Identifying Emerging Trends of Blockchain Technology Using a Topic-based Patent Mining Model

Chia Hung Liu, Juite Wang

AGENDA

01 Introduction

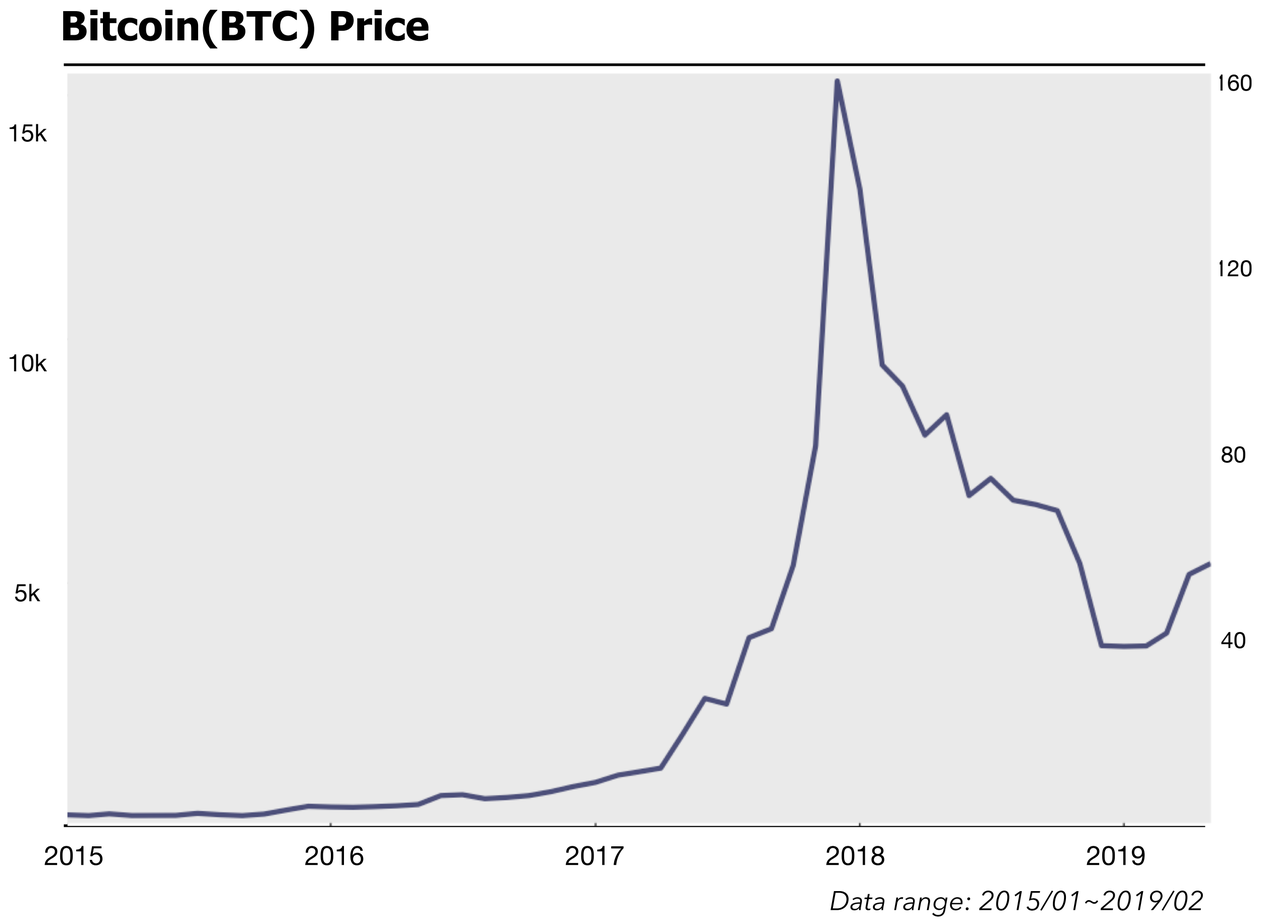
02 Literature Review

03 Methodology

04 Results

05 Conclusion

Introduction

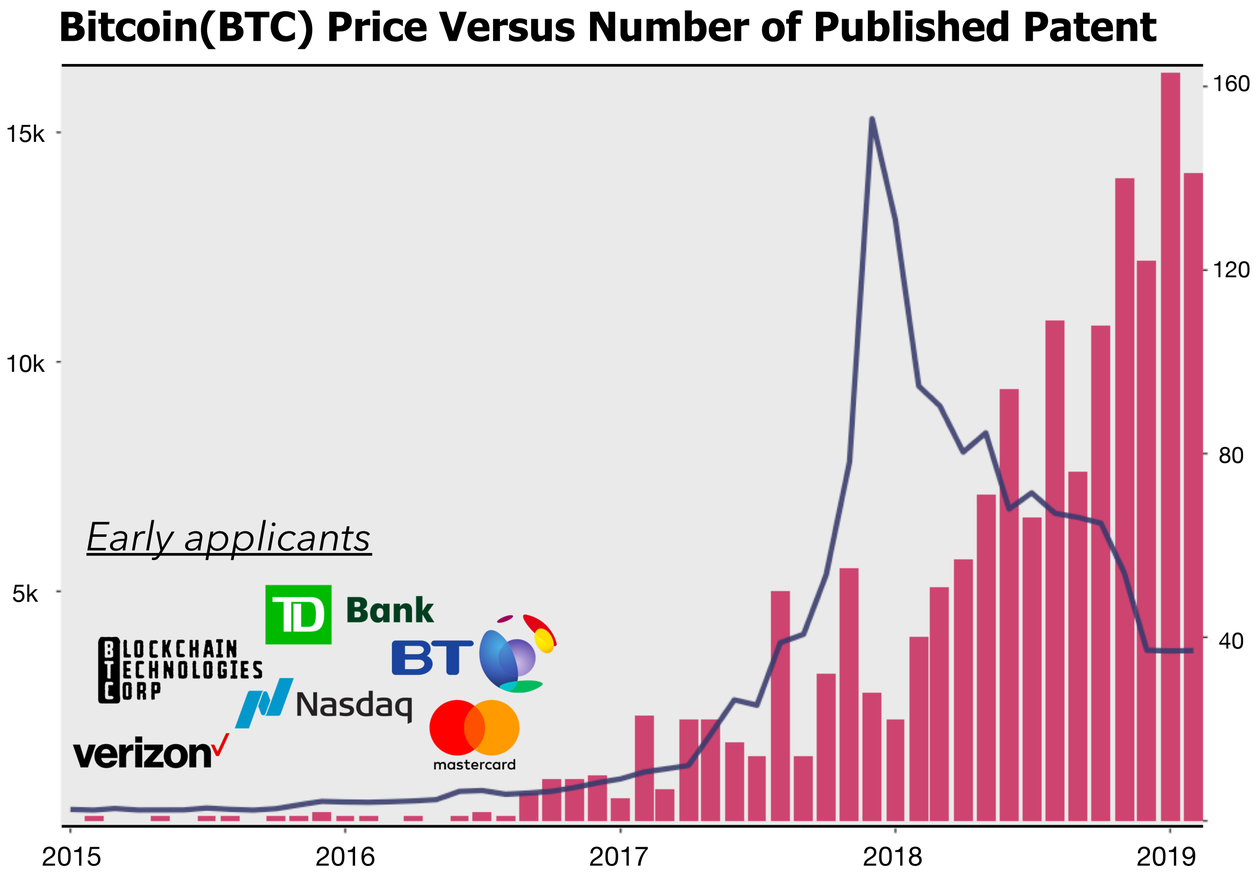


Since Satoshi Nakamoto coined the Blockchain idea in 2008, the interest of this technology has been increasing, especially at the end of 2018 as the price of Bitcoin rose dramatically.

In recent years, this promising technology has been gaining increasing attention by integrating with existing **business processes**, including **financial transactions**, **enterprise resource planning**, **supply chain management**, and so on.

Introduction

In this research, instead of taking the technical aspect to examine Blockchain technology, we proposes an application-oriented classification through patent mining to discover the underlying topics to further identify emerging trends for exploring potential technological opportunity.



Literature Review

PCT Biblio. Data | Description | Claims | Drawings | National Phase | Notices | Documents

Latest bibliographic data on file with the International Bureau

PermaLink

Pub. No.: WO/2015/024129 **International Application No.:** PCT/CA2014/050805
Publication Date: 26.02.2015 **International Filing Date:** 21.08.2014
IPC: ***G06F 21/10*** (2013.01) ,***G06F 21/16*** (2013.01) ,***H04L 12/58*** (2006.01) ,***H04L 12/16*** (2006.01)
Applicants: **ASCRIBE GMBH** [DE/DE]; Wichertstr 17 10439 Berlin, DE
Inventors: **MCCONAGHY, Trent, Lorne**; DE
MCCONAGHY, Maria; DE
Agent: **ALLARD, Louis**; CA
Priority Data: 61/868,256 21.08.2013 US
Title **(EN)** METHOD TO SECURELY ESTABLISH, AFFIRM, AND TRANSFER OWNERSHIP OF ARTWORKS
(FR) PROCÉDÉ POUR ÉTABLIR, CONFIRMER ET TRANSFÉRER DE MANIÈRE SÉCURISÉE LA PROPRIÉTÉ D'ŒUVRES D'ART
Abstract:

(EN) A method for establishing, confirming, and transferring ownership of digital artworks. The method associates a given artwork to a unique public ID and to an online service. The owner of the artwork is the person who has full access to the online service at the public ID. Confirming ownership of the artwork includes the owner performing an action that demonstrates the owner has full access to the online service at the public ID. Transferring ownership of the artwork includes (1) transferring access to the account itself, or (2) using the online service's transfer protocol, which may also update the public ID associated with the owner of the artwork.
(FR) La présente invention concerne un procédé pour établir, confirmer et transférer la propriété d'œuvres d'art numériques. Le procédé associe une œuvre d'art donnée à un identifiant (ID) public unique et à un service en ligne. Le propriétaire de l'œuvre d'art est la personne qui peut accéder sans restriction au service en ligne au niveau de l'ID public. La confirmation de la propriété de l'œuvre d'art comprend l'exécution, par le propriétaire, d'une action qui démontre que le propriétaire a un accès sans restriction au service en ligne au niveau de l'ID public. Le transfert de la propriété de l'œuvre d'art comprend (1) le transfert d'un accès au compte lui-même, ou (2) l'utilisation du protocole de transfert du service en ligne, qui peut également mettre à jour l'ID public associé au propriétaire de l'œuvre d'art.

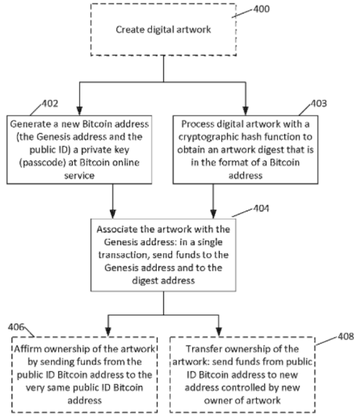


Figure 4

Designated States: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW
African Regional Intellectual Property Organization (ARIPO) (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW)

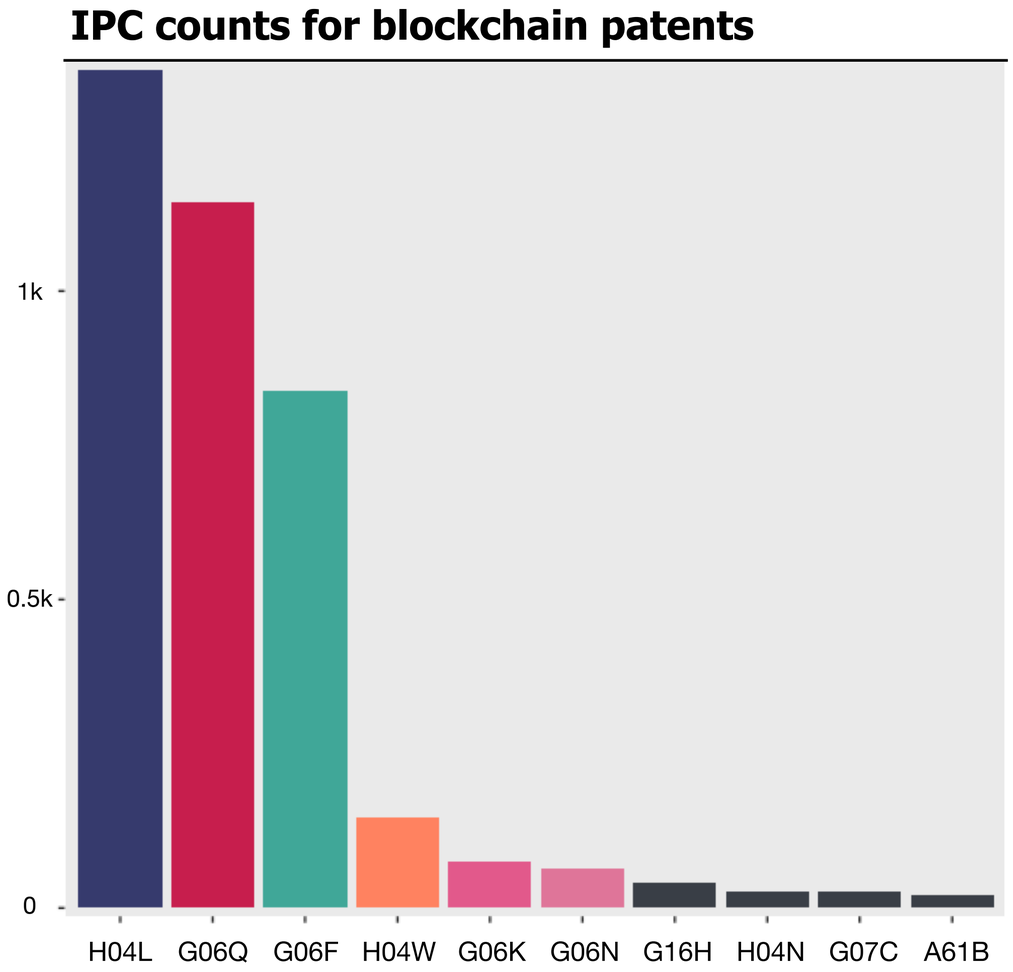
According to the statistics from WIPO, 90 ~ 95% of world's inventions can only be found in patented documents and 80% of these techniques do not appear in other professional articles.

Besides, due to information overload, text mining and natural language process comes in handy to discovery important information from patent corpus (Tseng et al., 2007).

Columns used in our research

Source: https://patentscope.wipo.int/search/en/detail.jsf?docId=WO2015024129

Literature Review



International Patent Classification is a hierarchical classification system, used primarily to classify and search patent documents according to the technical fields they pertain.

H - ELECTRICITY
H04 - ELECTRIC COMMUNICATION TECHNIQUE
H04L - TRANSMISSION OF DIGITAL INFORMATION
H04W - WIRELESS COMMUNICATION NETWORKS

G - PHYSICS
G06 - COMPUTING; CALCULATING; COUNTING
G06Q - DATA PROCESSING SYSTEMS OR METHODS
G06F - ELECTRIC DIGITAL DATA PROCESSING

While IPC code is more **function-oriented** instead of application-oriented, it may hardly match IPC code in reality. It seems hard to classify or interpret patents according to IPC in more specific field. 😔

Source: https://www.wipo.int/classifications/ipc/ipcpub/

Methodology - LDA

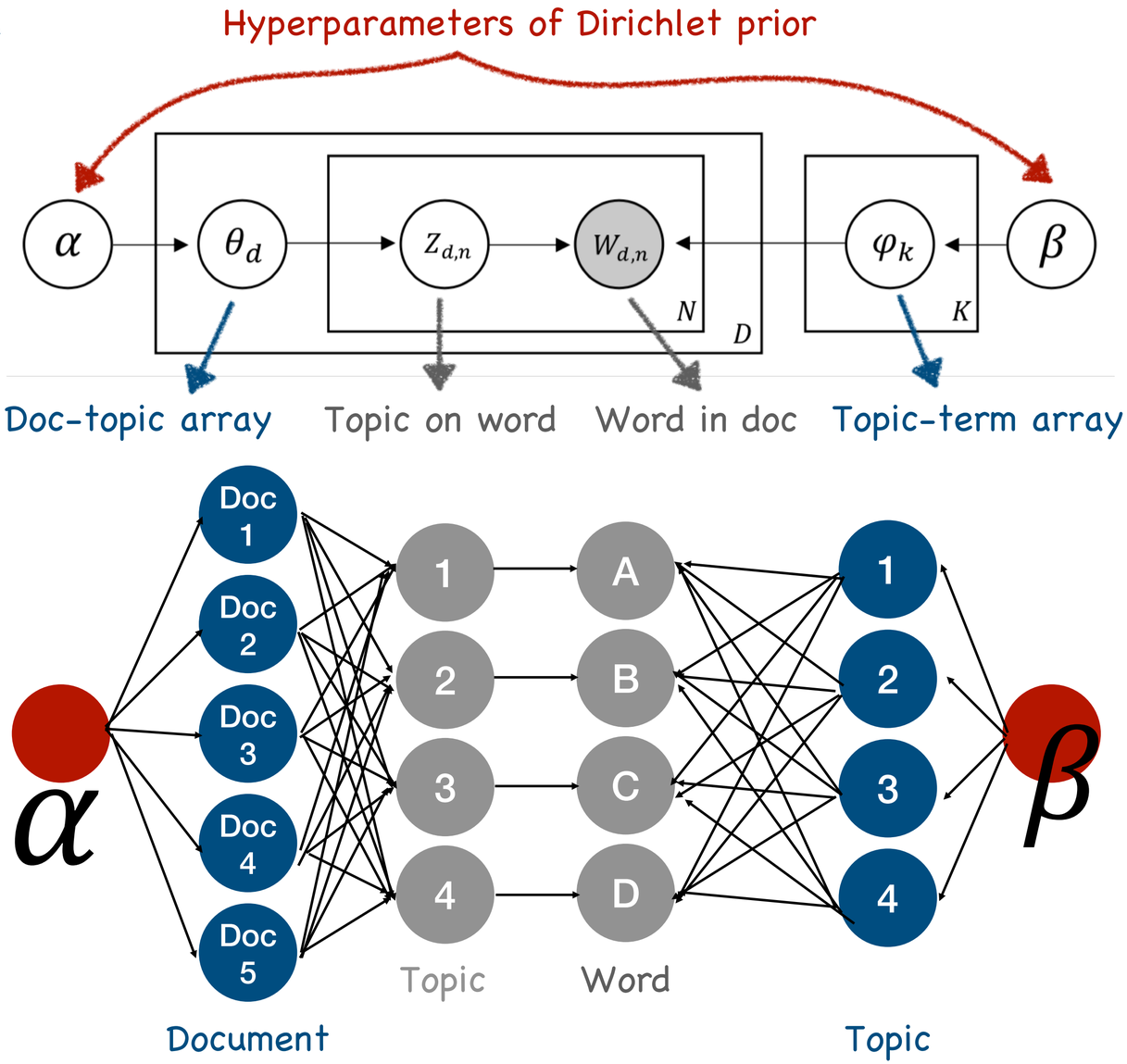
Latent Dirichlet allocation was proposed by Blei(2003), and widely used in several tasks to discover the main topic that pervade a large collection of documents, in order to facilitate information retrieval.

LDA could be seen as PCA in text mining field, but using bayesian approach (and it's also a generative model instead of discriminative one) and…

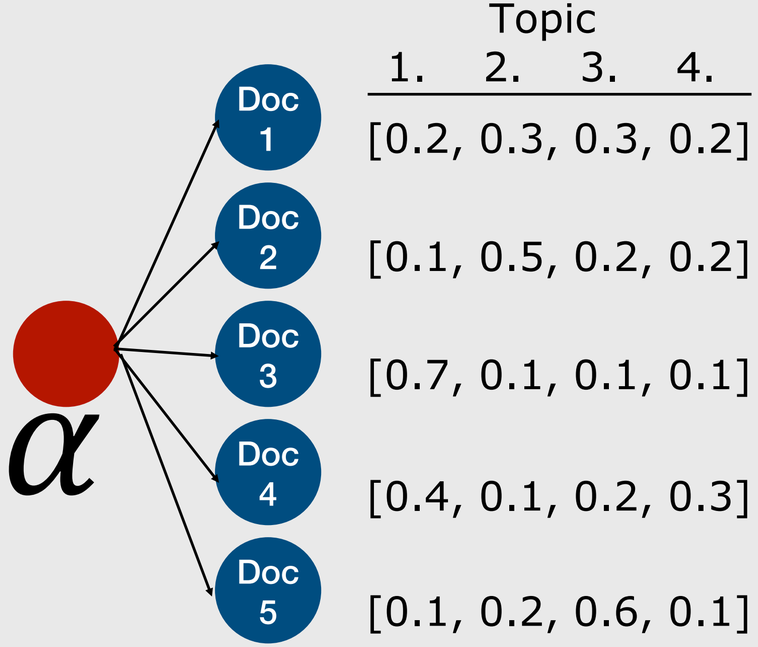
- Every documents is a mixture of topics
- Every topic is a mixture of words

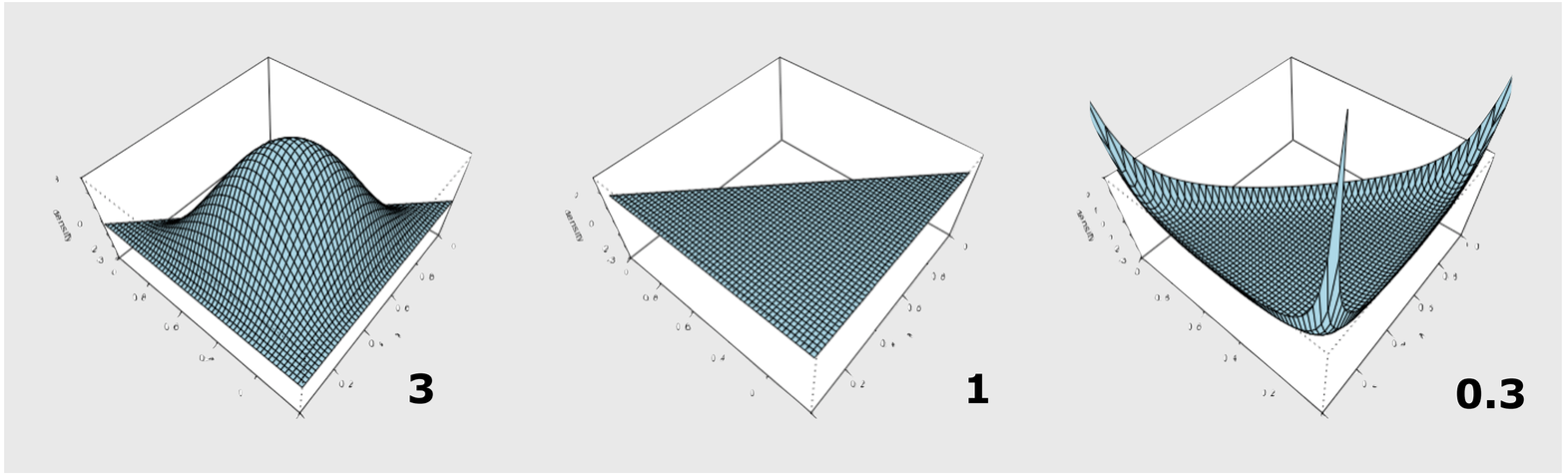
Like PCA, we need to define the number of clusters in the beginning, the hyperparameters setting(alpha, beta, number of topics) is also necessary in LDA. So, it's crucial to get sense on these greek symbols.

Plates notation for LDA (adapted from Blei and Lafferty (2009))



Methodology - hyperparameter





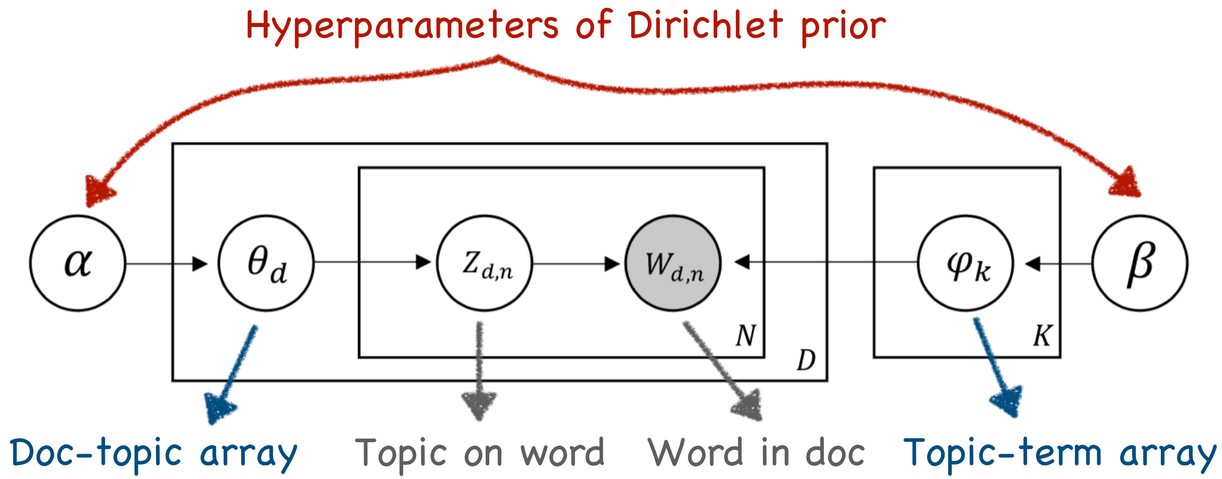
Two-simplex (K=3) visualization of Dirichlet distribution density plots listed above generated by alpha = (3, 3, 3), (1, 1, 1), (0.3, 0.3, 0.3) from the left side to the right.

Each probability distribution is sampled from Dirichlet distribution, which controlled by alpha hyperparameter! On the other hand, beta hyperparameter controls Dirichlet of topic-term array.

The figure above shows the probability density distribution of symmetric Dirichlet distribution, the more peak area the higher chance to be drawn.

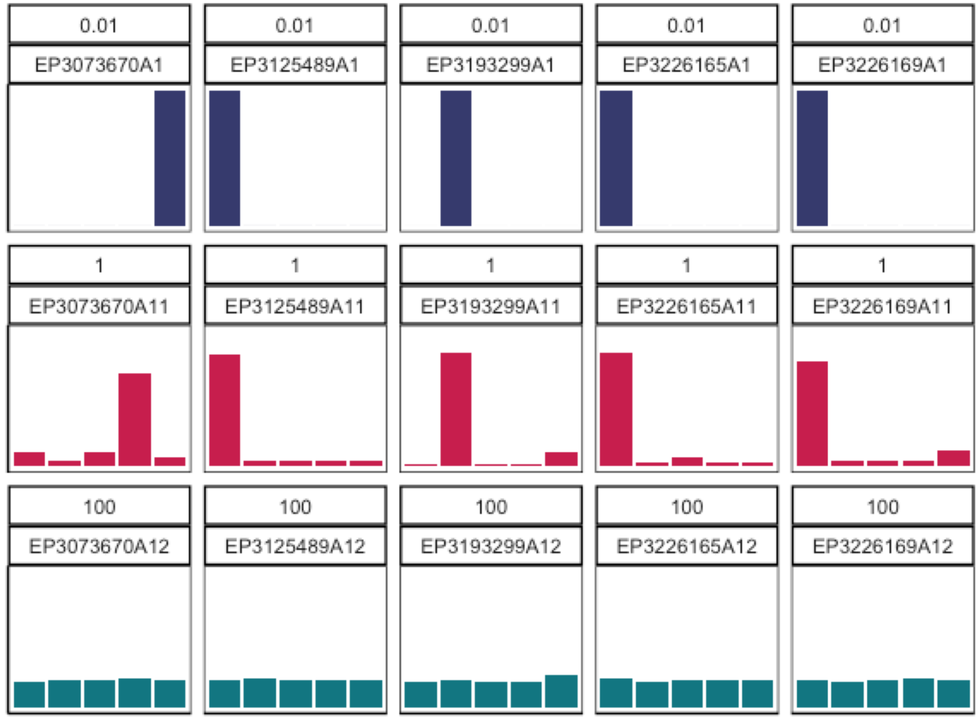
Namely, the doc-topic matrix will be more sparse by lowing the alpha value.

Methodology - hyperparameter



For example, if we have several documents need to be assigned to 5 topics, we could observe the distribution of documents over topics through lowering the alpha(doc-topic) value.

The intuition behind tuning hyperparameter is that when lowering value of alpha and beta, we assume **the more decisive topic association**.



Research framework

Define Search Keyword
Blockchain@TI, AB, CL
(Yli-Huumo et al. 2016)

EPO
WIPO
USPTO

Title,
abstract

Textual data preprocessing

Tokenization | Part-of-speech | Lemmatize | Noun phrase

Textual data cleaning

Patent freq used word | TF-IDF exclusion | Combine phrase

Topic modelling

Hyperparameter tuning | LDA modelling

1,397 patents & 1,237 phrases

Structured Data

Public date | Applicant | Country

Citation | Patent Family | ...

Topic-term matrix

D...x

Emerging topic analysis

Competitive analysis

Technology value analysis

Technology evolution map

Results - Topic model

Instead of default value of alpha(50/k) and beta(0.1), we conduct the grid search to measure the quality of mode, and the range of grid search was suggested by Zhong(2015) was between 0.01 to 0.1.

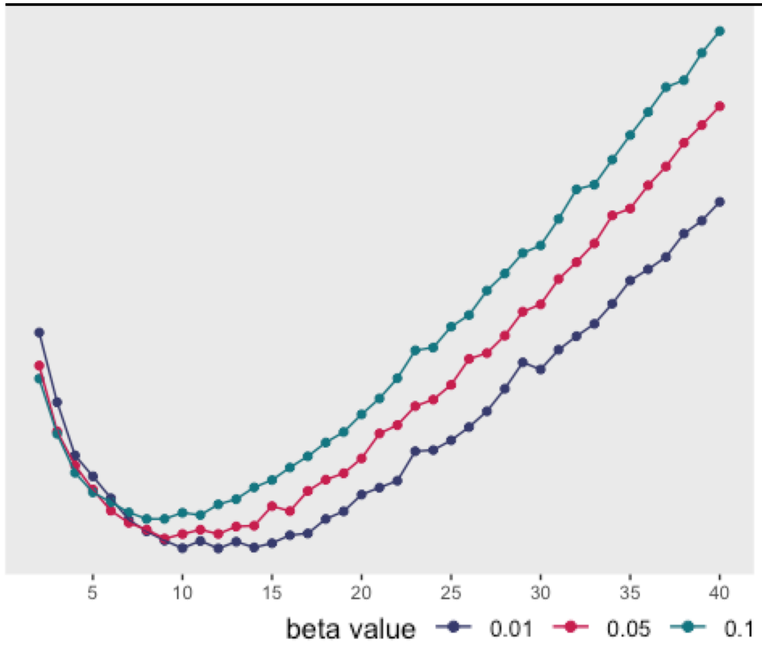
In this research, we applied Akaike information criterion (AIC) as a model quality estimator, and choosing the minimum AIC value point of topic number.

The shortcoming of LDA(or other text mining techniques) is that the model validation is mainly **subjective**.

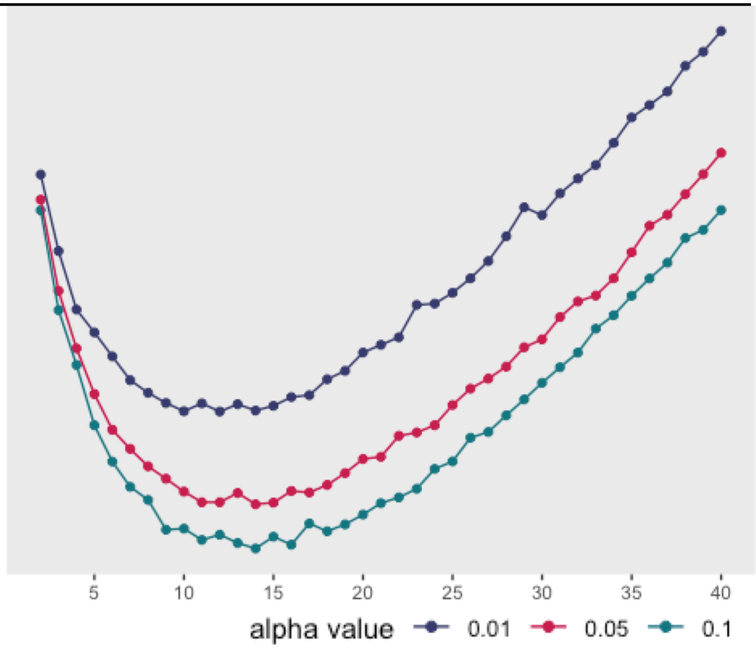
$$AIC = 2k - 2ln(L)$$

Grid Search of hyperparameters

with fixed alpha 0.01(doc-topic)



with fixed beta 0.01(topic-term)



Final parameters setting in LDA: alpha = 0.1, beta = 0.01, k = 12

Results - Topic interpretation

Word cloud for featured phrases

Patent number	Title
US20170279783A1	secure 3d model sharing using distributed ledger
US20170103167A1	blockchain system for natural language processing
WO2017207717A1	validating blockchain transactions regarding real money
US20180219683A1	possession and alteration of documents
US09855785B1	digitally encoded seal for document verification
EP3361433A1	system and method for interacting devices from different islands of trust
EP3285248A1	blockchain-based security threat detection method and system
WO2018134602A1	a method for resource allocation in a utility service network
WO2018104276A1	master blockchain

After examining the featured phrases from topic-term matrix and most related patents from doc-topic matrix, we could label the topic according them. In addition, we also interviewed related expert to evaluate the classification results. Taking topic1 for example, we label it with **“Document and payment verification related technologies on Blockchain”.**

Results - Topic interpretation

Topic label	Patent quantity
Topic 1 - Document and payment verification related technologies on Blockchain	134.19
Topic 2 - Database management and data storage based on Blockchain related technologies	97.91
Topic 3 - Consensus system building and digital asset management related technologies on Blockchain	118.70
Topic 4 - Cross-chain transaction related technologies on Blockchain	105.34
Topic 5 - Database information processing based on Blockchain related technologies	118.58
Topic 6 - Integrate information across various resources to Blockchain related technologies	109.20
Topic 7 - Identity management related technologies on Blockchain	106.21
Topic 8 - Identity verification related technologies on Blockchain	115.34
Topic 9 - Healthcare applications using Blockchain and electronic currency management related technologies	115.22
Topic 10 - Database access authentication and data synchronization related technologies on Blockchain	121.54
Topic 11 - Cryptocurrency payment, token distribution and Blockchain security related technologies	125.96
Topic 12 - Digital asset management using smart contract related technologies	128.80

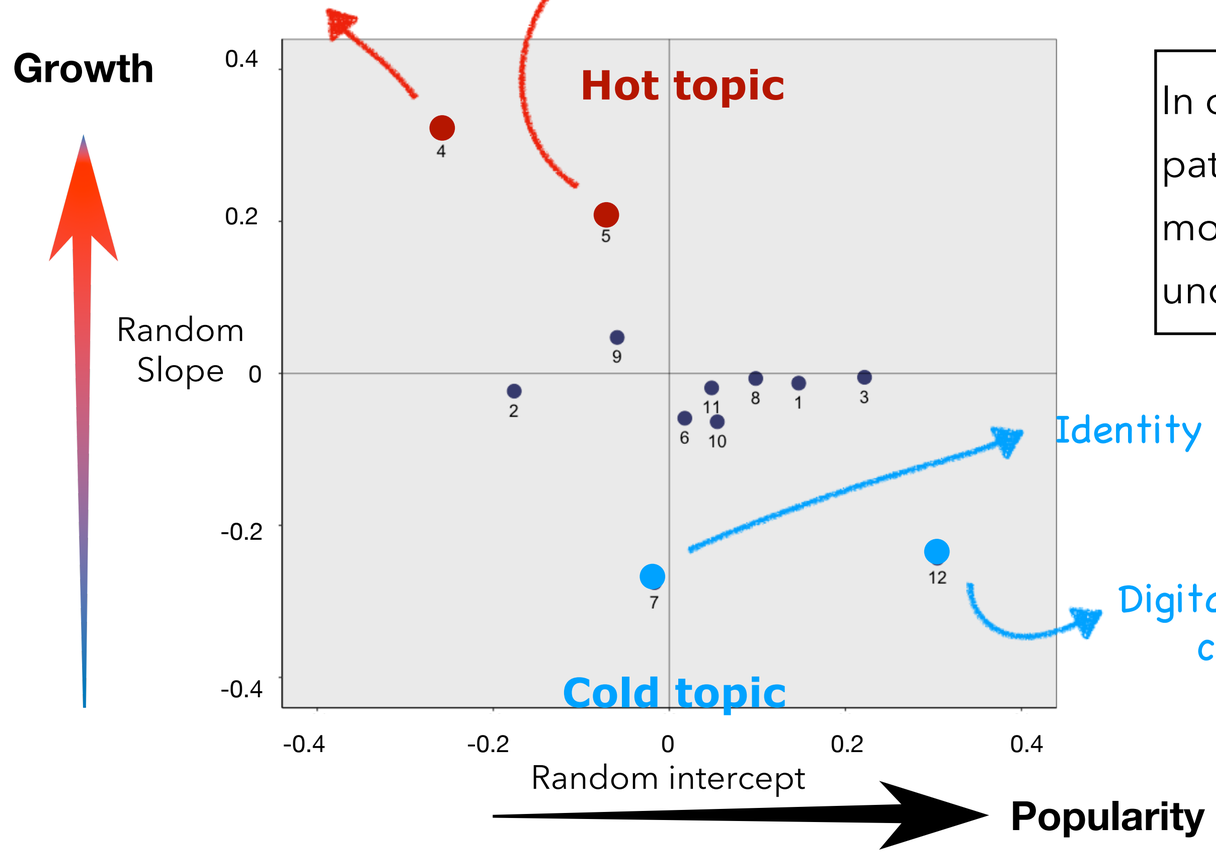
The same process for the other topics, however sometimes we need to reconsider whether the textual preprocessing and cleaning step have really extract the meaningful phrases.

On the other hand, we can observe the different patent quantity of each topic by lowing the alpha and beta value.

Results - Emerging topics

Cross-chain transaction related technologies on Blockchain

Database information processing based on Blockchain related technologies



In order to match the hierarchical structure of topic-patent data, we applied the generalized linear mixed model (GLMM) to identify the hot and cold topics underlying patents through time series.

Identity management related technologies on Blockchain

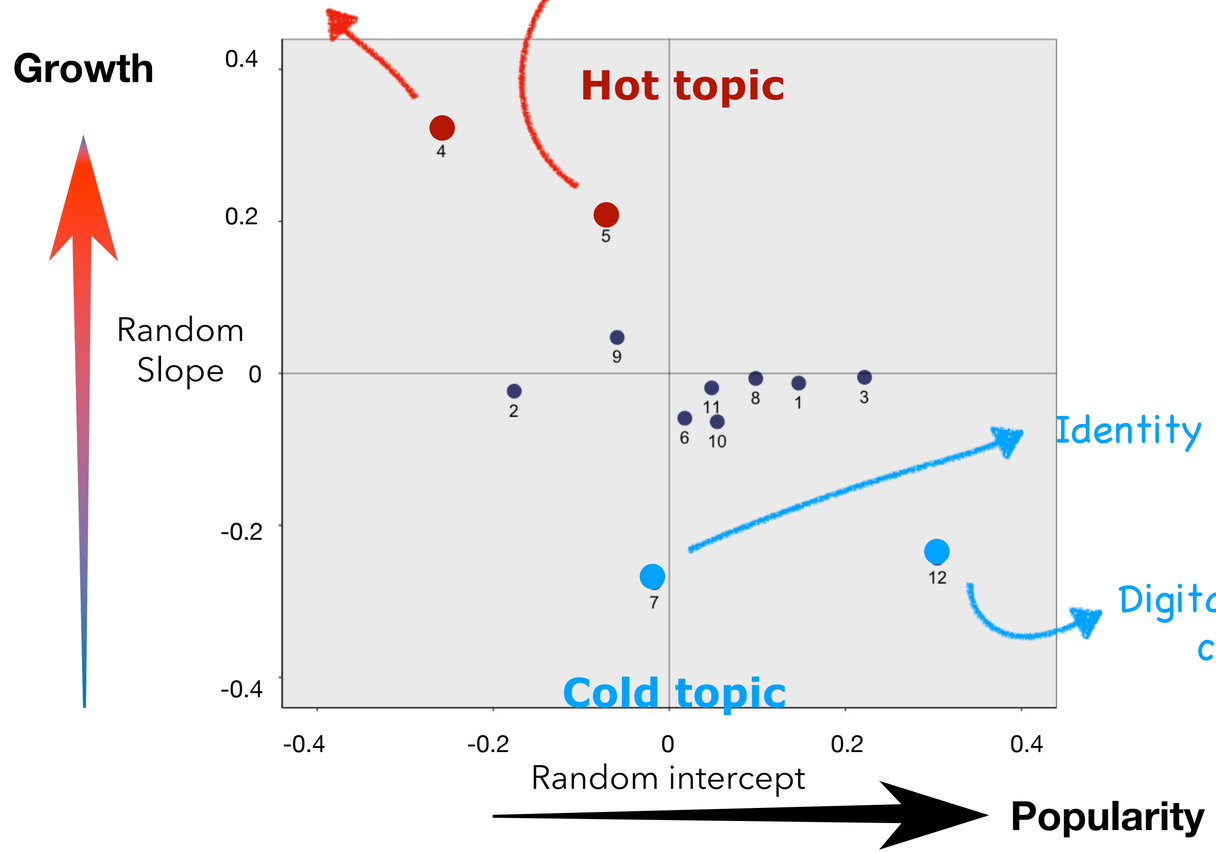
Digital asset management using smart contract related technologies

$$Y_{mt} = \beta_0 + S_{0m} + (\beta_1 + S_{1m})X_t +$$

Results - Emerging topics

Cross-chain transaction related technologies on Blockchain

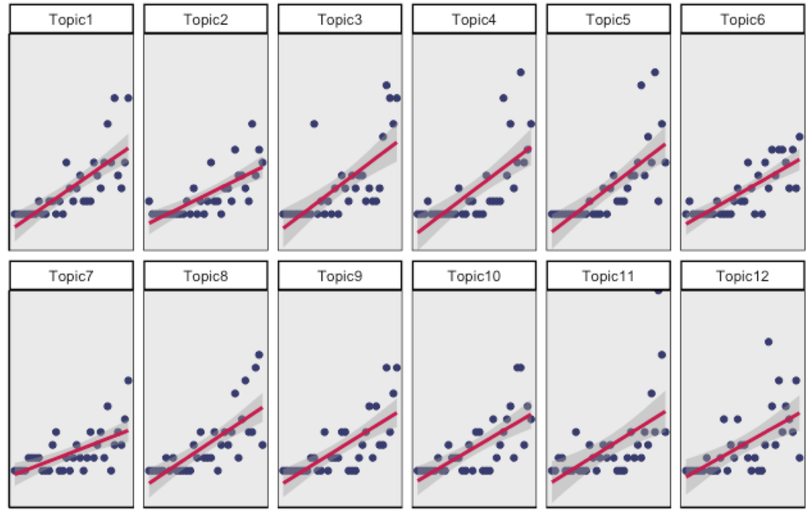
Database information processing based on Blockchain related technologies



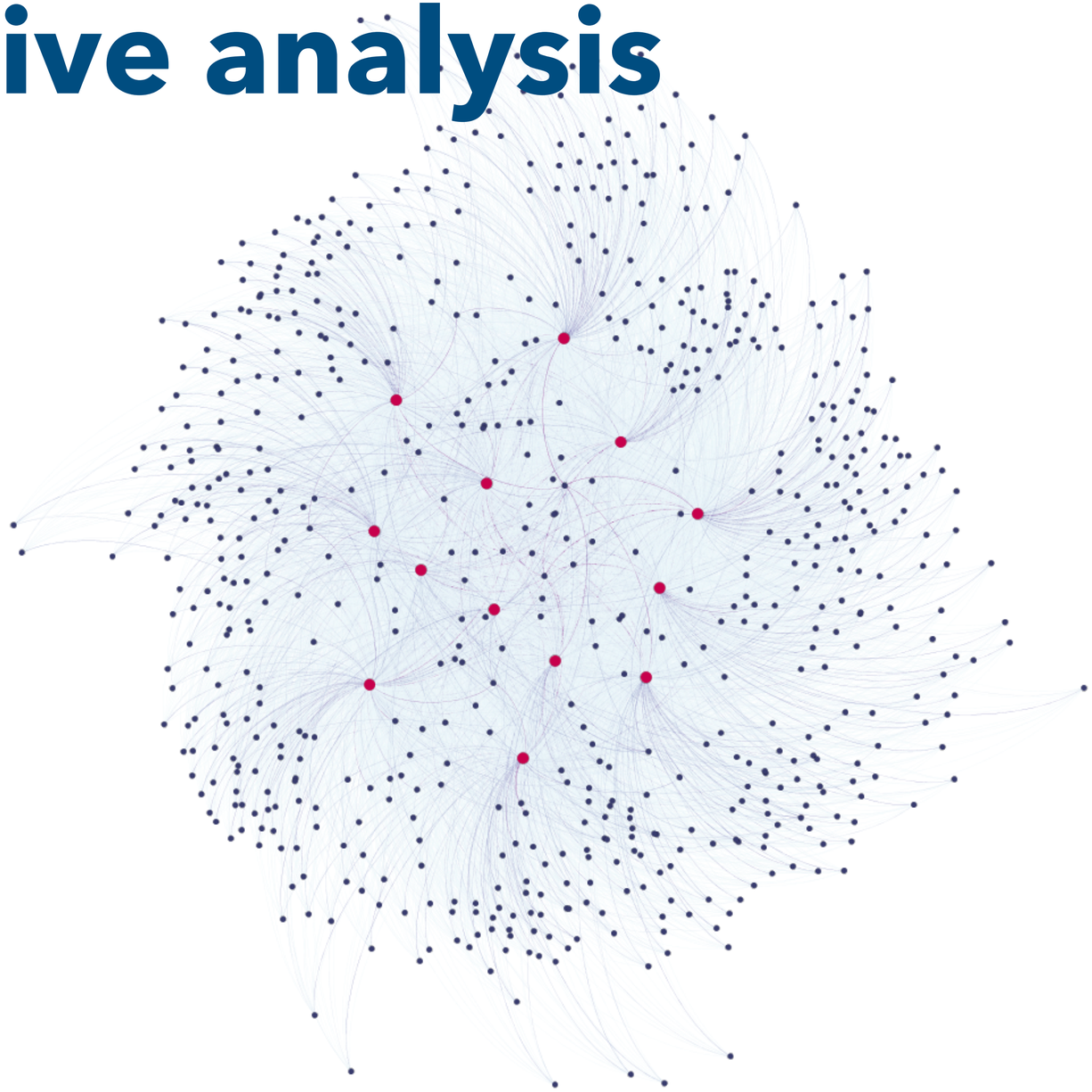
Identity mana

Digital ass
contr

Caveat: GLMM result means the relative hot or cold topics compared to mean slope, the figure below shows the linear regression on each topic.

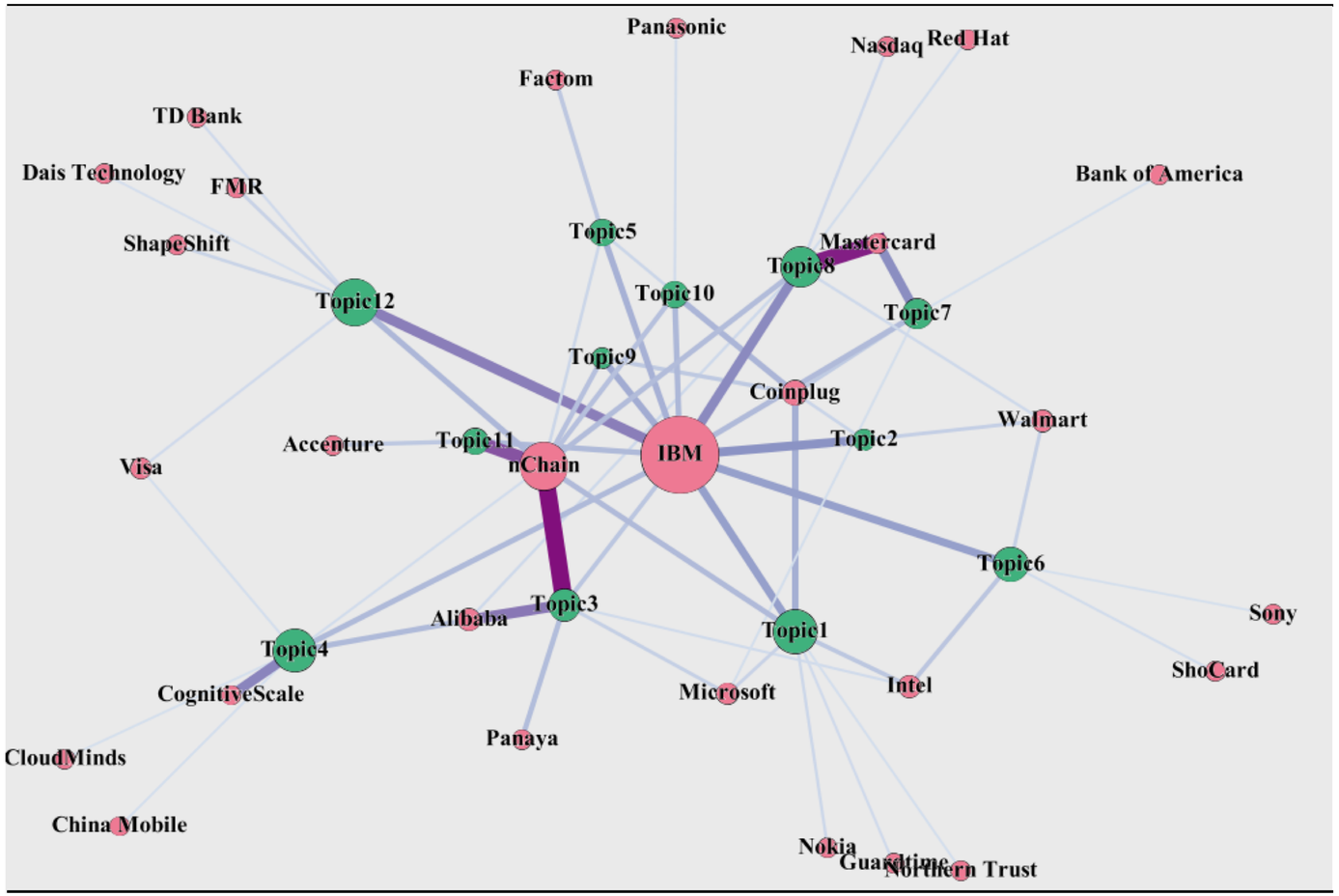


Results - Competitive analysis



Results - Competitive analysis

Topic-applicant network



$$g(v) = \Sigma_{s \neq v \neq t} \frac{\sigma_{st}(v)}{\sigma_{st}}$$

By adopting ForceAtlas 2 layout, we can further investigate the company patent strategy whether focusing on certain topic or not.

In addition, a node with a high betweenness centrality means that the node takes crucial position and owns more diverse technologies in Blockchain field. As shown in left, IBM, nChain, and Coinplug are positioned in the center of network holding the most diversification topics of patents.

Results - Technology value analysis

$$Technologyvalue = z\big(f(citationcount, year)\big) + z\big(\ln(claimcount)\big) + z\big(ln(patentfamilysize)\big)$$

$$f(citationcount, year) = \frac{citationcount}{theaveragecitationcountofthatyear}$$

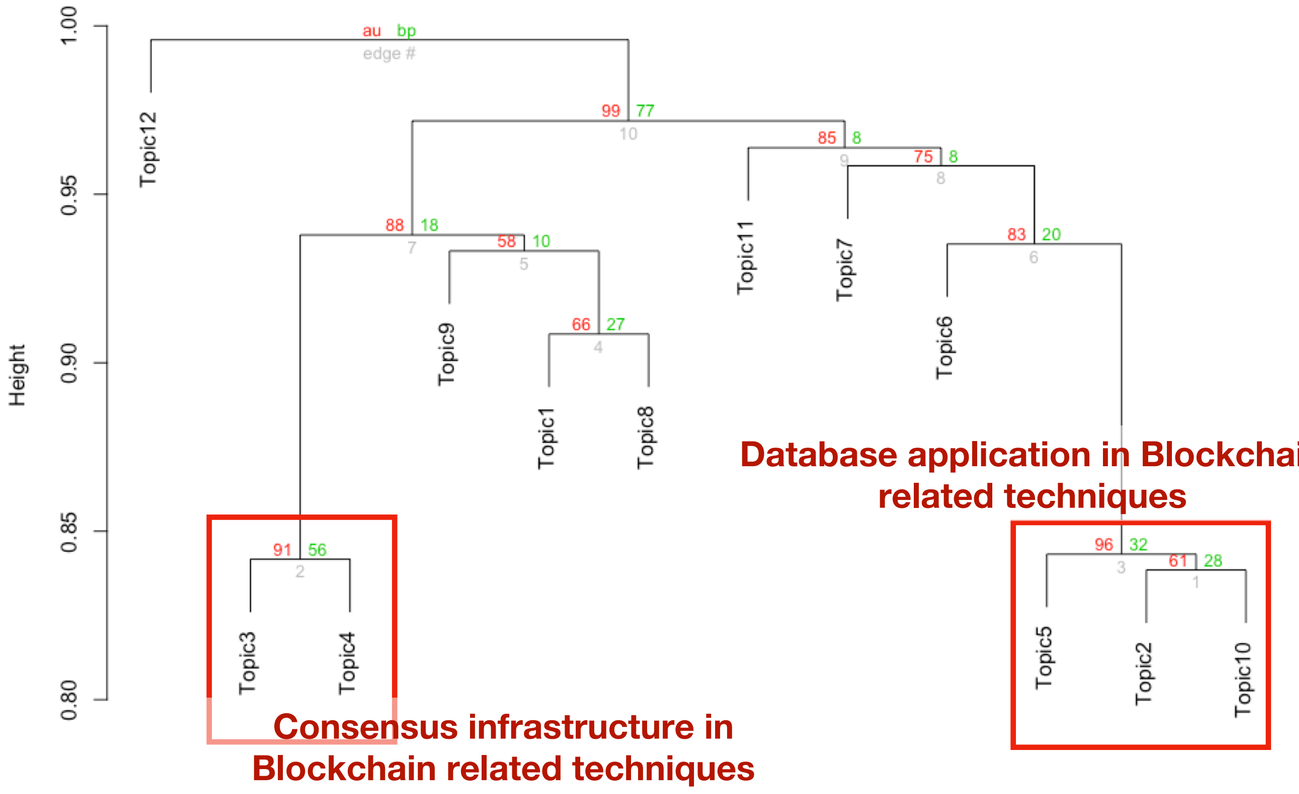
$$z(x) = \frac{x_i - x_{min}}{x_{max} - x_{min}}$$

- Forward citations, which reflect the economic value of inventions and influence power
- Claims, which determine the breadth of the rights granted by patents (the scope of protection).
- Patent family size, which refers to the number of countries in an invention is protected by patents or similar technical content.

Topic	Value	Amount of topic	Average value
Topic1	11.46	69.10	0.7712
Topic2	7.88	52.65	0.7415
Topic3	12.55	73.67	0.7392
Topic4	10.13	63.65	0.7607
Topic5	10.22	69.07	0.7257
Topic6	10.24	60.56	0.7350
Topic7	8.78	49.93	0.7571
Topic8	9.10	65.79	0.7172
Topic9	10.28	59.29	0.7548
Topic10	10.69	60.24	0.7539
Topic11	11.13	62.74	0.7677
Topic12	12.76	68.32	0.7641

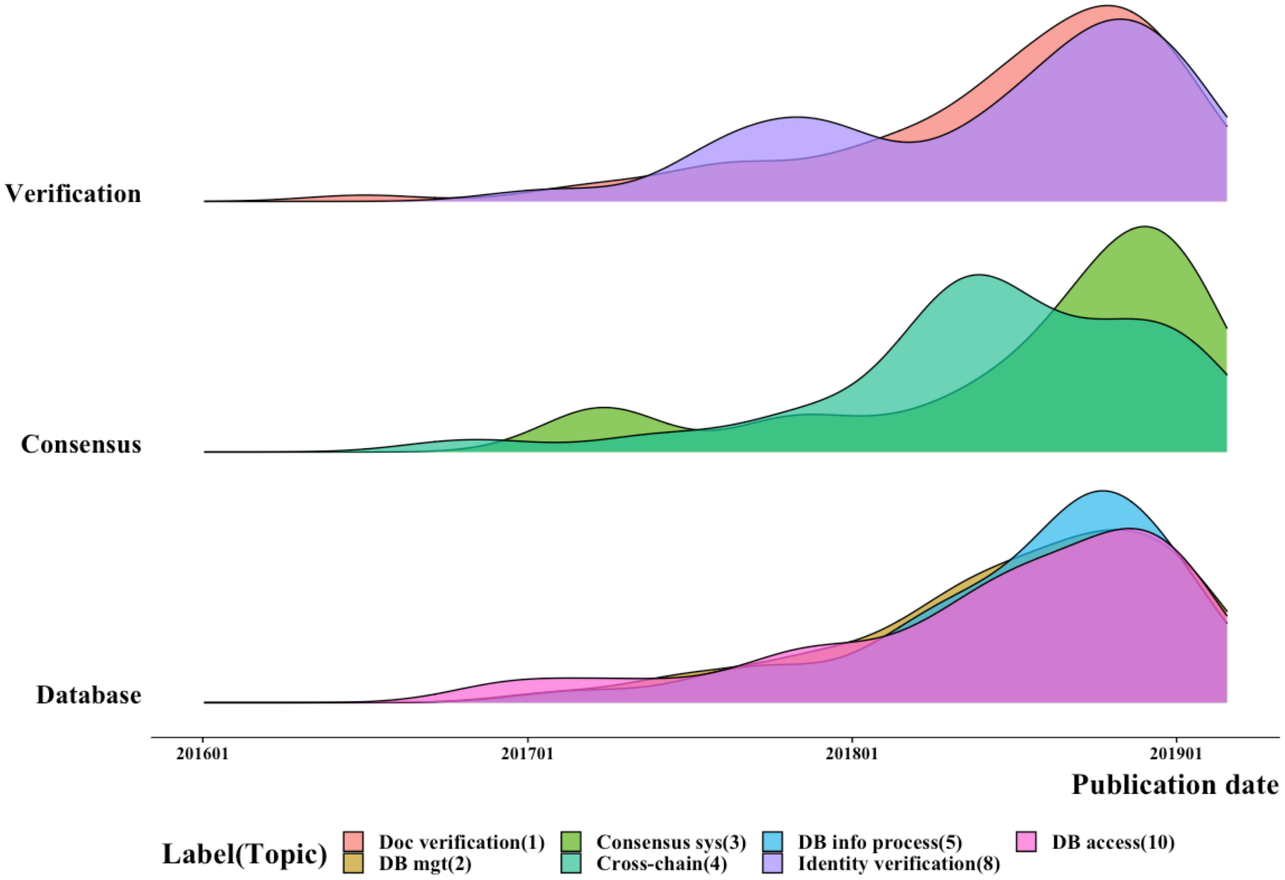
Results - Technology evolution map

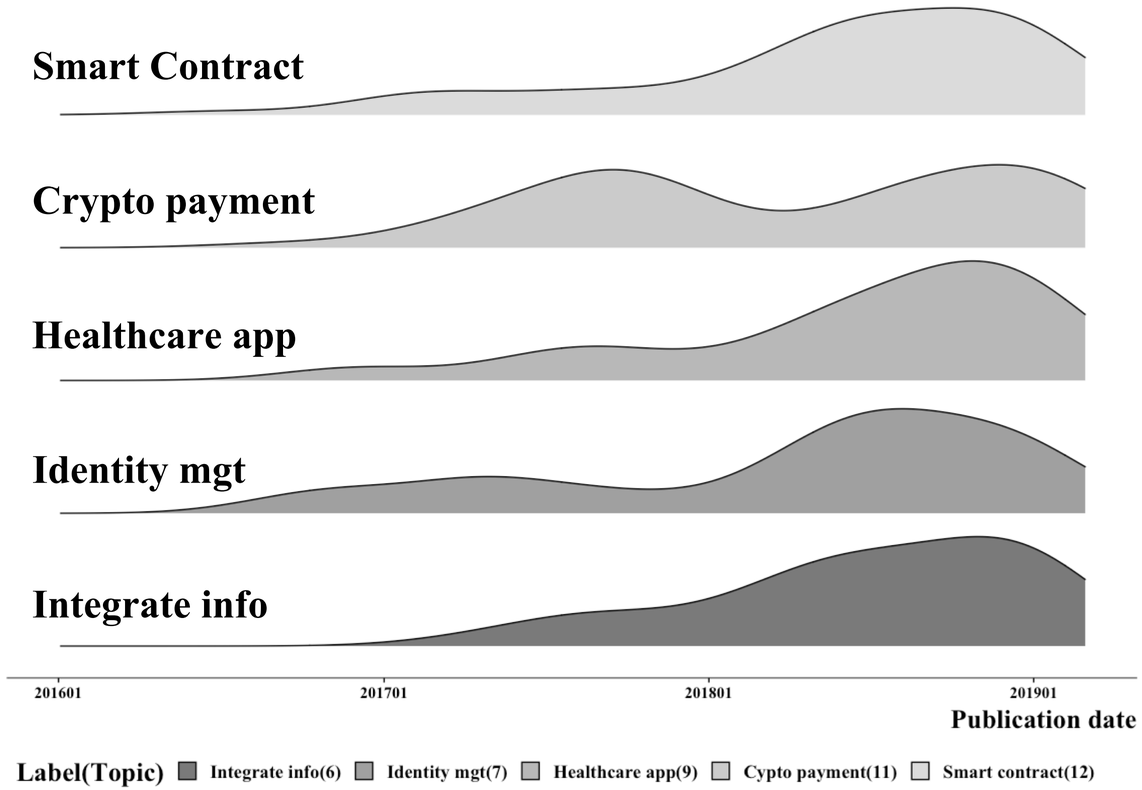
- Using topic-term matrix generated from LDA, the distance between topics can be computed by cosine similarity, and hierarchical clustering is also applied to identify more stable and distinct topics through bootstrap sampling. And the clustered group can further be labeled according to co-phrase appearing in both topics.



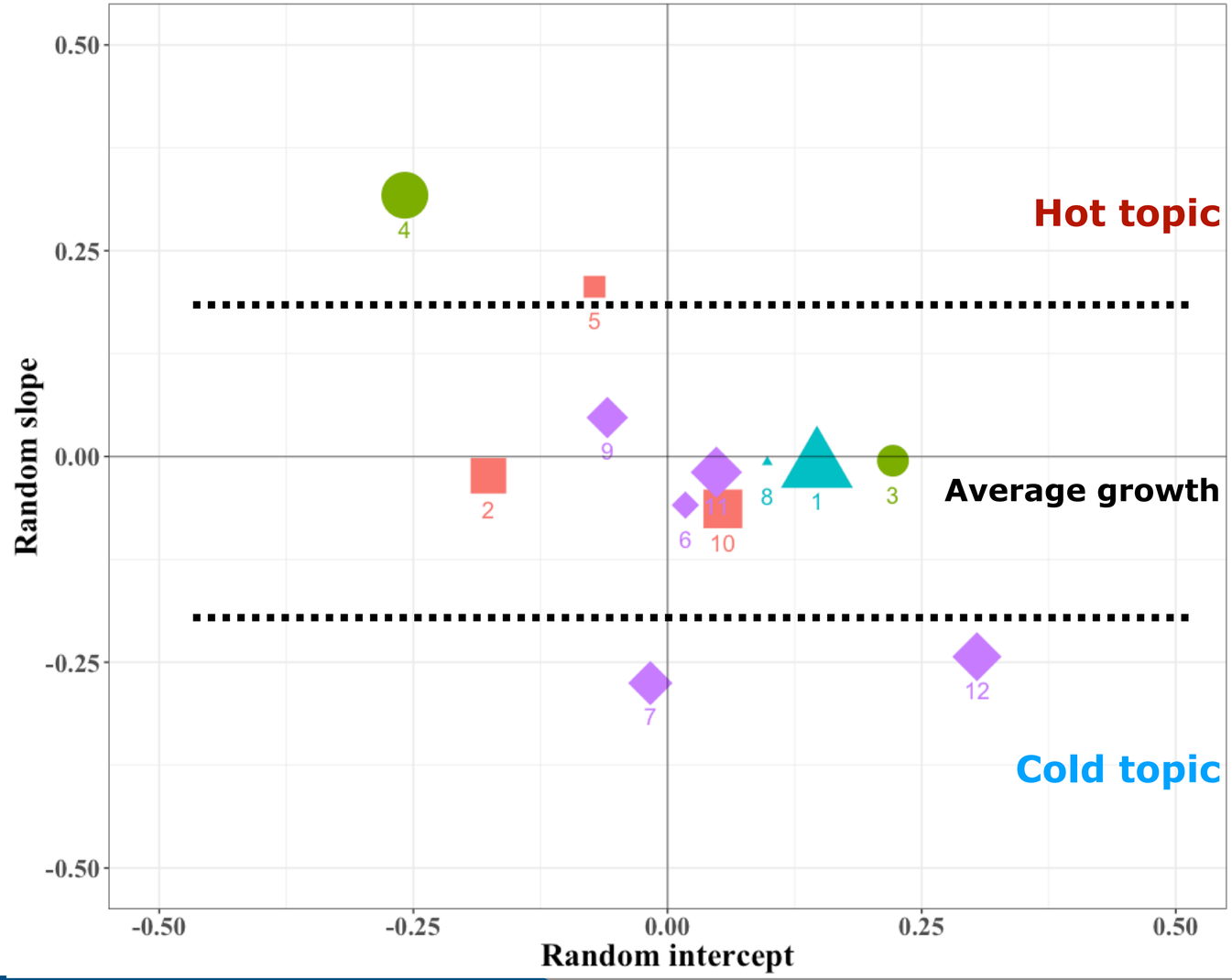
$$cosinesimilarity(Topic_1, Topic_2) = \frac{Topic_1 \times Topic_2}{|Topic_1| \times |Topic_2|}$$

Results - Technology evolution map





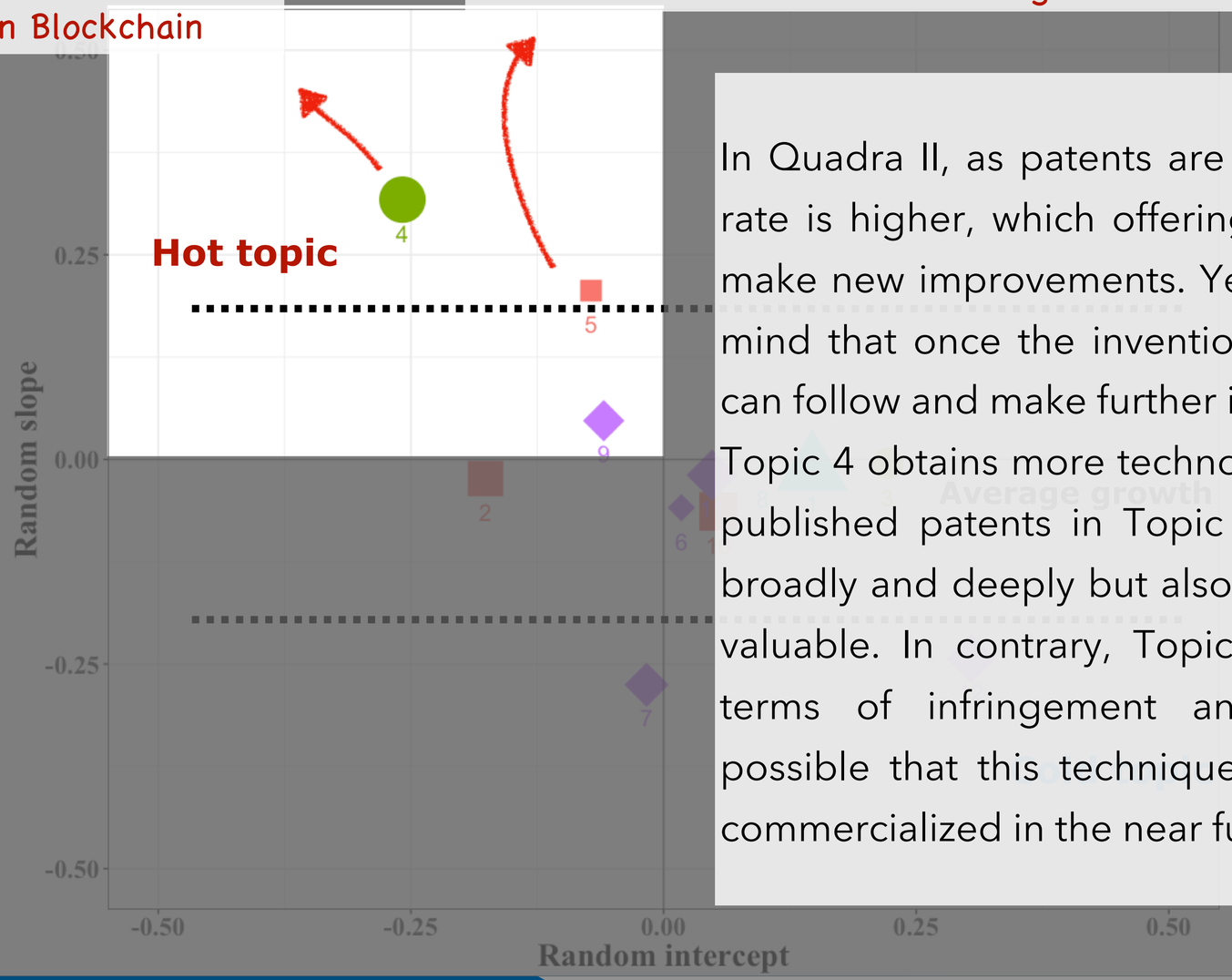
Conclusion


Hot topic
Average growth
Cold topic
Random slope
Random intercept
0.50
0.25
0.00
-0.25
-0.50
-0.50
-0.25
0.00
0.25
0.50
4
5
9
2
11
8
1
3
6
10
7
12

Conclusion

Cross-chain transaction related technologies on Blockchain

Database information processing based on Blockchain related technologies

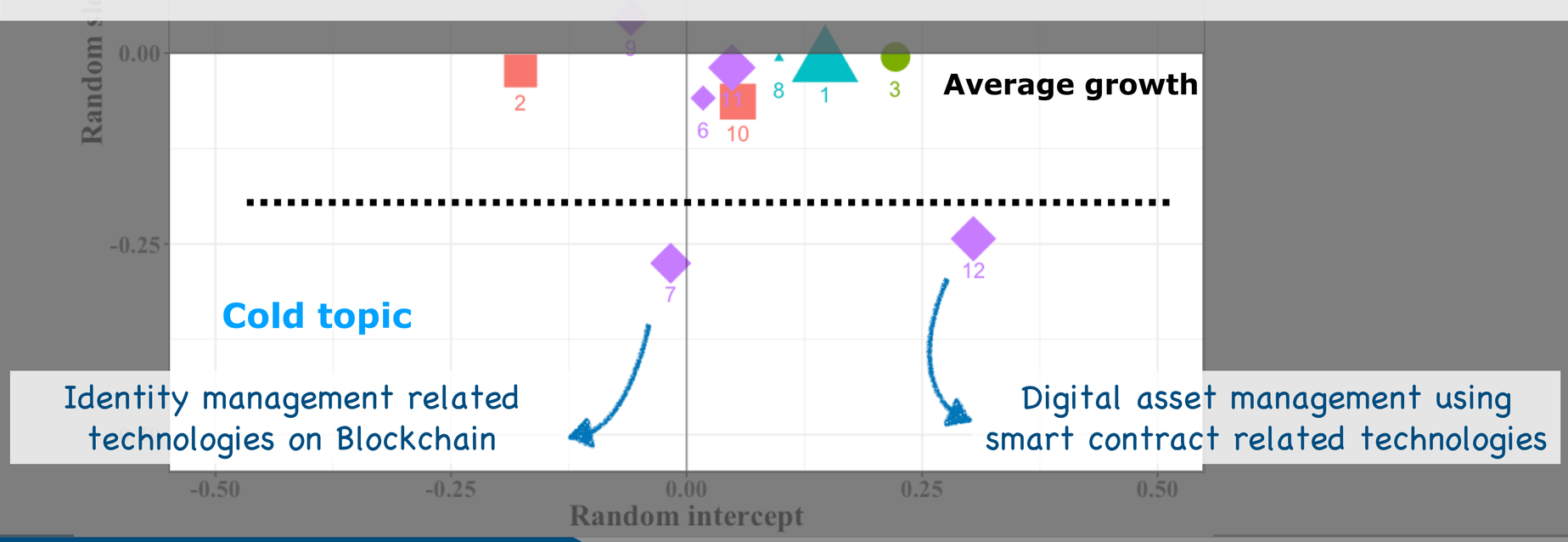


In Quadra II, as patents are less crowed and growth rate is higher, which offering a good opportunity to make new improvements. Yet it also need to keep in mind that once the invention is public, many others can follow and make further improve on it.

Topic 4 obtains more technology value which implies published patents in Topic 4 were protected more broadly and deeply but also reflect this techniques is valuable. In contrary, Topic 5 may be less risky in terms of infringement and litigation, yet more possible that this techniques cannot be realized or commercialized in the near future.

Conclusion

Two cold topics receiving lower growth rate are Topic 7 and Topic12 which lie in Quadrant III and IV respectively. Especially Quadrant IV is crowed with patents, and the change rate is slow, which imply that it might well be a highly competitive technology field and unlikely to be a source of competitive advantage. Besides, Topic 12 also hold higher technology value, therefore companies who sought to launch related technology can be achieved via licensing to mitigate R&D expense. Topic 7 is a total different case owing to lower number of published patents. For topics in Quadrant III, companies should understand the reasons of sparse patenting, and the technologies in this field might be old-fashioned or the potential of this technology is still blurred.



Conclusion

- These two hot topics shows the necessity of the interoperability and compatibility across different blockchain. Especially through increasing usage of established blockchain network like Bitcoin, Ethereum, Ripple and so on and most blockchains operate on isolated ecosystems.
- The results also indicate the smart contract (also called Blockchain 2.0 technique) is relative mature, so companies who sought to develop this technology can through licensing or cooperate with related companies which can be discover via topic-applicant network.

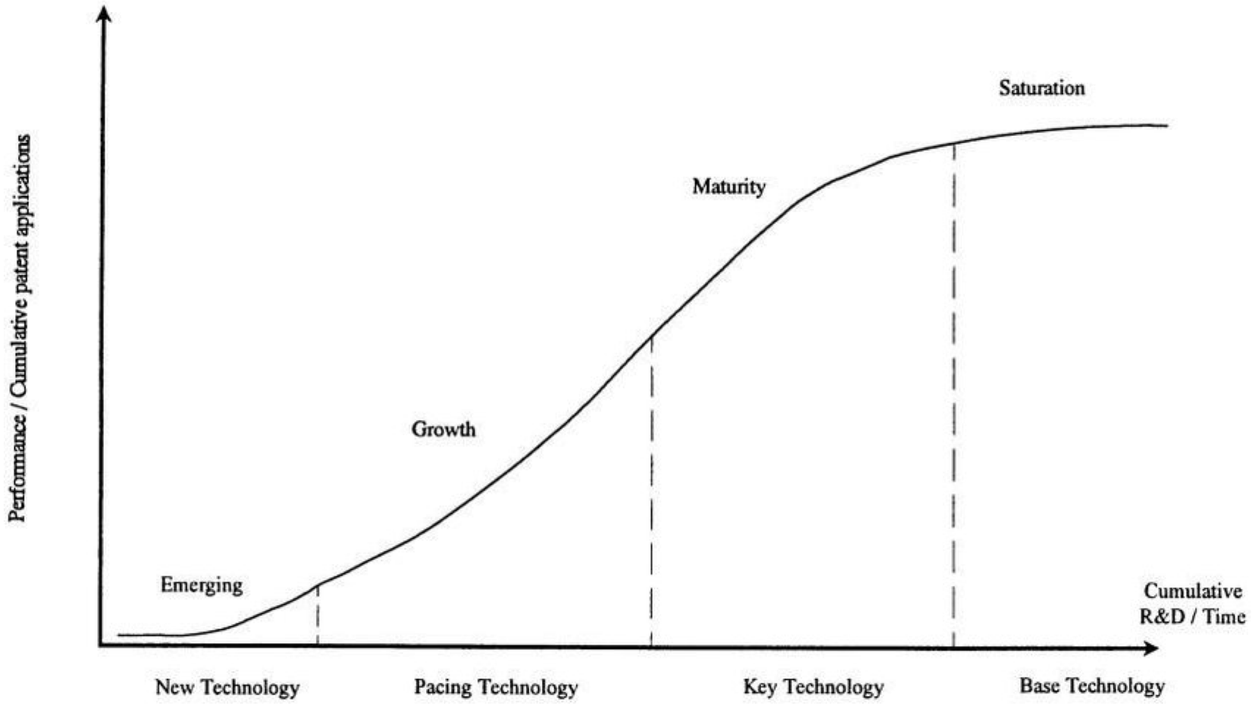
Limitation & future work

- Instead of using applicant in network analysis, it'd be better to adapt assignees who really have the transferred patent.
- The hyperparameters tuning section still need more discussion, especially when LDA applied to more specific field with less data.
- Last, this research applied patent mining method to identify emerging trends, but in Blockchain technology field the cryptocurrencies and related services may not be presented in patent documents and have immense influence.

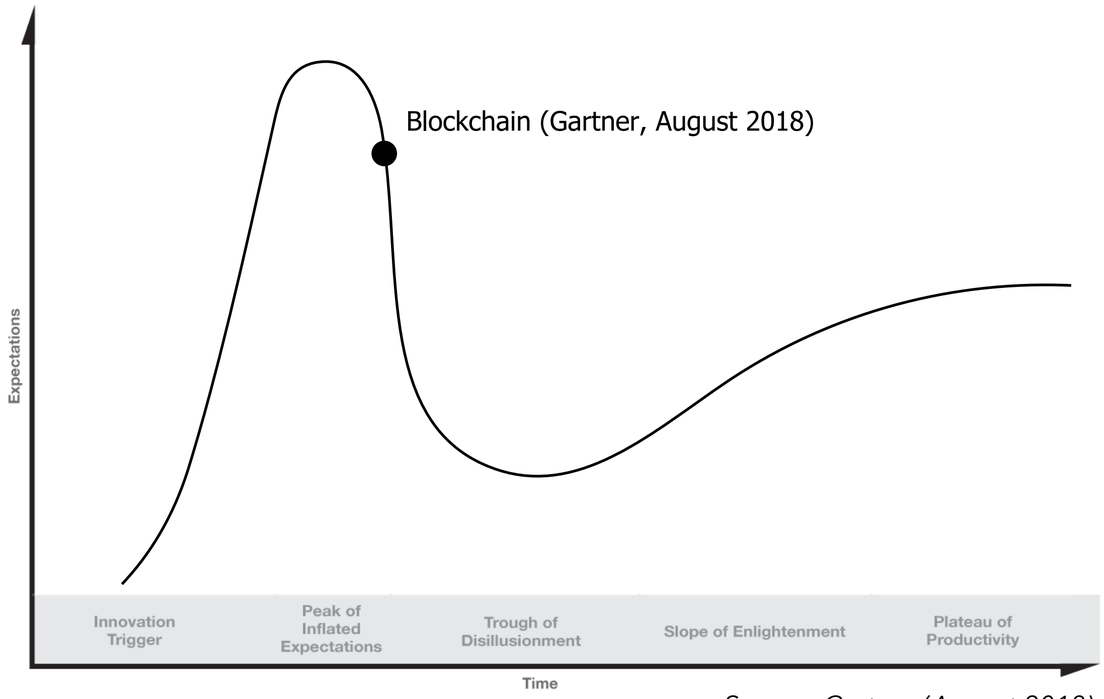
Thanks for your attention 😄

Cycle mapping

S-curve technology life cycle

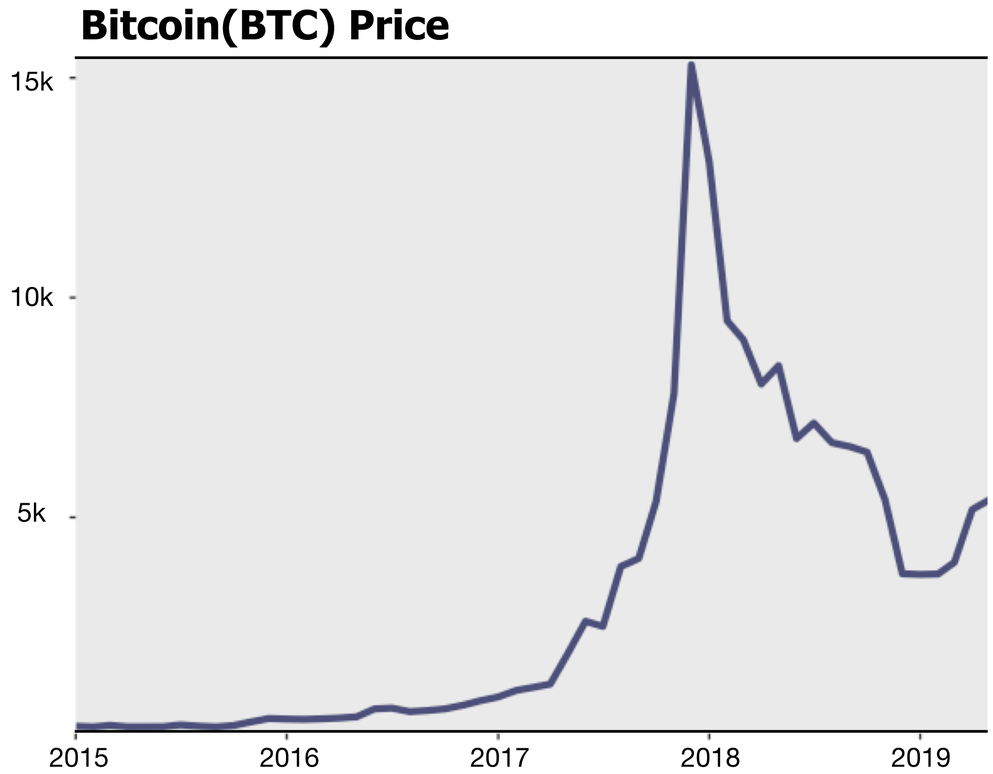


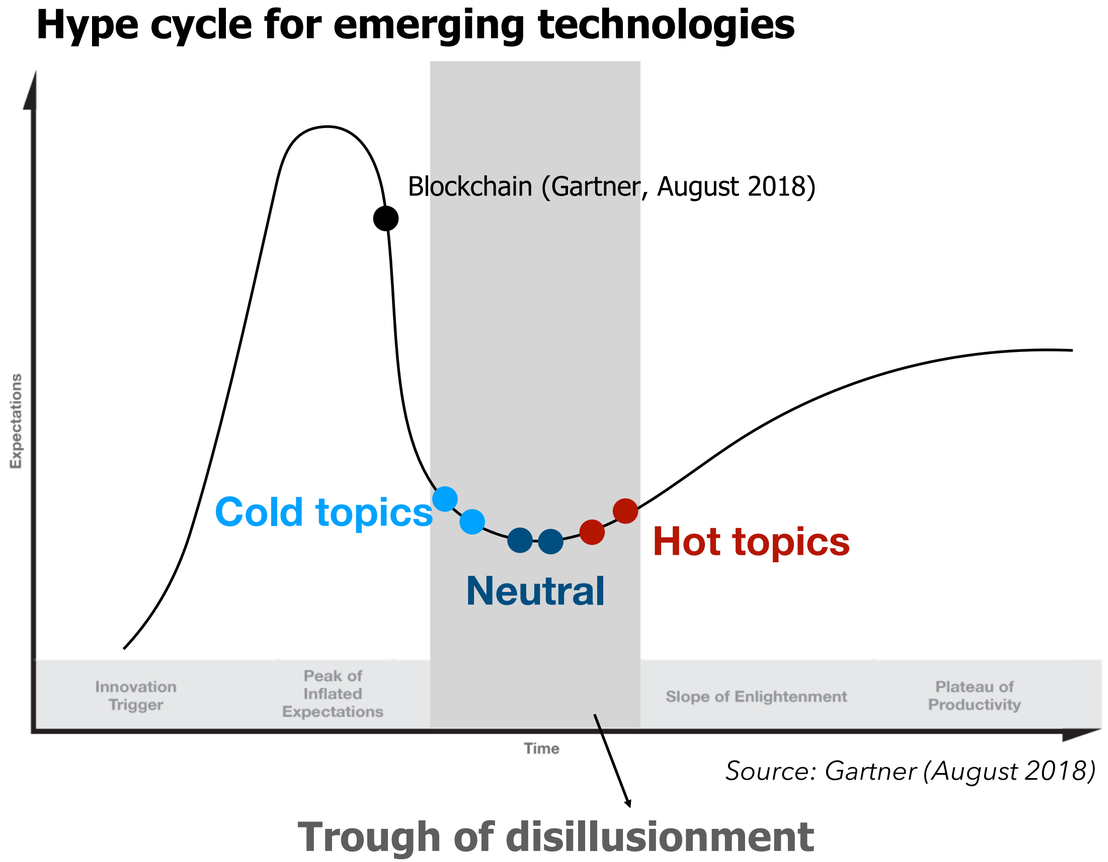
Hype cycle for emerging technologies



Source: Gartner (August 2018)

Hype cycle mapping





Source: Gartner (August 2018)

Trough of disillusionment