

# La *blockchain* vue sous l'angle des brevets : cartographie d'une technologie émergente

## Blockchain from a patent perspective: Mapping an emerging technology

Raffaele Anedda<sup>1</sup>

<sup>1</sup> Unité d'Economie Appliquée, ENSTA Paris, Institut Polytechnique de Paris, France, raffa.anedda@gmail.com

**RÉSUMÉ.** Les données de brevets commencent à être utilisés pour cartographier le système d'innovation de la *blockchain*. Nous utilisons l'analyse des réseaux pour représenter ce système à partir de données de la base PATSTAT. Nous enrichissons cette approche tout d'abord en considérant une nomenclature technologique plus adaptée aux technologies émergentes. Ensuite, nous augmentons la liste des outils utilisés pour l'analyse des réseaux (centralité, petits mondes, citations entre déposants). Nous montrons que les inventions autour de la *blockchain*, en dépit d'une forme d'autonomie des réseaux d'acteurs distinguant une dynamique occidentale d'une dynamique chinoises, connaissent des trajectoires technologiques similaires explorées dans ces deux univers.

**ABSTRACT.** Patent data is beginning to be used to map blockchain innovation systems. We use network analysis to study this system using data from the PATSTAT database. We enrich this method first by considering a classification of technologies that would be more suitable for emerging ones. Then, we expand the list of tools used for network analysis (centrality, small worlds, citations between filers). We demonstrate that inventions around blockchain – despite a type of autonomy of networks of actors that distinguish a Western dynamic from a Chinese one – experience similar technological trajectories explored in these two universes.

**MOTS-CLÉS.** *Blockchain*, Brevet, Analyse des réseaux.

**KEYWORDS.** Blockchain, Patent, Network analysis.

### 1. Introduction

Le brevet est-il un support pertinent pour représenter les processus et les écosystèmes technologiques d'innovation de la *blockchain* ? Une exploration rapide de la plateforme de données Orbit Intelligence, isolant le simple mot clé « *blockchain* » dans les titres, résumés et revendications de brevets, indique à la mi-2021 l'existence de plus de 12 000 « familles » de brevets déposées dans le monde comprenant ce terme. Ce titre de droit de propriété intellectuelle semble donc constituer un matériau riche pour proposer une cartographie de l'activité d'innovation autour de la *blockchain*. C'est d'ailleurs le constat effectué par [MAR 20] et [CLA 20]. Dans notre article, cette masse de données nous conduit à appliquer la démarche empirique introduite par [KRA 11] sur la structure des bases de connaissances associées à des technologies particulières (dans leur contribution, les biotechnologies). Chez ces auteurs, les bases de connaissances sont représentées sous la forme de réseaux, et leurs propriétés sont révélées par l'usage d'indicateurs de l'analyse des réseaux sociaux (ARS ; [WAS 94]). Nous procéderons de même ici.

Dans ces réseaux, les nœuds sont des unités élémentaires de connaissances, ces « briques » étant reliées entre elles lorsqu'elles ont une utilisation conjointe. L'utilisation est conjointe lorsqu'un même brevet est associé simultanément à plusieurs codes d'une nomenclature technologique. Par exemple, au brevet *p* associé simultanément aux codes (« briques ») A et B correspond un réseau composé de deux « nœuds » reliés entre eux par un « lien ».

La représentation des relations entre briques de connaissances sous la forme d'un réseau à partir de données de brevets est donc, pour [KRA 11], un moyen de caractériser la structure des bases de connaissances technologiques. A partir de là, il s'agit pour ces auteurs notamment de détecter l'impact de ce réseau sur l'organisation industrielle, et ce dans un cadre évolutionniste. En effet,

dans ce cadre, la connaissance y est caractérisée par un degré élevé de cumulativité et de recombinaisons, en même temps qu'elle les déborde en engendrant des innovations plus radicales. La représentation sous la forme d'un réseau a justement pour particularité de tenir compte de ces caractéristiques. L'émergence de nœuds additionnels dans le réseau traduit une forme de radicalité de l'innovation, tandis que la création de nouveaux liens entre des nœuds existants correspond à de la recombinaison de briques de connaissances.

L'objectif de notre article est de cartographier la *blockchain* à partir de données de brevets. Par rapport à [KRA 11], nous proposons deux approfondissements. Le premier porte sur la construction du réseau. Nous utilisons une nomenclature technologique plus riche que celle mobilisée par ces auteurs : la CCB (Classification Coopérative des Brevets) en lieu et place de la CIB (Classification Internationale des Brevets), cette dernière étant utilisée par [KRA 11]. La CIB compte environ 70 000 codes dans son découpage le plus fin, contre 260 000 pour la CCB. Cette dernière est gérée conjointement par l'OEB (Office Européen des Brevets) et l'USPTO (son équivalent étasunien). La CCB est considérée comme étant plus pertinente pour identifier les technologies émergentes [OEB 17].

Le deuxième approfondissement porte sur l'usage des outils de l'ARS. Au-delà des indicateurs de centralités des nœuds du réseau [FRE 79], nous proposons des indicateurs propres à la « nouvelle science des réseaux » [WAT 04], notamment pour caractériser la *blockchain* en termes de « petit monde ». Si ce réseau est un petit monde (un réseau très clusterisé et où la distance moyenne entre les nœuds est faible), cela signifie qu'il est multipolaire, et en ce sens qu'il est caractérisé par des dynamiques d'innovation spécifiques à certaines de ses composantes. Ce qui induit que les briques technologiques de la *blockchain* ne tirent par toutes dans le même sens, qu'elles n'innovent pas nécessairement au même rythme, et que des recombinaisons internes radicales peuvent potentiellement se produire.

L'article est composé de deux sections. Dans la première, nous définissons le périmètre technologique de la *blockchain* en nous référant à la nomenclature CCB. Dans la deuxième, nous procédons à l'analyse empirique de l'écosystème technologique « *blockchain* » sur la période 2000-2015 à partir des brevets déposés à l'OEB et à l'USPTO. Dans cette section, nous faisons également une comparaison entre la structure des connexions technologiques en occident et celle que l'on constate sur le marché chinois des technologies sur cette même période. [CLA 20] ont en effet constaté que les réseaux d'acteurs participant aux dynamiques d'invention sur la *blockchain* sont spécifiques en occident et en Chine. Est-ce que cela se traduit par des bases de connaissances elles-mêmes spécifiques à ces territoires ?

## 2. Le périmètre « brevets » de la *blockchain*

En 2017, l'Office Européen des Brevets (OEB) édite un rapport sur les technologies de la quatrième révolution industrielle (4RI) telles qu'elles apparaissent dans les données de brevets. Pour l'OEB, les technologies de la 4RI comprennent les objets intelligents connectés, la création de la valeur par les données (*big data*, informatique distribuée, stockage de données sur le *cloud*, nouveaux outils d'analyse de données, systèmes de simulation 3D, fabrication additive) et l'innovation tirée par le logiciel. Ces technologies trouvent des champs applicatifs très diversifiés et croissants : l'organisation de la production, les transports, la gestion des systèmes énergétiques, la finance, etc. Dans cet exercice, la *blockchain* n'apparaît pas comme telle. Elle est citée, cependant, à travers de nombreux champs d'application (dans la finance, la santé, l'industrie agroalimentaire, etc.).

Dans le prolongement de ce rapport se développe alors une littérature faisant directement le lien entre brevets et *blockchain*. Cette littérature reste aujourd'hui cependant embryonnaire, car très récente. Deux contributions nous semblent plus particulièrement marquantes : [MAR 20] et [CLA 20]. Chacune développe une stratégie spécifique pour isoler les brevets qui ont trait à la *blockchain*.

[MAR 20] privilégient une démarche mobilisant prioritairement le contenu textuel des brevets. Dans leurs titres, leurs résumés et leurs revendications, ils requêtent un ensemble de mots clés associés à la technologie : « *blockchain* », « *block chain* », etc. Ils repèrent, à partir du portefeuille obtenu, les codes de la nomenclature CIB les plus présents. Puis ils procèdent à un *clustering* sur les cooccurrences technologiques présentes dans les brevets faisant référence à ces codes pour aboutir finalement à une carte technologique polycentrique de la *blockchain*.

La démarche de [CLA 20] est toute autre. Ils cherchent en effet à définir le périmètre *blockchain* le plus strictement possible. Dans la démarche précédente, en effet, il se peut que la recherche textuelle intègre des brevets citant la *blockchain* mais ne faisant pas de celle-ci le cœur de l'invention. Ainsi, contrairement à [MAR 20], leur point de départ réside dans un travail sur la nomenclature technologique et, en prolongement des travaux de l'OEB cités plus haut, c'est la nomenclature CCB qui se prête le mieux à cet exercice selon ces auteurs.

Ce travail sur la CCB se fait en lien avec des experts brevets de l'OEB, afin d'avoir la lecture la plus opérationnelle possible de cette nomenclature, qui est très difficile à manipuler car en évolution permanente. En effet, celle-ci cherche à coller au plus près du progrès technique. Cette expertise d'ingénieurs brevets est également nécessaire pour appréhender la dimension systémique de l'invention *blockchain*. Cela signifie qu'un brevet individuel ne peut être qualifié de brevet *blockchain* que lorsque certains postes de la CCB sont présents simultanément dans le brevet en question. Dans ce cas, si l'on représente l'écosystème technologique de la *blockchain* sous la forme d'un réseau à la [KRA 11], c'est dans les liens entre certains nœuds du réseau que se révèle l'existence de l'invention *blockchain*, dans une combinaison de nœuds, et non pas dans la seule présence des nœuds (i.e. des postes de la CCB).

Pour [CLA 20], le travail sur la nomenclature technologique n'est cependant pas suffisant, même s'il intègre la composante systémique. Il faut lui adjoindre une dimension textuelle. Toujours avec l'expertise d'ingénieurs brevets de l'OEB, une série de mots clés sont associés aux codes de la CCB pour limiter le bruit (les faux positifs) transmis par la seule information sur la nomenclature. Les auteurs définissent des requêtes complexes qui permettent d'interroger les bases de données de brevets (dans leur cas, il s'agit de la plateforme Orbit Intelligence de l'éditeur Questel, et dans le cas de [MAR 20], il s'agit de la plateforme Derwent, de l'éditeur Clarivate) pour isoler les inventions que l'on peut attribuer le plus strictement à la *blockchain*.

Une fois le portefeuille des brevets *blockchain* identifié, [CLA 20] se contentent de présenter des résultats quantitatifs sur celui-ci : évolution temporelle des dépôts, poids des offices nationaux et internationaux, identification des déposants et relations de citations entre eux. Nous proposerons dans la section suivante d'aller plus loin dans la phase d'analyse, et le cadre dessiné par [KRA 11] sur la structure des bases de connaissances technologiques va nous aider à enrichir celle-ci.

Nous faisons le choix, dans cet article, de combiner les approches de [MAR 20] et de [CLA 20] pour identifier notre propre portefeuille de brevets « autour » de la *blockchain*. En effet, notre objectif n'est pas tant d'augmenter l'analyse du périmètre strict à la [CLA 20] par de l'ARS que d'identifier l'écosystème technologique (ou « technosystème ») dans lequel baigne l'invention *blockchain*. De ce point de vue, la démarche de [MAR 20] a l'avantage d'essayer d'analyser la *blockchain* comme *General Purpose Technology*, c'est-à-dire par son potentiel de diffusion au sein du réseau technologique global, diffusion qui se conçoit comme une capacité à se combiner à d'autres technologies pour irriguer l'activité d'invention.

Cette conception de la diffusion d'une technologie consistant à l'assimiler à la diversité des processus d'invention dans lesquels elle s'intègre est, en définitive, celle qui transparaît dans le rapport de [OEB 17] : la *blockchain* n'est nulle part en propre, mais elle s'associe à d'autres technologies qui relèvent de champs d'application multiples. D'un point de vue d'analyse économique, toujours à partir de données de brevets, cette conception de la diffusion se retrouve

chez [GKO 16], chez [MEU 21] et chez [LEB 21] : un technosystème autour d'une brique de connaissance unique est un réseau en étoile de cooccurrences de technologies au sein des brevets intégrant cette brique, réseau quelquefois simplifié pour ne prendre en compte que l'existence de relations entre technologies indépendamment de l'intensité de ces relations. [GKO 16] et [MEU 21] montrent, sur des périmètres technologiques différents, et à partir de la représentation simplifiée du réseau, que la dynamique historique des connexions des technologies de la nomenclature CIB suit une traditionnelle courbe en S (logistique) et, dans ce cadre, qu'il est possible de qualifier à un moment donné du temps chacune de ces technologies comme étant soit émergente, soit mature et, chez [MEU 21], d'anticiper le potentiel de diffusion d'une technologie sur le temps long (i.e. son ubiquité).

Chez [LEB 21], la démarche est un peu différente, puisque l'information sur les intensités des connexions au sein des technosystèmes est intégrée à l'analyse. Si l'on se centre sur une seule technologie (un seul code de la CIB, par exemple), le technosystème est donc représenté graphiquement par une étoile au centre de laquelle se situe la technologie étudiée, alors qu'aux extrémités de chaque branche de l'étoile se trouve l'ensemble des technologies avec lesquelles elle se combine dans les inventions. Les liens entre les technologies (les branches de l'étoile) sont quantifiés par le nombre de brevets au sein desquels on retrouve ces technologies ensemble. L'auteur cherche à produire un outil prédictif sur l'extension du technosystème de la technologie étudiée, i.e. l'ajout d'une branche à l'étoile. Pour cela, il identifie dans un premier temps l'ensemble des écosystèmes des technologies connectées à celle située au centre de l'étoile. Certaines technologies appartenant à ces écosystèmes ne sont pas connectées directement à la technologie évaluée. Il s'agit alors de définir des probabilités de connexion future avec ces technologies non connectées en faisant l'hypothèse que cette probabilité est d'autant plus forte que les intensités des connexions au sein des technosystèmes existants sont elles-mêmes fortes. Cette hypothèse relève de la distinction entre « liens forts » et « liens faibles » chez [GRA 73], et peut se résumer par l'analogie suivante : j'ai plus de chance à court terme de devenir ami avec les bons amis (que je ne connais pas encore) de mes amis les plus proches qu'avec ceux de connaissances plus vagues.

Donc, la démarche « large » du technosystème blockchain, telle que [MAR 20] la conçoivent, est privilégiée ici. Cependant, nous limitons cette largeur au périmètre technologique de la *blockchain* identifié par [CLA 20]. La tableau 1 reprend les 25 codes de la CCB qui sont liés à la blockchain selon ces auteurs. Rappelons que dans leur démarche, certains de ces codes doivent se combiner, ou bien s'associer à certains mots clés pour être considérés comme étant strictement *blockchain*. Nous faisons le choix de considérer ces codes isolément. C'est de cette manière que nous sommes capable de dessiner « autour » de la *blockchain* un environnement technologique riche. Par construction, l'ensemble des brevets *blockchain* identifiés par [CLA 20] sont compris dans ceux que nous repérons, puisque nous utilisons des conditions moins strictes pour isoler les brevets : pas de combinaisons de code, et absence de requête sémantique. Nous reprenons en définitive précisément la méthode mobilisée par [KRA 11] pour représenter une base de connaissances à partir de données de brevets.

Codes CCB	Intitulés
<b>H04L 9/3247</b>	Cryptographic mechanisms or cryptographic arrangements for secret or secure communication including means for verifying the identity or authority of a user of the system involving digital signatures ...
<b>H04L 9/3249</b>	... using RSA or related signature schemes, e.g. Rabin scheme
<b>H04L 9/3252</b>	... using DSA or related signature schemes, e.g. elliptic based signatures,

	ElGamal or Schnorr schemes
<b>H04L 9/3255</b>	... using group based signatures, e.g. ring or threshold signatures
<b>H04L 9/3257</b>	... using blind signatures
<b>G06Q 20/065</b>	Payment architectures, schemes or protocols, Payment circuits, Private payment circuits, e.g. involving electronic currency used among participants of a common payment scheme using e-cash ...
<b>G06Q 20/0652</b>	... with decreasing value according to a parameter, e.g. time
<b>G06Q 20/0655</b>	... managed centrally
<b>G06Q 20/0658</b>	... managed locally
<b>G06Q 20/02</b>	Payment architectures, schemes or protocols involving a neutral party, e.g. certification authority, notary or trusted third party ...
<b>G06Q 20/023</b>	... the neutral party being a clearing house
<b>G06Q 20/4014</b>	Payment protocols authorisation, identity check for transactions
<b>G06Q 20/4016</b>	Payment protocols authorisation involving fraud or risk level assessment in transaction processing
<b>G06Q 20/4018</b>	Payment protocols authorisation using the card verification value [CVV] associated with the card
<b>H04L 9/3236</b>	Cryptographic mechanisms or cryptographic arrangements for secret or secure communication including means for verifying the identity or authority of a user of the system using cryptographic hash functions ...
<b>H04L 9/3239</b>	... involving non-keyed hash functions, e.g. modification detection codes [MDCs], MD5, SHA or RIPEMD
<b>H04L 9/3242</b>	... involving keyed hash functions, e.g. message authentication codes [MACs], CBC-MAC or HMAC
<b>H04L 9/0637</b>	Cryptographic mechanisms or cryptographic arrangements for secret or secure communication, encryption apparatus using shift registers or memories for block-wise coding, Block ciphers, i.e. encrypting groups of characters of a plain text message using fixed encryption transformation, modes of operation, e.g. cipher block chaining [CBC], electronic codebook [ECB] or Galois/counter mode [GCM]
<b>H04L 9/0643</b>	Cryptographic mechanisms or cryptographic arrangements for secret or secure communication, encryption apparatus using shift registers or memories for block-wise coding, Block ciphers, i.e. encrypting groups of characters of a plain text message using fixed encryption transformation, Hash functions, e.g. MD5, SHA, HMAC or f9 MAC



<b>H04L2209/38</b>	Additional information or applications relating to cryptographic mechanisms or cryptographic arrangements for secret or secure communication, chaining, e.g. hash chain or certificate chain
<b>H04L2209/56</b>	Additional information or applications relating to cryptographic mechanisms or cryptographic arrangements for secret or secure communication, financial cryptography, e.g. electronic payment or e-cash
<b>H04L2209/30</b>	Additional information or applications relating to cryptographic mechanisms or cryptographic arrangements for secret or secure communication, compression, e.g. Merkle-Damgard construction
<b>H04L2209/46</b>	Additional information or applications relating to cryptographic mechanisms or cryptographic arrangements for secret or secure communication, secure multiparty computation, e.g. millionaire problem ...
<b>H04L2209/463</b>	... electronic voting
<b>H04L2209/466</b>	... electronic auction

**Tableau 1.** Le périmètre blockchain (codes CCB de [CLA 20])

Les codes de la CCB qui apparaissent dans le tableau 1 sont au degré de désagrégation le plus poussé de cette nomenclature. Autrement dit, les entrées *blockchain* ne représentent qu'environ 1 / 10 000e du technosystème global (25 entrées sur 260 000). Remarquons qu'à chaque entrée correspond une définition complète de la technologie correspondant au code en question. Nous avons affaire, plutôt qu'à une simple nomenclature, à un véritable dictionnaire, extrêmement précis pour que les experts des offices de brevets puissent attribuer les inventions à des domaines les moins ambigus possibles. Cette richesse dans la description conduit actuellement certains experts brevets et chercheurs à utiliser de nouvelles techniques, fondées sur l'intelligence artificielle, pour allouer automatiquement des inventions à ces codes par le traitement systématique des textes contenus dans les brevets (voir par exemple [TRA 20]), pour la constitution de leurs états de l'art [SET 21] et pour, plus globalement, améliorer le système de gestion de ces droits de propriété intellectuelle [ALD 19]. Dans cet article, nous nous en tiendrons à des données catégorielles (i.e. limitées aux informations sur les codes CCB incluses dans les brevets), sans usage de ces techniques de traitement novatrices.

### 3. Le technosystème « *blockchain* » révélé

Nous utilisons la base de données PATSTAT de l'OEB dans sa version automne 2018. Cette base recense les informations d'environ 130 millions de brevets déposés dans tous les offices nationaux ou internationaux. Nous sélectionnons les brevets déposés à l'OEB et à l'USPTO entre les années 2000 et 2015, et nous isolons l'ensemble de ceux qui contiennent au moins l'un des 25 codes CCB listés dans le tableau 1. Ces brevets sont regroupés en familles INPADOC pour éviter de compter plusieurs fois une même invention déposée dans plusieurs offices (les membres d'une famille de brevets INPADOC ont au moins une priorité en commun avec au moins un autre membre, directement ou indirectement). Ce socle constitue notre portefeuille pour l'analyse. Pour chaque brevet de ce portefeuille, nous repérons l'ensemble des codes CCB qui y sont présents. Pour construire le réseau, nous envisageons alors l'ensemble des liens possibles entre ces codes. Par exemple, s'il existe 4 technologies CCB notées A, B, C et D dans le brevet *p*, il en résulte 6 liens non dirigés entre elles : entre A et B, entre A et C, entre A et D, entre B et C, entre B et D, et entre C et D. Ces liens peuvent se trouver dans plusieurs brevets, et nous les pondérons par le nombre de

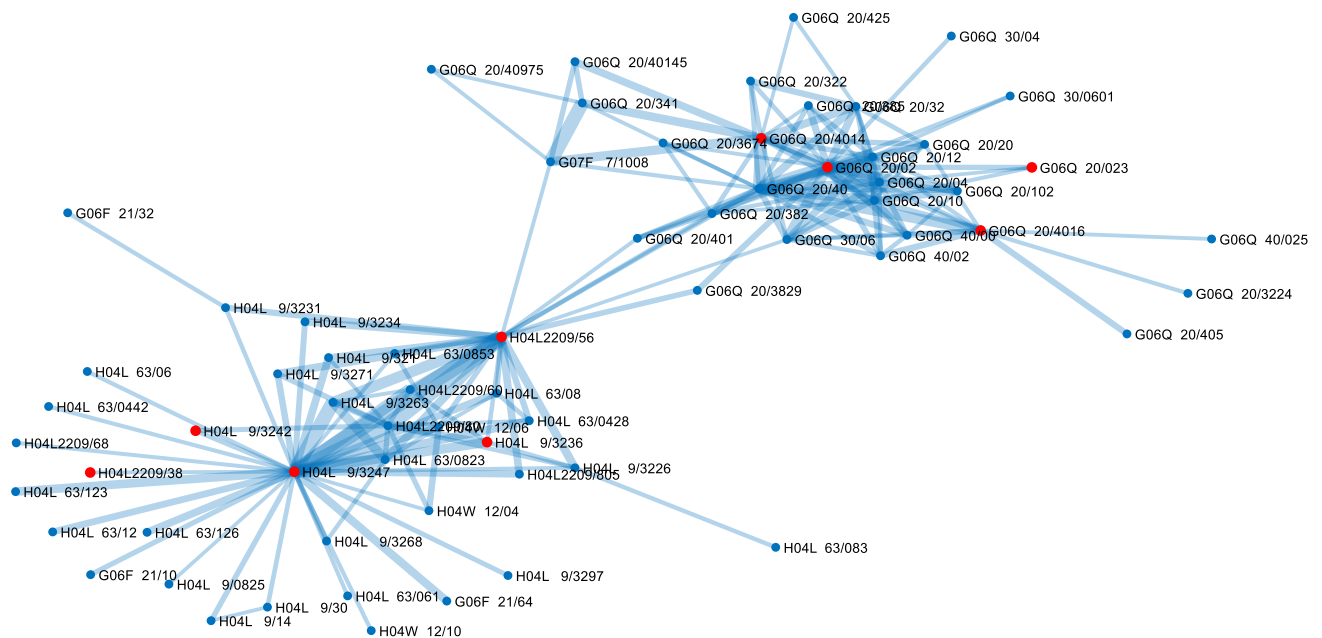
brevets dans lesquels ils apparaissent. Le tableau 2 propose des statistiques descriptives sur ce portefeuille par années de dépôt.

Année	nb_tech	nb_liens	pondér.	val_moy	val_min	val_max
2000	2031	226643	295998	1,143	1	170
2001	1743	284803	308940	1,042	1	101
2002	1325	77153	96101	1,116	1	68
2003	1362	62084	74288	1,094	1	47
2004	1290	92573	105998	1,070	1	44
2005	1750	104589	123768	1,088	1	50
2006	1077	23254	41300	1,333	1	42
2007	931	14488	28663	1,407	1	42
2008	1192	27585	44533	1,271	1	42
2009	1682	235645	252548	1,035	1	47
2010	1043	21720	30863	1,192	1	29
2011	1353	50638	64932	1,132	1	25
2012	1337	28423	46583	1,280	1	47
2013	1340	41652	58758	1,188	1	35
2014	1257	20517	38147	1,364	1	47
2015	1111	22920	43372	1,376	1	35

**Tableau 2.** *Statistiques descriptives*

Pour l'année 2010, par exemple, les 25 technologies de la CCB s'associent à 1 018 autres technologies de cette nomenclature (pour un total de 1 043 codes). Le réseau des connexions entre ces technologies comprend 21 720 liens dont la somme des pondérations est égale à 30 863, soit une valeur moyenne par lien de 1.192, avec une valeur minimale de 1 et une valeur maximale de 29. La densité du réseau, c'est-à-dire le nombre de liens effectifs sur le nombre maximal de liens possibles entre les 1 043 technologies, est donc très faible (de l'ordre de 0.00006).

Pour la phase d'analyse, nous sommons les réseaux de toutes les années et ne retenons que les liens possédant une « forte » intensité. Ici, nous considérons qu'une intensité est forte lorsque des technologies se retrouvent ensemble dans au moins 100 brevets sur la période cumulée 2000-2015. En ne tenant compte que de ces liens forts, nous obtenons le sous-réseau de la figure 1, qui comprend 67 nœuds et 181 liens. C'est précisément de cette manière que [KRA 11] procèdent pour représenter le technosystème d'innovation des biotechnologies.



**Figure 1.** *Technosystème d'innovation de la blockchain – 2000-2015 – OEB et USPTO*

Dans la figure 1, les technologies du tableau 1 apparaissent en rouge. Neuf d'entre elles sont présentes, ce qui signifie que les seize restantes ne se combinent pas de manière suffisamment intense à d'autres (du périmètre *blockchain* ou non) pour pouvoir apparaître dans cette figure. Elles semblent donc être moins au cœur des dynamiques d'invention sur la période de l'étude. Les technologies représentées par des nœuds bleus n'appartiennent pas au périmètre *blockchain*, mais font partie de son technosystème d'invention. Chaque nœud est nommé par son code CCB (pour obtenir les définitions des codes en dehors de ceux qui apparaissent dans le tableau 1, il est possible d'effectuer des requêtes sur le site <https://www.uspto.gov/web/patents/classification/cpc/html/cpc.html>).

La phase d'analyse de ce réseau comprend deux étapes. Dans un premier temps, nous nous intéressons à chacune des technologies prises individuellement et nous déterminons leur importance relative dans le réseau, i.e. leur « centralité » dans la terminologie usuelle de l'ARS. Dans un deuxième temps, nous étudions le réseau dans son ensemble et produisons des indicateurs testant l'hypothèse de « petit monde » pour le technosystème *blockchain*.

Rang	Noms	Deg	Clos	Betw	Eigen	Katz	h_ind	Lapl	RWCC	RWBC
1	H04L2209/56	0,333	0,564	0,531	0,133	10,594	6	816	0,041	0,611
2	H04L 9/3247	0,515	0,512	0,465	0,056	9,145	5	1422	0,026	0,563
3	G06Q 20/40	0,303	0,500	0,236	0,353	16,970	9	808	0,027	0,229
4	G06Q 30/06	0,152	0,462	0,094	0,239	11,617	8	402	0,018	0,137
5	G06Q 20/4016	0,182	0,369	0,090	0,221	10,726	7	380	0,016	0,149
6	G06Q 20/02	0,303	0,384	0,080	0,328	15,623	8	734	0,025	0,211
7	G07F 7/1008	0,091	0,434	0,059	0,084	5,234	4	178	0,011	0,137
8	G06Q 20/382	0,136	0,449	0,051	0,200	10,037	7	342	0,018	0,148
9	G06Q 20/4014	0,242	0,384	0,049	0,288	13,819	9	578	0,022	0,170
10	H04L 9/3226	0,076	0,415	0,030	0,024	3,561	4	180	0,009	0,093

**Tableau 3.** *Mesures de centralité des technologies (classement sur l'intermédiarité)*



	Deg	Clos	Betw	Eigen	Katz	h_ind	Lapl	RWCC	RWBC
Deg	1	0,766	0,852	0,867	0,952	0,974	0,950	0,984	0,935
Clos		1	0,777	0,504	0,665	0,718	0,839	0,797	0,877
Betw			1	0,719	0,793	0,790	0,835	0,834	0,906
Eigen				1	0,958	0,887	0,776	0,852	0,771
Katz					1	0,955	0,901	0,942	0,867
h_ind						1	0,933	0,967	0,883
Lapl							1	0,967	0,900
RWCC								1	0,927
RWBC									1

**Tableau 4.** *Corrélations de Spearman entre les indices de centralité*

Les tableaux 3 et 4 présentent des mesures de la centralité des nœuds qui apparaissent en figure 1. Dans un réseau, un nœud peut être important / central de plusieurs manières. Il revient à [FRE 79] de proposer une première tentative pour clarifier ces significations. Un nœud peut être central en fonction de son activité potentielle de communication dans le réseau. C'est la notion de « degré » (ici, la variable nommée Deg) qui reflète cette importance. Le degré se mesure comme le nombre normalisé de connexions d'un nœud dans le réseau (normalisé par le nombre maximal de connexions possibles qu'il peut entretenir). Alternativement, un nœud peut être central s'il est « proche » des autres nœuds. Cette proximité traduit le potentiel d'indépendance et d'efficacité de la communication d'un nœud, la moins perturbée possible par la multiplication de relais de transmission (ici, la variable est nommée Clos, pour *closeness*, également normalisée). Enfin, un nœud peut être important lorsqu'il est un passage obligé de communication entre les autres nœuds du réseau. Cette intermédiation est un indice du potentiel de contrôle de la communication par ce nœud (ici, la variable est nommée Betw, pour *betweenness*, normalisée pareillement).

Ces trois premiers indicateurs sont construits selon les mêmes principes : simplification de la structure des connexions pour ne prendre en compte que l'existence de celles-ci en dehors de leurs intensités, et mesure à partir de l'identification des « chemins géodésiques » entre les nœuds (i.e. des chemins les plus courts), les connexions plus longues n'intervenant pas dans les calculs. D'autres indicateurs se situent dans le respect d'au moins certains de ces principes, comme le h-index, avec lequel un nœud est considéré comme étant important s'il est connecté directement avec d'autres nœuds possédant des degrés élevés (variable ici nommée h-ind), et le laplacien, qui est une mesure d'intermédiation valable pour les structures pondérées (ici, variable nommée Lapl).

Les autres indicateurs (Eigen, Katz, RWCC, RWBC) prennent en compte l'ensemble des connexions directes et indirectes possibles entre les nœuds du réseau. RWCC et RWBC représentent respectivement la proximité et l'intermédiation lorsque les nœuds communiquent en utilisant les chemins les reliant de manière aléatoires au sein de la structure (les pondérations jouant un rôle dans la probabilité d'utiliser tel ou tel chemin). Eigen et Katz utilisent la complexité des connexions concrètes entre les nœuds et synthétisent la position de chacun d'eux dans la structure. Ces quatre derniers indicateurs possèdent donc une portée plus générale (l'ensemble de la structure des connexions est prise en compte dans la mesure de la centralité des nœuds) que ceux présentés aux

deux paragraphes précédents. [OLD 19] proposent une synthèse sur le sens et la mesure de tous ces indicateurs. Ils en présentent également beaucoup d'autres. Mais l'essentiel des informations sur l'importance relative des nœuds d'une structure est présent dans ceux que nous avons retenus. Le tableau 4 montre que ces indicateurs restent fortement positivement corrélés (corrélation de rang de Spearman) : lorsqu'un nœud est relativement important selon un critère, il y a de fortes chances qu'il le soit également selon de nombreux autres.

Le tableau 3 présente une hiérarchie des dix nœuds les plus importants selon l'indicateur Betw (centralité d'intermédierité normalisée au sens de [FRE 79]). Selon ce critère, la technologie la plus importante est H04L2209/56, qui est une technologie *blockchain* liée à la cryptographie financière (tableau 1). Cependant, cinq des dix technologies principales ne font pas partie de la *blockchain* (surlignées en gris dans le tableau 3). Elles ont trait à l'authentification des parties de la transaction (G06Q 20/40, qui est un agrégat de certaines technologies citées dans le tableau 1, mais qui déborde le strict domaine de la *blockchain*), aux modalités de transaction en e-commerce (G06Q 30/06), aux moyens de paiement avec codes (G07F 7/1008, H04L 9/3226) et aux protocoles de sécurisation des transactions (G06Q 20/382).

L'importance de H04L2209/56 en termes d'intermédierité est directement visible sur la figure 1. Imaginons que nous supprimions cette technologie du réseau. Nous aurions alors deux composantes déconnectées, incapables de communiquer entre elles. La cryptographie financière joue donc un rôle majeur dans la manière dont s'organise l'innovation dans la *blockchain*. C'est la seule brique qui soit capable de dialoguer entre deux composantes autonomes du progrès technique, l'une centrée sur des codes à dominante G06Q (systèmes de traitement de données, à droite de la figure 1), l'autre centrée sur des codes à dominante H04L (transmission d'informations numériques, à gauche de cette figure). De ce point de vue, nous pourrions dire que cette technologie est « absorbante » au sein du technosystème *blockchain*, i.e. capable d'intégrer les technologies de ces deux grands blocs.

Est-ce que ce technosystème d'innovation possède des propriétés globales particulières ? Plus précisément, peut-on le classer dans la catégorie des « petits mondes » ? [WAT 98] montrent que des réseaux de natures très diverses (sociaux, naturels, artificiels) possèdent des propriétés de petits mondes, en combinant un degré élevé de *clustering* avec une longueur moyenne des chemins géodésiques (i.e. les plus courts) entre les nœuds qui reste faible. Cette caractéristique de petit monde éloigne le réseau en étude d'une structure dite régulière, i.e. qui possède une faible hétérogénéité (le nombre de connexions de chaque nœud est équivalent) et une faible probabilité que deux nœuds choisis au hasard soient connectés l'un à l'autre. Les graphes réguliers ont tendance à avoir des chemins géodésiques de longueur importante et un niveau élevé de *clustering* (les nœuds sont connectés en groupes denses). Elle éloigne également les petits mondes d'un réseau aléatoire, qui a les propriétés inverses d'un réseau régulier : faible longueur moyenne des chemins courts, et faible présence de *clusters* de nœuds. Le tableau 5 synthétise les propriétés des différents types de réseaux, avec  $C$  un indice de *clustering* de la structure concrète ( $C_r$  pour sa valeur dans un réseau aléatoire) et  $L$  un indice de la longueur moyenne des chemins géodésiques / courts ( $L_r$  pour sa valeur dans un réseau aléatoire). Le technosystème *blockchain* a des propriétés de réseau aléatoire si  $C$  et  $L$  ont des valeurs équivalentes respectivement à  $C_r$  et  $L_r$ , des propriétés de réseau régulier si ces deux valeurs sont très supérieures à celles d'un réseau aléatoire, et de petit monde s'il se distingue de l'aléatoire par le seul indice de clustering  $C$ .

	$C = C_r$	$C > C_r$
$L = L_r$	Réseau aléatoire	Petit monde
$L > L_r$	-	Réseau régulier

**Tableau 5.** Propriétés des réseaux réguliers, aléatoires et des petits mondes

Pour sa mesure, l'indice  $C$  est égal au nombre de « triangles » dans la structure non pondérée des connexions (i.e. le nombre de fois où trois nœuds sont directement connectés entre eux) divisé par le nombre total de chemins de longueur 2 dans la structure. On mesure la possibilité que deux nœuds qui sont connectés tous les deux à un même autre nœud soient aussi connectés directement entre eux. Le triangle, dans ce cas, est l'élément premier d'un *cluster*, i.e. d'une connexion dense entre les membres d'un réseau. Plus ce nombre de triangles est important, plus la structure est clusterisée, toute chose égale par ailleurs.

L'indice  $L$  correspond donc à la moyenne de la longueur des chemins les plus courts (géodésiques) reliant les nœuds entre eux. L'algorithme qui permet de produire cette valeur repose sur la manipulation de la matrice adjacente associée au réseau des connexions. Cette matrice est carrée (disons ici de format  $n \times n$ ) et symétrique par rapport à sa diagonale principale, et comprend un « 1 » au croisement de la ligne d'un nœud et de la colonne d'un autre si ces nœuds sont directement connectés, et un « 0 » sinon. Trouver les chemins les plus courts entre les nœuds revient à mettre progressivement à la puissance 1, 2, ...,  $n - 1$  cette matrice adjacente. Nous remplissons une matrice de même format  $n \times n$  par les valeurs de ces puissances lorsque cette mise en puissance remplit d'une valeur non nulle pour la première fois les cellules correspondant aux connexions entre les nœuds. On pose qu'un nœud ne peut être en connexion avec lui-même dans cette matrice des chemins courts, et une fois remplie, on divise la somme de ses éléments par le nombre de cellules non nulles.

Pour obtenir ces valeurs pour une structure aléatoire,  $C_r$  et  $L_r$ , nous extrayons le nombre de nœuds et le nombre moyen de connexion des nœuds de la structure concrète (leur degré moyen), et nous procédons aux connexions de manière uniforme entre les nœuds en gardant le degré moyen.

C	$C_r$	L	$L_r$
0,358	0,075	2,777	2,497

**Tableau 6.** Indices des « petits mondes »

Le tableau 6 reporte les valeurs permettant d'en déduire le type de structure de la figure 1. Il s'agit bien d'un petit monde : la longueur moyenne de ses chemins géodésiques est proche d'un réseau aléatoire, alors que l'indice de *clustering* y est bien plus élevé.

Dans leur étude quantitative sur le périmètre strict des brevets *blockchain*, [CLA 20] constatent qu'il existe une séparation très nette dans les manières d'inventer entre, de façon générale, les pays occidentaux d'une part, et la Chine d'autre part.

Cette distinction occident vs Chine repose essentiellement sur les relations qu'entretiennent les déposants de brevets entre eux. Ces relations sont synthétisées par des citations de brevets, c'est-à-dire par le contenu des états de l'art dans les inventions déposées par ces acteurs. Clarke et al. (2020) constatent que ces acteurs se citent de manière très localisée, donnant naissance à deux univers d'invention très différents. La figure 2 reproduit ces univers d'invention, avec à gauche une concentration d'acteurs chinois, et à droite d'acteurs occidentaux.





industriels de la *blockchain* (notamment dans l'agriculture, le commerce électronique et l'énergie, volonté affirmée dans le plan quinquennal de 2016 ; il faudrait donc s'attendre, avec cette nouvelle impulsion, à voir le technosystème *blockchain* chinois s'enrichir très prochainement dans les données de brevets). Le résultat auquel nous aboutissons en exploitant des données un peu plus anciennes peut soutenir une hypothèse selon laquelle les trajectoires technologiques qui sont explorées dans les univers occidentaux et chinois sont similaires, faisant écho à la diversité des champs applicatifs qui y sont simultanément envisagés et des moyens en R&D qui y ont été consacrés.

#### 4. Conclusion

Les brevets commencent à être utilisés pour tracer les contours du système d'innovation de la *blockchain*. Nous avons présenté dans cet article une application de la méthode d'analyse bien éprouvée de [KRA 11] à cette technologie, en l'enrichissant à la fois du côté des sources de données employée pour délimiter le périmètre *blockchain* (nomenclature CCB plutôt que CIB) et de celui des outils d'analyse des réseaux (multiplicité des indices de centralité, petits mondes, relations de citations entre déposants de brevets). Les inventions autour de la *blockchain* sont marquées à la fois par une forme d'autonomie des réseaux d'acteurs distinguant une dynamique occidentale d'une dynamique chinoises [CLA 20], en même temps que par une similarité des trajectoires technologiques explorées dans ces deux univers.

Nous percevons deux manières d'enrichir l'analyse empirique de ces réseaux. Premièrement, comme le proposent [MAR 20] et [CLA 20], en combinant l'approche textuelle et l'approche catégorielle pour identifier le périmètre des brevets pertinents à prendre en compte. Nous avons fait le choix, ici, de nous en tenir aux catégories technologiques de la CCB. Procéder comme le proposent [CLA 20] viendrait réduire la taille de notre portefeuille. Nous obtiendrions un sous-réseau partiel (i.e. amputé de nœuds et de liens) et l'idée serait d'étudier la *blockchain* au sens strict (tel que l'envisagent ces auteurs) au sein de son technosystème d'innovation au sens large (tel que nous l'envisageons nous-mêmes). La *blockchain*, dans la perception de ces auteurs, ne se limite pas aux nœuds du tableau 1. Elle comprend également un ensemble de connexions jugées nécessaires. Le contenu textuel des brevets viendrait ajouter un troisième niveau de complexité dans le traitement et l'analyse, cernant le bruit introduit par la seule donnée catégorielle. En définitive, trois niveaux d'étude s'interpénétreraient : le réseau tel que nous l'avons nous-même défini, celui prenant en compte les seules combinaisons technologiques, puis celui intégrant en plus les éléments textuels.

La deuxième extension que nous pouvons envisager est de prendre en compte les données de citations. Nous pourrions ainsi définir les processus d'invention de la *blockchain*, augmentant le périmètre des technologies jugées nécessaires à maîtriser pour développer un technosystème cohérent et efficace. A notre connaissance, il n'existe pas encore d'étude centrée sur un système technique utilisant des données de brevets qui combine l'analyse des cooccurrences technologiques (comme c'est le cas dans cet article) et les flux de citations qui produisent cette cooccurrence. Cela conduit à concevoir un réseau qui associe, entre les briques de connaissances, des liens non dirigés et pondérés, comme nous l'avons envisagé ici, en même temps que des liens dirigés et pondérés comme l'envisage traditionnellement la représentation des processus d'invention sous la forme de réseaux [DOS 20].

#### Bibliographie

- [ALD 19] ALDERUCCI D., SICKER D., « Applying artificial intelligence to the patent system », *Technology & Innovation*, vol. 20, p. 415-425, 2019.
- [CLA 20] CLARKE N.S., JÜRGENS B., HERRERO-SOLANA V., « Blockchain patent landscaping: An expert based methodology and search query », *World Patent Information*, vol. 61, 101964, 2020.



- [DOS 20] DOSSO M., LEBERT D., « The centrality of regions in corporate knowledge flows and the implications for Smart Specialisation Strategies », *Regional Studies*, vol. 54, p. 1366-1376, 2020.
- [FRE 79] FREEMAN L.C., « Centrality in social networks: conceptual clarification », *Social Networks*, vol. 1, p. 215-239, 1979.
- [GKO 16] GKOTSIS P., VEZZANI A., « The combinatorial dimension of technological diffusion: evidence from patent data », *Innovation Forum*, RNI, Paris, 2016.
- [GRA 73] GRANOVETTER M.S., « The strength of weak ties », *American Journal of Sociology*, vol. 78, p. 1360-1380, 1973.
- [KRA 11] KRAFFT J., QUATRARO F., SAVIOTTI P.-P., « The knowledge-base evolution in biotechnology: a social network analysis », *Economics of Innovation and New Technology*, vol. 20, p. 445-475, 2011.
- [LEB 21] LEBERT D., « Sur les modèles structuraux de diffusion technologique », *Club Patent Analytics*, INPI, 3 mars, 2021.
- [MAR 20] MARKU E., CASTRIOTTA M., LOI M., DI GUARDO M.C., « General Purpose Technology: the blockchain domain », *International Journal of Business and Management*, vol. 15, p. 192-205, 2020.
- [MEU 21] MEUNIER F.-X., « La diffusion des technologies d'ingénierie biomédicale : évidences sur données des brevets », *Technologie et Innovation*, vol. 6, n° 2, p. 1-12.
- [OEB 17] OEB, *Patents in the Fourth Industrial Revolution*, European Patent Office, 2017.
- [OLD 19] OLDHAM S., FULCHER B., PARKES L., ARNATKEVICIUTE A., SUO C., FORNITO A., « Consistency and differences between centrality measures across distinct classes of networks », *PLoS ONE*, vol. 14, e0220061, 2019.
- [SET 21] SETCHI R., SPASI I., MORGAN J., HARRISON C., CORKEN R., « Artificial intelligence for patent prior art searching », *World Patent Information*, vol. 64, 102021, 2021.
- [TRA 20] TRAPPEY A.J.C., TRAPPEY C.V., WU J.-L., WANG J.W.C., « Intelligent compilation of patent summaries using machine learning and natural language processing techniques », *Advanced Engineering Informatics*, vol. 43, 101027, 2020.
- [WAS 94] WASSERMAN S., FAUST K., *Social network analysis: Methods and applications*, Cambridge University Press, Cambridge (Mass.), 1994.
- [WAT 04] WATTS D.J., « The “new” science of networks », *Annual Review of Sociology*, vol. 30, p. 243-270, 2004.
- [WAT 98] WATTS D.J., STROGATZ S.H., « Collective dynamics of “small-world” networks », *Nature*, vol. 393, p. 440-442, 1998.
- [ZHA 20] ZHAO R., « The blockchain ecosystem in China », *OECD On The Level*, 22 juillet, 2020.