BRUNET ALEXANDRE

RAPPORT DE PROJET

2024

LES DONNÉES DE LA BLOCKCHAIN DE BITCOIN ET SES UTILISATIONS DANS UN CONTEXTE DE BIG DATA

URCA REIMS
M2 APE SEP

SOMMAIRE

INTRODUCTION

En 2009, la première blockchain a été créée par le mystérieux Satoshi Nakamoto. Cette nouvelle technologie est conçue pour transmettre des informations entre les membres d'un même réseau de manière décentralisée, sécurisée et immuable. Chaque transaction, qu'elle soit monétaire ou non, est consignée dans un grand livre qui est dupliqué auprès de tous les membres du réseau. La blockchain lancée par Satoshi Nakamoto avait pour seul objectif d'échanger de la monnaie numérique, donnant ainsi naissance au Bitcoin. Aujourd'hui, il s'agit d'une des plus grandes blockchains publiques, renfermant une quantité considérable de données en raison de leur immuabilité.

Cette exploration se concentrera sur le contexte du Big Data au sein de la blockchain publique Bitcoin. Ainsi, nous retracerons l'histoire des données qui s'y trouvent, mettant en avant leur extraction, leur traitement, leurs utilisations et les problèmes liés à la non-centralisation de l'information dans un contexte du Big Data. Plus particulièrement, nous allons nous intéresser au problème de validation d'un bloc étant donné le grand nombre de mineurs qui sont en concurrence. Autrement dit, tout le monde à accès aux données car les membres du réseau blockchain sont connectés en « peer-to-peer » donc ont tous la même base de données. Cependant les données sont régulièrement mises à jour, notamment lorsqu'un bloc est validé par deux mineurs qui valident à quelques millisecondes d'écart. Etant donné qu'il y a un certain temps pour que l'information se diffuse, nous pouvons nous demander (dans ce cas très précis servant d'exemple) qui a validé le bloc. C'est sur ce type de question que nous allons explorer ce champ d'application : en quoi le Big Data peut poser des problèmes dans un réseau blockchain ?

Ce rapport est fourni avec le lien GitHub[1] permettant d'accéder à la fois à ce rapport individuel, mais aussi au code permettant la visualisation des résultats du modèle, au rapport de TD et au résumé du symposium.

[1] https://github.com/Alfex-1/Outils_big_data

1. Le choix, la collecte puis le nettoyage des données d'une blockchain

Dans la sélection des données pour étudier l'impact du Big Data au sein de la blockchain Bitcoin, nous allons inclure les données sur les transactions, la validation des blocs, la diffusion de l'information, la concurrence entre mineurs, ainsi que la taille et la croissance de la blockchain.

Premièrement, concernant les données des transactions, une compréhension détaillée des caractéristiques essentielles telles que le montant des transactions, il nous faut récolter les adresses des expéditeurs et des destinataires, ainsi que les horodatages associés. Selon Antonopoulos (2014), ces informations sont accessibles sur le site web de blockchain.info[2] ou le logiciel Bitcoin Core. En se focalisant sur le montant des transactions, il est possible d'évaluer la dynamique économique au sein de la blockchain, identifiant les tendances de transaction et les variations de volume. Les adresses des expéditeurs et des destinataires offrent des informations sur les participants du réseau, permettant ainsi de comprendre la diversité des acteurs impliqués dans les transactions. Les horodatages associés à chaque transaction sont importants pour comprendre la chronologie des événements. Cette temporalité des transactions peut révéler des périodes d'activité intense, ou pas. Ces données sont pertinentes pour évaluer la charge de travail sur la blockchain, notamment en identifiant des pics d'activité qui peuvent entraîner des défis de traitement. Dans un contexte non-centralisé, telle que la blockchain, les données transactionnelles permet de comprendre la dynamique économique d'une blockchain telle que Bitcoin sous un angle temporel.

Deuxièmement, l'étude de la validation des blocs implique de recueillir des informations exhaustives sur le processus de validation des transactions. Cette démarche englobe une analyse approfondie des mineurs impliqués, des temps de validation, ainsi que des blocs concurrents. Selon les travaux d'Eyal et Sirer (2016), des données provenant du journal de transaction Bitcoin ou d'explorateurs de blocs en ligne se présentent comme des sources incontournables pour comprendre les défis liés à la validation dans un environnement non-centralisé. En identifiant les mineurs actifs, leur part relative dans le processus de validation, et la fréquence de leur participation, il devient possible de comprendre la dynamique concurrentielle entre les

[2] https://www.blockchain.com/explorer

acteurs du réseau. Les temps de validation des blocs sont des paramètres essentiels à considérer, influençant directement la capacité du réseau à traiter un grand volume de transactions. En analysant ces temps, on peut déterminer des tendances de performance du réseau, identifier des goulots d'étranglement potentiels, et anticiper des défis liés à la synchronisation des données, en particulier dans un contexte de concurrence accrue entre mineurs. Le temps de validation d'une transaction peut fortement varier selon l'affluence : lorsqu'il y a trop de transactions par rapport au nombre de mineurs, certaines sont placées en file d'attente et donc une semaine peut s'être écoulée avant qu'elle soit confirmée (Cryptonaute, 2020). L'identification et l'analyse des blocs concurrents ajoutent de la complexité à la compréhension du processus de validation. Les sources recommandées par Eyal et Sirer offrent des informations détaillées sur les transactions et les blocs, permettant ainsi d'approfondir l'analyse de la validation des blocs dans un environnement non-centralisé.

Troisièmement, l'examen de la diffusion de l'information au sein d'un réseau blockchain, nous devons explorer le délai entre la validation d'un bloc par deux mineurs concurrents et la propagation de cette information à l'ensemble du réseau. Selon les travaux de Narayanan et al. (2016), l'analyse de la propagation de l'information peut être approfondie en utilisant des données horodatées provenant de nœuds du réseau Bitcoin. Le délai de propagation entre la validation d'un bloc par deux mineurs concurrents offre des informations sur la résilience du réseau face à des événements tels que des validations simultanées. En comprenant les temps de diffusion, on peut évaluer la robustesse du protocole de communication du réseau et identifier des points potentiels de latence qui pourraient pertuber la synchronisation de l'information. L'analyse de la propagation de l'information repose sur des données horodatées provenant des nœuds du réseau Bitcoin. Ces données offrent une vision temporelle importante, permettant ainsi de suivre le cheminement de l'information à travers le réseau. L'utilisation de ces horodatages facilite également l'identification de décalages potentiels dans la diffusion de l'information, contribuant ainsi à une compréhension plus approfondie des défis liés à la communication décentralisée. Les travaux de Narayanan et al. mettent en avant l'importance de considérer la dynamique temporelle dans l'analyse de la propagation de l'information. En se basant sur ces recommandations, il devient possible d'élaborer des modèles plus précis de transmission de l'information au sein du réseau, fournissant ainsi des informations plus approfondis sur la manière dont les données se propagent à travers la blockchain Bitcoin.

Ensuite, lors de l'analyse des données liées à la concurrence entre mineurs, une exploration approfondie du nombre de mineurs en compétition pour valider un bloc et de la puissance de calcul qu'ils apportent s'avère indispensable. Selon les recherches d'Eyal et Sirer (2016), les pools miniers et les forums Bitcoin constituent des sources essentielles pour comprendre cette compétition. L'examen du nombre de mineurs en compétition pour valider un bloc offre un aperçu significatif de la dynamique concurrentielle au sein du réseau. En identifiant les fluctuations dans le nombre de participants, on peut évaluer la distribution de la compétition et comprendre comment celle-ci évolue au fil du temps. Cette information est déterminante pour anticiper des changements dans la structure de la concurrence et leurs implications sur la sécurité et la décentralisation du réseau. La puissance de calcul apportée par les mineurs est un facteur essentiel dans la compétition pour la validation des blocs. Une exploration détaillée de cette puissance de calcul permet de quantifier la contribution de chaque mineur, d'identifier des acteurs majeurs, et de détecter des éventuelles concentrations de pouvoir. Ces données sont souvent disponibles sur les pools miniers, qui regroupent les ressources de plusieurs mineurs pour augmenter leurs chances de valider des blocs. Les pools miniers offrent des informations détaillées sur la compétition entre mineurs. Ces plates-formes regroupent les ressources de plusieurs mineurs vers un objectif commun, et leur analyse permet de comprendre la collaboration et la compétition au sein de ces regroupements. Les forums Bitcoin, en tant que canaux de communication entre les mineurs, fournissent également des perspectives sur les défis et les enjeux auxquels ils font face. En se basant sur les recommandations d'Eyal et Sirer, l'exploration des données liées à la concurrence entre mineurs offre une compréhension approfondie des dynamiques concurrentielles. En utilisant les informations provenant des pools miniers et des forums Bitcoin, il est possible de discerner les nuances de cette compétition, de quantifier l'impact des mineurs individuels et d'anticiper les évolutions potentielles dans la structure du réseau.

Pour finir, en abordant la taille et la croissance de la blockchain, il est impératif de suivre ces paramètres au fil du temps afin de mettre en lumière les défis de gestion du Big Data. Selon les travaux de Lelarge et al. (2019), l'utilisation de données historiques provenant de sources telles que CoinMetrics[3] ou Blockchain Size offre des éléments pour analyser l'évolution de la blockchain Bitcoin. Le suivi de la taille de la blockchain au fil du temps permet de quantifier

[3] https://coinmetrics.io/

l'ampleur de l'ensemble des données stockées sur la blockchain Bitcoin. En observant les variations de taille, on peut identifier des tendances de croissance, des périodes d'expansion rapide, ainsi que des moments de stabilité. Ces informations servent à anticiper les exigences de stockage futures et comprendre la charge que cela impose au réseau. La croissance de la blockchain, évaluée par le nombre de transactions accumulées, offre une perspective sur la dynamique d'utilisation du réseau. L'analyse de cette croissance permet de déterminer si la blockchain suit une trajectoire continue, si elle rencontre des fluctuations importantes, etc. Ces éléments sont fondamentaux pour évaluer les capacités du réseau face à un flux constant de transactions. L'utilisation de données historiques provenant de sources telles que CoinMetrics ou Blockchain Size est préconisée par Lelarge *et al.* pour une analyse approfondie de la taille et de la croissance de la blockchain Bitcoin. Ces sources fournissent des métriques précises et des visualisations détaillées, permettant de contextualiser l'évolution de la blockchain dans le temps. En s'appuyant sur ces recommandations, l'analyse de la taille et de la croissance de la blockchain devient un outil essentiel pour comprendre les défis de gestion du Big Data. Les informations obtenues permettent d'anticiper les besoins futurs en matière de ressources, de développer des stratégies d'évolutivité adaptées, et de mieux appréhender l'impact de la croissance constante des données sur la décentralisation du réseau.

Après avoir collecté ces données, il n’est pas nécessaire de procéder à du nettoyage car les données sont par nature bien formatées et structurées, cela vient du fait que les sources formatent déjà au préalable les données.

2. La sélection des modèles d'analyse

Avec toutes ces données, nous devons trouver des modèles statistiques qui nous permettent d'analyser les caractéristiques des transactions, de la concurrence des mineurs, etc. Pour rappel, nous avions récolter des donner pour appréhender la dynamique économique et concurrentielle d'une blockchain telle que Bitcoin. Un des modèles qui nous permettront d'analyser cela c'est un modèle de série temporelle approprié au contexte financier : GARCH-X qui est une variante du modèle GARCH qui modélise la volatilité conditionnelle d'une série temporelle financière. Le modèle GARCH-X prend en compte, en plus, des variables explicatives.

On peut également analyser la distribution des temps de validation via la densité de Kernel (Kernel Density Estimation, KDE). Le but étant d'abord des événements exceptionnels dans le réseau lorsque le noyau est à faible densité. Ensuite, si la distribution des temps de validation présente des modes multiples (plusieurs pics), la KDE peut aider à identifier ces modes, indiquant ainsi la présence de différents comportements ou régimes de fonctionnement dans le réseau. Enfin, la KDE peut servir de première étape dans la modélisation statistique des données en fournissant des informations sur la structure de la distribution, ce qui peut orienter le choix du modèle statistique approprié.

En ce qui concerne la diffusion des informations au sein d'une blockchain, nous pouvons utiliser le modèle de diffusion épidémique avec composante temporelle. Il peut être adapté pour modéliser la propagation temporelle de l'information à travers un réseau, notamment dans le contexte de la validation des blocs dans une blockchain.

Pour modéliser la concurrence minière, nous pouvons nous appuyer sur le « Modèle de Jeu de la Preuve de Travail » (Proof-of-Work Game Model), particulièrement bien adapté à la blockchain Bitcoin qui fonctionne avec cet algorithme de consensus-ci. Dans ce modèle, les mineurs sont considérés comme des joueurs rationnels qui cherchent à maximiser leurs gains en résolvant des preuves de travail. En réalité c'est de la théorie des jeux dans lequel Les mineurs peuvent ajuster leurs stratégies en fonction de la concurrence, des coûts de production, des récompenses potentielles, etc.

3. Mise en forme

La mise en forme des données est une étape essentielle pour maximiser l'utilité des données dans le cadre de l'analyse. Elle prépare le terrain pour des modèles statistiques précis, des visualisations informatives, et une compréhension approfondie des patterns et des relations au sein des données.

Dans le contexte des données sur les transactions, une attention particulière est portée à la normalisation et à la standardisation des montants des transactions. Cette étape vise à rendre les données comparables, compte tenu des variations significatives de montants. De plus, les horodatages associés à chaque transaction sont décomposés en composantes temporelles (mois, jour, année) pour permettre une analyse fine de l'évolution temporelle des transactions. Les données provenant de sources telles que le site web de blockchain.info ou le Bitcoin Core sont souvent bien formatées, réduisant ainsi la probabilité de valeurs manquantes. Cependant, toute donnée manquante, si elle survient, est traitée de manière appropriée.

Concernant la validation des blocs, les temps de validation peuvent être soumis à une normalisation pour faciliter la comparaison. De nouvelles variables, telles que la fréquence de participation des mineurs, sont créées pour enrichir les informations extraites. La gestion des outliers, notamment les temps de validation exceptionnellement longs, est effectuée pour atténuer leur impact sur l'analyse.

Dans l'exploration de la diffusion de l'information au sein du réseau blockchain, les données horodatées sont transformées en formats appropriés pour faciliter l'analyse temporelle. Tout cas de valeur manquante, qui peut parfois être présent dans les données provenant des nœuds du réseau Bitcoin, est traité avec attention.

L'analyse des données liées à la concurrence entre mineurs implique la création de nouvelles variables, telles que la puissance de calcul agrégée des mineurs à partir des données des pools miniers. La gestion des outliers est essentielle pour atténuer l'impact des mineurs exceptionnellement puissants.

En ce qui concerne la taille et la croissance de la blockchain, la normalisation de la taille permet des comparaisons significatives. La création de nouvelles variables, comme la croissance de la blockchain basée sur le nombre de transactions accumulées, fournit des perspectives sur l'utilisation du réseau. La vérification constante de la cohérence des données garantit que les variations de taille correspondent aux attentes.

4. Analyse, interprétation

Nous avons précédemment discuté du modèle de diffusion épidémique avec composante temporelle, également connu sous le nom de modèle épidémique de Kermack–McKendrick (Brauer & Castillo-Chavez, 2012). Ce modèle est applicable non seulement dans le contexte de la propagation d'un virus, mais également lorsque des informations circulent. Supposons que $I(t)$ représente le nombre de nœuds ayant reçu l'information à un moment t (notre variable cible).

Dans ce cadre, supposons que β représente le taux de transmission de l'information, soit la probabilité qu'un nœud n'ayant pas l'information la reçoive d'un nœud la possédant. *N* est le nombre total de nœuds dans la blockchain, $R(t)$ représente le nombre de nœuds qui ont reçu ou validé une information à un moment donné t et ne peuvent plus la transmettre (« nœuds récupérés »), et $\gamma(t)$ représente le taux de récupération, soit la vitesse à laquelle un nœud ayant l'information l'a validée.

L'analyse temporelle implique l'examen de la dynamique de $\gamma(t)$ au fil du temps. Une diminution rapide de $\gamma(t)$ pourrait indiquer une propagation rapide de l'information dans les premiers stades, tandis qu'une diminution plus lente pourrait représenter une propagation progressive.

Ce modèle peut être décrit par des équations différentielles qui décrivent la propagation de l'information au fil du temps. Les équations du modèle peuvent être formulées de la manière suivante :

1. $\hat{c}\frac{\partial I(t)}{\partial t} = \beta I(t)(N - I(t) - R(t))e^{-\gamma(t)}$: Cette équation traduit la variation du nombre de nœuds infectés au fil du temps. Elle tient compte du nombre de nœuds susceptibles, représentés par $S(t) = N - I(t) - R(t)$.
2. $\frac{\partial S(t)}{\partial t} = -\beta S(t)I(t)e^{-\gamma(t)}$: Cette équation décrit la diminution du nombre de nœuds susceptibles. Un β plus élevé entraîne une transmission d'information plus rapide, conduisant ainsi à une diminution plus rapide du nombre de nœuds susceptibles.
3. $\frac{\partial I(t)}{\partial t} = \beta S(t)I(t)e^{-\gamma(t)} - \alpha I(t)$: Cette équation reflète l'augmentation du nombre de nœuds récupérés au fil du temps. Les nœuds récupérés sont ceux qui ont validé ou

récupéré l'information. Le taux α (« taux de validation ») détermine la vitesse à laquelle cela se produit.

4. $\frac{\partial R(t)}{\partial t} = \alpha I(t)$: Cette équation exprime le taux de changement des nœuds récupérés par rapport au temps. Elle représente la vitesse à laquelle de nouveaux nœuds récupérés apparaissent au fil du temps en raison de la propagation de l'information. Le taux α détermine cette vitesse.

$S(t)$ diminue à mesure que les nœuds susceptibles reçoivent l'information, $I(t)$ augmente à mesure que l'information se propage, puis diminue à mesure que les nœuds récupèrent. Le paramètre β reflète les délais de diffusion, modélisant la dynamique temporelle de la transmission d'information.

5. Visualisation

Il ne nous reste plus qu'à visualiser les résultats du modèle analysé, c'est-à-dire du modèle épidémique de Kermack–McKendrick. Pour cela, nous avons représenté le nombre de nœuds qui est susceptibles (qui n'ont pas encore reçu l'information), infectés (qui ont reçu l'information mais qui ne l'ont pas validé) et récupérés (qui ont reçu et validé l'information) sur une période 100 unités de temps (qui peuvent être des heures, des jours, etc.) (figure 1). Nous avons initialisé le modèle avec les conditions suivantes :

- $I(0) = 1$: un seul nœud possède l'information au temps $t = 0$;
- $S(0) = N - 1$: tous les nœuds sauf le nœud infecté ne possède pas encore l'information au temps $t = 0$;
- $R(0) = 0$: aucun nœud n'a encore validé l'information au temps $t = 0$.

En plus de ces conditions, nous avons paramétrer le modèle ainsi :

- $N = 1000$: le nombre total de nœuds dans le réseau ;
- $\beta = 0.3$: le taux de transmission de l'information ;
- $\gamma = 0.1$: le taux de récupération (vitesse d'adoption) ;
- $\alpha = 0.05$: le taux de validation.

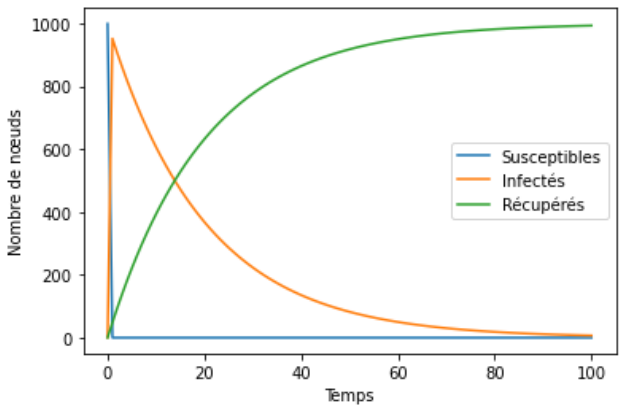

Figure 1 : Modèle épidémique de Kermack–McKendrick.

Bien évidemment, les paramètres ne sont pas déterminés arbitrairement et sont estimés. Mais l'objectif ici est de voir l'intérêt d'une telle visualisation et conclusions on peut en tirer. En effet l'intérêt majeur réside dans la prise de connaissance de la rapidité de l'information. Autrement dit, nous pouvons voir à partir de combien de temps, l'information circulera et sera validé au sein d'un réseau entier. Au-delà, cela nous permettra aussi de savoir à partir de quand $I(t) = R(t)$ et donc quand $I(t) = R(t)$.

6. Discussions

6.1. Limitations du modèle étudié

La discussion sur la sensibilité du modèle et ses limitations révèle des éléments essentiels pour évaluer la pertinence des résultats obtenus. L'analyse de ces aspects cruciaux permet une compréhension approfondie des nuances et des conditions d'application du modèle de diffusion épidémique avec composante temporelle.

En ce qui concerne la sensibilité aux paramètres, le modèle repose sur des facteurs clés tels que β, γ, et α, et il réagit de manière significative à leurs variations respectives. Par exemple, des ajustements dans β influent directement sur la rapidité de transmission de l'information, tandis que des changements dans γ impactent la vitesse de récupération des nœuds. Une analyse de sensibilité approfondie devient ainsi cruciale pour déterminer l'influence de ces paramètres et guider les ajustements nécessaires.

Cependant, le modèle présente des limitations importantes. Tout d'abord, il suppose une homogénéité parmi les nœuds, les considérant avec des caractéristiques similaires. Cette simplification peut négliger la diversité des participants dans l'écosystème complexe des blockchains. De plus, les paramètres β, γ, et α sont traités comme constants, ne prenant pas en compte les fluctuations potentielles dans un environnement blockchain dynamique. Enfin, le modèle ne tient pas compte des influences externes, telles que des événements spécifiques à la blockchain, des mises à jour de protocoles, ou des incidents de sécurité, qui peuvent fortement impacter la diffusion de l'information.

Dans une perspective d'amélioration du modèle, plusieurs stratégies peuvent être envisagées. L'introduction de paramètres dynamiques, la prise en considération de sous-groupes de nœuds avec des comportements différents, ou encore l'incorporation de données historiques pour refléter les variations temporelles peuvent renforcer la validité du modèle. Expérimenter avec différentes configurations représente une voie prometteuse pour mieux aligner le modèle sur les spécificités des blockchains.

En dépit de ses limitations, le modèle offre une base solide pour comprendre la dynamique de la diffusion de l'information au sein d'une blockchain. Les résultats obtenus peuvent orienter les discussions sur l'optimisation des protocoles de diffusion d'informations et fournir des perspectives pour anticiper les scénarios potentiels.

6.2. Les phénomènes de fork

La blockchain peut faire face à des conflits. En effet, des phénomènes de « forks » (ou fourche en français) peuvent apparaître. Un fork de blockchain se produit lorsque sa communauté apporte un changement qui modifie le fonctionnement du protocole. Il existe les « soft forks » et les « hard forks »[4] mais ici nous seulement globaliser en fork. Ainsi, une deuxième blockchain se sépare de la première (figure 2). Cette dernière partage une histoire commune avec l'originale, mais sa propre histoire débute lorsqu'elle se sépare de l'originale.

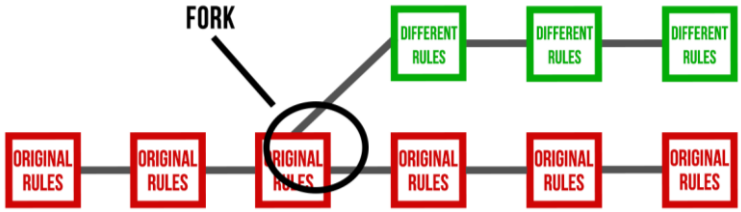

Figure 2 : Un fork de blockchain. **Source** : Good Audience, 2019.

Cela peut donc avoir plusieurs problèmes en termes de prise de décision et de récolte des données. En fait, la blockchain (que ça soit Bitcoin ou une autre) se base sur la prise de décision non-centralisée grâce aux copies du grand livre que tous les membres possèdent. Et dans un contexte de big data, cette prise de décision est répartie. Cependant, si un fork se produit, les décisions peuvent ne pas être cohérent avec la blockchain originale, car les règles auront changé. Ainsi, la synchronisation est difficile, voire impossible. Cela impacte également les données en elles-mêmes, car nous ne pouvons pas vérifier leur véracité. À cause des forks, les informations sont temporairement (voire définitivement) différentes, ce qui pose des problèmes de cohérence : introduction d'un arbitrage entre vraie et fausse vérité.

[4] Selon le blog BitPay, en août 2017, des phénomènes de soft fork et de hard fork ont été observés. Le 23 août, le protocole Bitcoin a été mis à jour pour inclure plus de transactions dans chaque bloc (soft fork). Le 1er août, un certain nombre d'individus souhaitent augmenter la taille limite des blocs de transactions, ce qui a conduit à la création de la blockchain Bitcoin Cash et de la crypto-monnaie (hard fork).

CONCLUSION

En conclusion, bien que l'analyse du Big Data au sein d'une blockchain, telle que Bitcoin, puisse offrir des perspectives enrichissantes sur la dynamique économique et concurrentielle, il est crucial de reconnaître les défis potentiels que le Big Data peut poser dans un tel réseau décentralisé.

La collecte et le traitement massif de données transactionnelles, de validations de blocs, de la diffusion de l'information, de la concurrence entre mineurs, ainsi que de la taille et de la croissance de la blockchain, peuvent entraîner des problèmes liés à la gestion du volume de données. La charge de travail sur la blockchain peut connaître des pics d'activité, entraînant des défis de traitement et de stockage. Ces défis peuvent influencer la scalabilité du réseau, la synchronisation des données, et potentiellement compromettre l'efficacité opérationnelle de la blockchain.

Le choix des modèles statistiques appropriés joue un rôle central dans l'analyse des caractéristiques des transactions, de la concurrence des mineurs, et plus encore. Le modèle GARCH-X, adapté au contexte financier, offre une perspective temporelle sur la volatilité de la blockchain. La densité de noyau (KDE) permet d'explorer la distribution des temps de validation, tandis que le modèle de diffusion épidémique avec composante temporelle offre une approche innovante pour comprendre la propagation de l'information au sein de la blockchain.

Par ailleurs, la sensibilité des modèles statistiques choisis, tels que le modèle épidémique de Kermack–McKendrick, aux paramètres β, γ, et α souligne l'importance d'une estimation précise de ces paramètres. Les variations significatives dans ces paramètres peuvent entraîner des résultats divergents, mettant en évidence la nécessité d'une compréhension approfondie des mécanismes sous-jacents.

Bien que l'analyse du Big Data puisse apporter des informations très intéressantes, les défis liés à la gestion des données massives et à la sensibilité des modèles soulignent l'importance d'une approche équilibrée. Il est impératif de développer des solutions innovantes pour surmonter ces défis et maximiser les avantages potentiels du Big Data dans le contexte complexe et dynamique des réseaux blockchain, surtout pour une des plus grande blockchain du monde : Bitcoin.

BIBLIOGRAPHIE

Antonopoulos, Andreas M. *Mastering Bitcoin: Unlocking Digital Cryptocurrencies*. O'Reilly Media, Inc., 2014.

Brauer, Fred, et Carlos Castillo-Chavez. « Epidemic Models ». In *Mathematical Models in Population Biology and Epidemiology*, par Fred Brauer et Carlos Castillo-Chavez, 345-409. Texts in Applied Mathematics. New York, NY: Springer New York, 2012. https://doi.org/10.1007/978-1-4614-1686-9_9.

Croman, Kyle, Christian Decker, Ittay Eyal, Adem Efe Gencer, Ari Juels, Ahmed Kosba, Andrew Miller, et al. « On scaling decentralized blockchains (A position paper): International Workshops on Financial Cryptography and Data Security, FC 2016 and 3rd Workshop on Bitcoin and Blockchain Research, BITCOIN 2016, 1st Workshop on Advances in Secure Electronic Voting Schemes, VOTING 2016, and 4th Workshop on Encrypted Computing and Applied Homomorphic Cryptography, WAHC 2016 ». Édité par Kurt Rohloff, Jeremy Clark, Sarah Meiklejohn, Dan Wallach, Michael Brenner, et Peter Y.A. Ryan. *Financial Cryptography and Data Security - International Workshops, FC 2016, BITCOIN, VOTING, and WAHC, Revised Selected Papers*, Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2016, 106-25. https://doi.org/10.1007/978-3-662-53357-4_8.

Narayanan, Arvind, Joseph Bonneau, Edward Felten, Andrew Miller, et Steven Goldfeder. *Bitcoin and Cryptocurrency Technologies: A Comprehensive Introduction*. USA: Princeton University Press, 2016.

Olleros, F. Xavier, et Majlinda Zhegu. *Research Handbook on Digital Transformations*. Edward Elgar Publishing, 2016.

Rathee, N., Singh Ankita, T. Sharda, N. Goel, Aggarwal Mansi, et S. Dudeja. « Analysis and Price Prediction of Cryptocurrencies for Historical and Live Data Using Ensemble-Based Neural Networks ». *Knowledge and Information Systems* 65, n° 10 (2023): 4055-84. https://doi.org/10.1007/s10115-023-01871-0.

WEBOGRAPHIE

BitPay Blog. « What is a Blockchain Fork? Hard Forks vs Soft Forks Explained | BitPay », 12 décembre 2023. https://bitpay.com/blog/blockchain-forks/.

« Blockchain Explorer - Bitcoin Tracker & More | Blockchain.Com ». Consulté le 17 février 2024. https://www.blockchain.com/explorer.

« Blockchain.Com | Charts - Average Confirmation Time ». Consulté le 17 février 2024. https://www.blockchain.com/explorer/charts/[id].

Blockgenic. « Blockchain Forks Explained ». *Medium* (blog), 10 janvier 2020. https://medium.com/@blockgenic/blockchain-forks-explained-17f22efbf5d3.

Cryptonaute. « Le temps de confirmation d'une transaction Bitcoin ». Cryptonaute, 22 septembre 2018. https://cryptonaute.fr/delai-temps-confirmation-transaction-bitcoin/.

Metrics, Coin. « Home ». Coin Metrics. Consulté le 17 février 2024. https://coinmetrics.io/.

TABLE DES FIGURES