] Raeesi, R. (2015). The silk road, bitcoins and the global prohibition regime on the international trade in illicit drugs: Can this storm be weathered? *Glendon Journal of International Studies/Revue d'études internationales de Glendon*, 8(1-2).

[**23**] Siegenthaler, J. (2020). Blockchain clustering with machine learning. Master's thesis, Center for Innovative Finance, University of Basel. adviser Prof. Dr. Fabian Schär.

[**24**] Syafruddin, W. A., Dadkhah, S., and Köppen, M. (2019). Blockchain scheme based on evolutionary proof of work. In *2019 IEEE Congress on Evolutionary Computation (CEC)*, pages 771–776. IEEE.

[**25**] Tikhomirov, S. (2020). *Security and Privacy of Blockchain Protocols and Applications*. PhD thesis, University of Luxembourg, Esch-sur-Alzette, Luxembourg.

[**26**] TradeIX (2018). The difference between blockchain and distributed ledger technology. URL: https://tradeix.com/distributed-ledger-technology/. Accessed Feb. 2021.

[**27**] Treiblmaier, H. and Clohessy, T. (2020). *Blockchain and Distributed Ledger Technology Use Cases*. Springer Nature, Cham.

[**28**] Turner, A. and Irwin, A. S. M. (2018). Bitcoin transactions: a digital discovery of illicit activity on the blockchain. *Journal of Financial Crime*, 25(1):109–130.

[**29**] Wang, F., Li, T., Wang, X., Zhu, S., and Ding, C. (2011). Community discovery using nonnegative matrix factorization. *Data Mining and Knowledge Discovery*, 22(3):493–521.

[**30**] Zheng, Z., Xie, S., Dai, H., Chen, X., and Wang, H. (2017). An overview of blockchain technology: Architecture, consensus, and future trends. In *IEEE international congress on big data (BigData congress)*, pages 557–564. IEEE.

[**31**] Zheng, Z., Xie, S., Dai, H.-N., Chen, X., and Wang, H. (2018). Blockchain challenges and opportunities: A survey. *International Journal of Web and Grid Services*, 14(4):352–375.