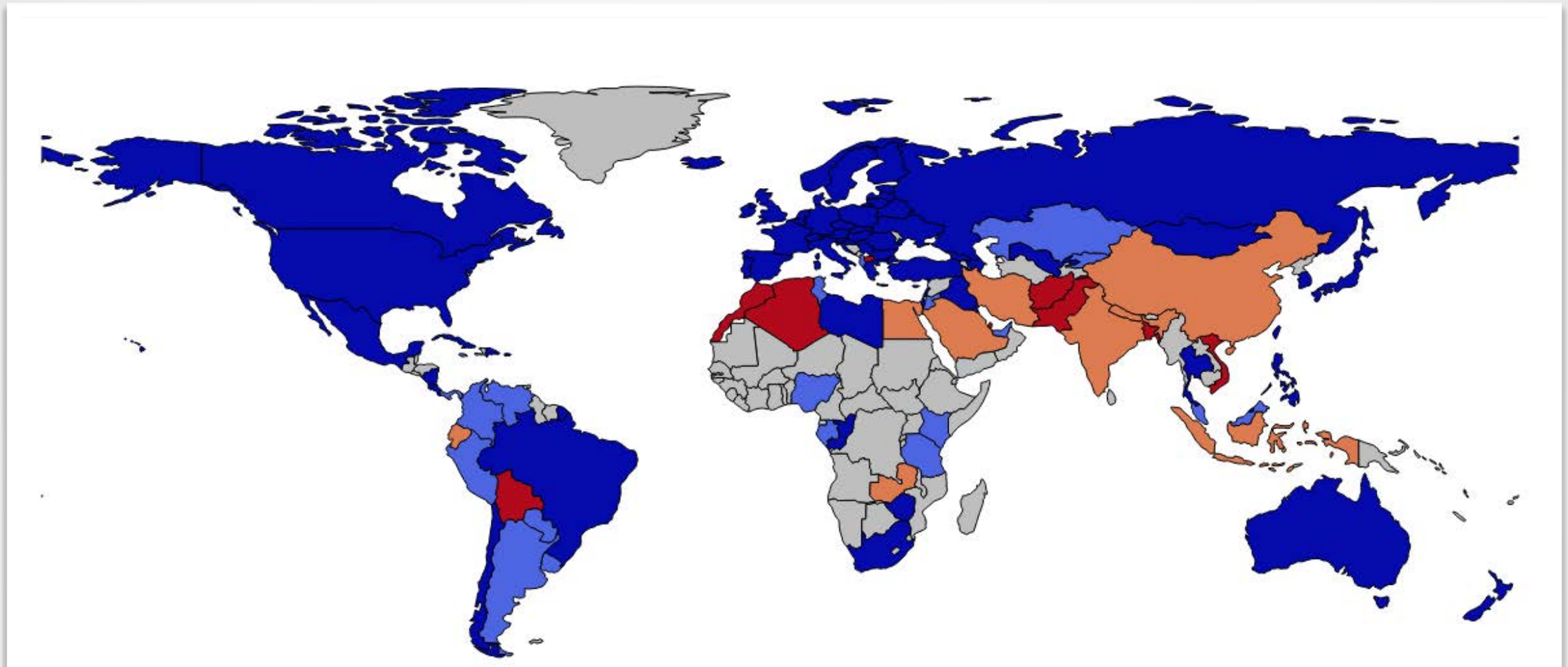
CRRIX: A Machine Learning Based Regulatory Risk Index for Cryptocurrencies

Xinwen Ni
Wolfgang Karl Härdle
Taojun Xie

Ladislaus von Bortkiewicz Professor of Statistics
Humboldt-Universität zu Berlin
BRC Blockchain Research Center
lvb.wiwi.hu-berlin.de
Charles University, WISE XMU, NCTU 玉山学者

20210609 Ni Hae CRRIX (VPC).key

Bitcoin Legality by Country

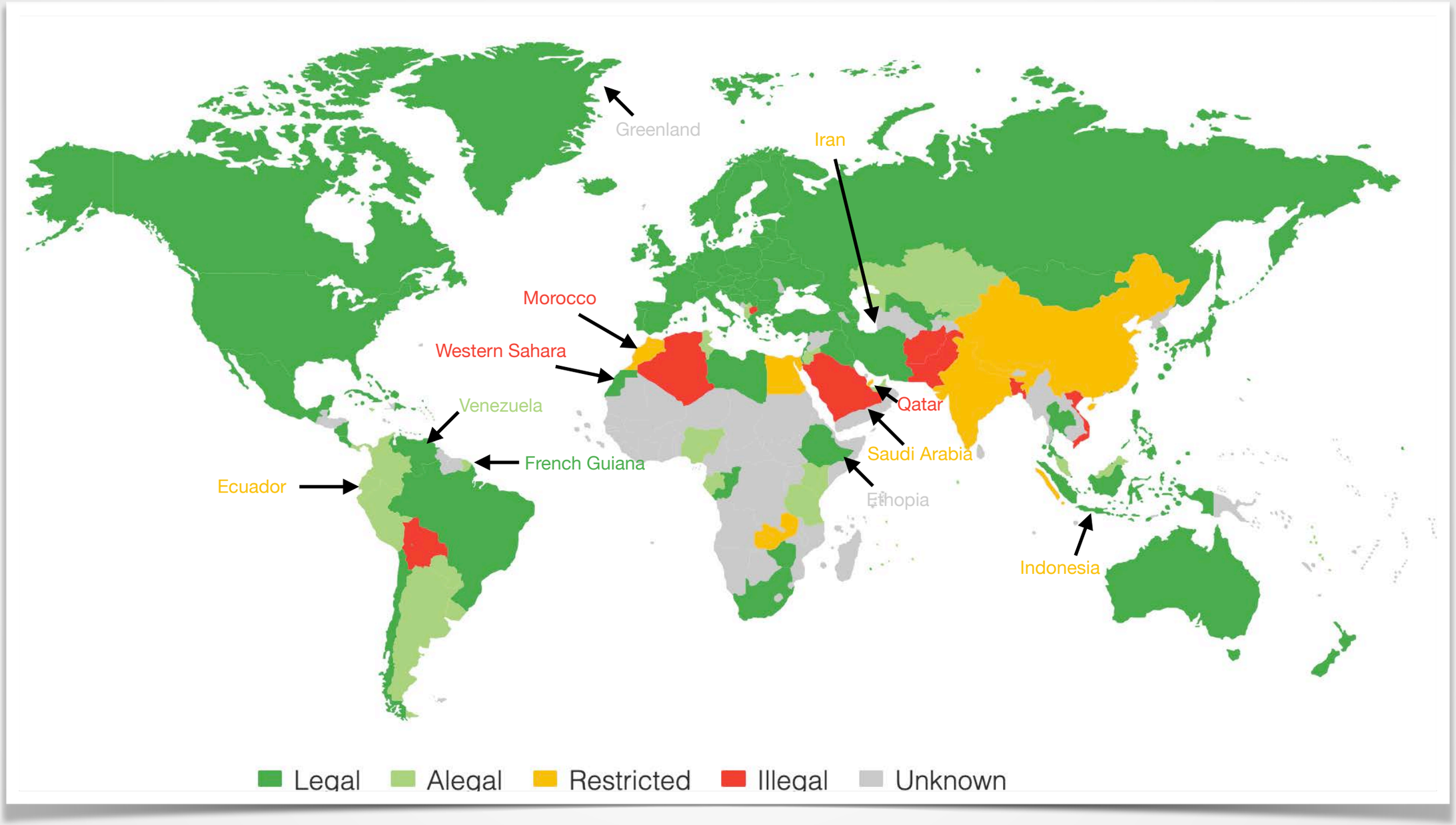


Legal, Alegal, Unknow, Restricted, Illegal, on 9th Jul 2018

Source: https://coin.dance/poli

?

Changing of Bitcoin Legality



On 24th Oct 2020

Source: https://coin.dance/poli

?

Cryptocurrency Regulations and Risk

16 April 2021 FRIDAY

16 April 2021 FRIDAY **Official newspaper** Number: 31456

REGULATION

The Central Bank of the Republic of Turkey:

NOT USING CRYPTO ASSETS IN PAYMENTS
CIRCLE REGULATION

Purpose and scope

ARTICLE 1 - (1) The purpose of this Regulation is not to use crypto assets in payments, not to use crypto assets directly or indirectly in the provision of payment services and electronic money issuance, and payment and electronic money institutions to platforms that offer trading, custody, transfer or issuance services for crypto assets or It is the determination of the procedures and principles regarding not mediating the fund transfers from these platforms.

Rest

ARTICLE 2 - (1) These Regulations, 01/14/1970 dated and No.1211 Central Bank of the Republic of Turkey of the Law Article 4 of the third paragraph (I) of subparagraph (f) The fourth paragraph of the subparagraph and 06.20.2013 dated and 6493 No. Numbered, on Payment and Securities Settlement Systems, Payment Services and Electronic Money Institutions, the third paragraph of Article 12 and the sixth paragraph of Article 18.

Not using crypto assets in payments

ARTICLE 3 - (1) In the implementation of this Regulation, a crypto asset is created virtually using distributed ledger technology or a similar technology and distributed over digital networks, but as a fiat money, deposit money, electronic money, payment instrument, security or other capital market instrument Refers to intangible assets that are not qualified.

(2) Crypto assets cannot be used directly or indirectly for payments.

(3) No service can be provided for direct or indirect use of crypto assets in payments.

Not using crypto assets in the provision of payment services and electronic money issuance

ARTICLE 4 - (1) Payment service providers cannot develop business models in a way that crypto assets are used directly or indirectly in the provision of payment services and electronic money issuance, and cannot provide any services related to such business models.

(2) Payment and electronic money institutions cannot mediate on platforms offering trading, custody, transfer or issuance services regarding crypto assets or fund transfers from these platforms.

Force

ARTICLE 5 - (1) This Regulation enters into force on 30/4/2021.

Executive

ARTICLE 6 - (1) These Regulations are enforced by the Chairman of the Central Bank of the Republic of Turkey.

TÜRKİYE CUMHURİYET MERKEZ BANKASI

Press Release on Payments Area

16 April 2021 **No: 2021-17**

Studies on the regulation regarding the disuse of crypto assets in payments have been completed.

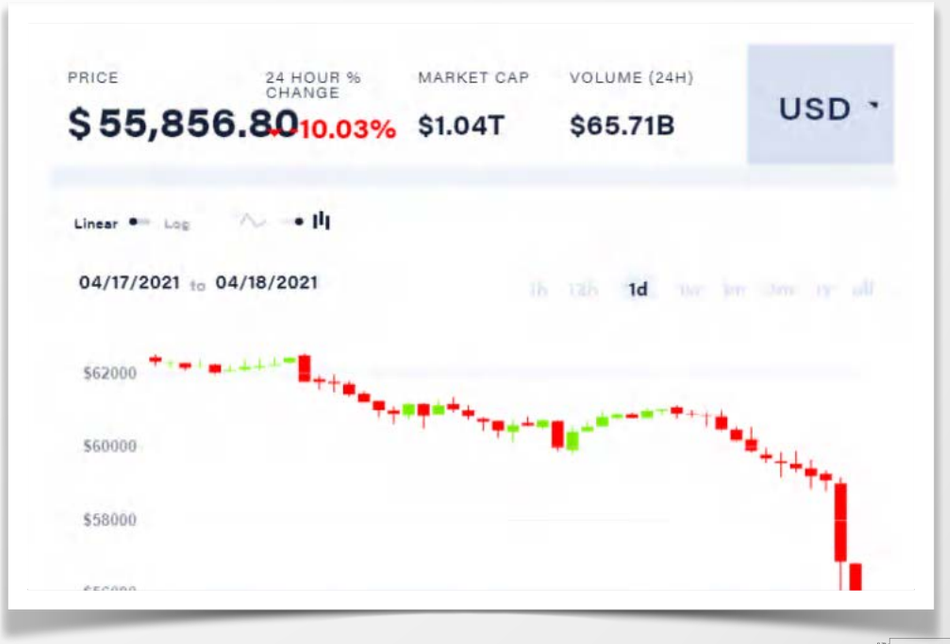
Crypto assets entail significant risks to the relevant parties due to the following reasons:

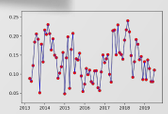
- they are neither subject to any regulation and supervision mechanisms nor a central regulatory authority,
- their market values can be excessively volatile,
- they may be used in illegal actions due to their anonymous structures,
- wallets can be stolen or used unlawfully without the authorization of their holders, and
- transactions are irrevocable.

Recently, some initiatives have emerged regarding the use of these assets in payments. It is considered that their use in payments may cause non-recoverable losses for the parties to the transactions due to the above-listed factors and they include elements that may undermine the confidence in methods and instruments used currently in payments.

Accordingly, pursuant to the authority vested by the Law No:1211 on the Central Bank of the Republic of Turkey (CBRT) and the Law No. 6493 on Payment and Securities Settlement Systems, Payment Services and Electronic Money Institutions, the CBRT has introduced "Regulation on the Disuse of Crypto Assets in Payments".

Contact
For further information, you may send an e-mail to basin@tcmb.gov.tr.





Cryptocurrency Regulations and Risk

Bitcoin plunges 14 per cent after Twitter rumor that US Treasury is planning crackdown on crypto money-laundering schemes

By REUTERS and ANNABEL GROSSMAN FOR DAILYMAIL.COM
PUBLISHED: 15:23 BST, 18 April 2021 | **UPDATED:** 14:37 BST, 19 April 2021

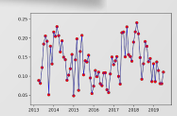
Share 52 shares 681 View comments

Bitcoin fell by as much as 14 per cent to $51,541 on Sunday amid unconfirmed Twitter speculation that the US Treasury may crack down on money laundering that's carried out through cryptocurrency.

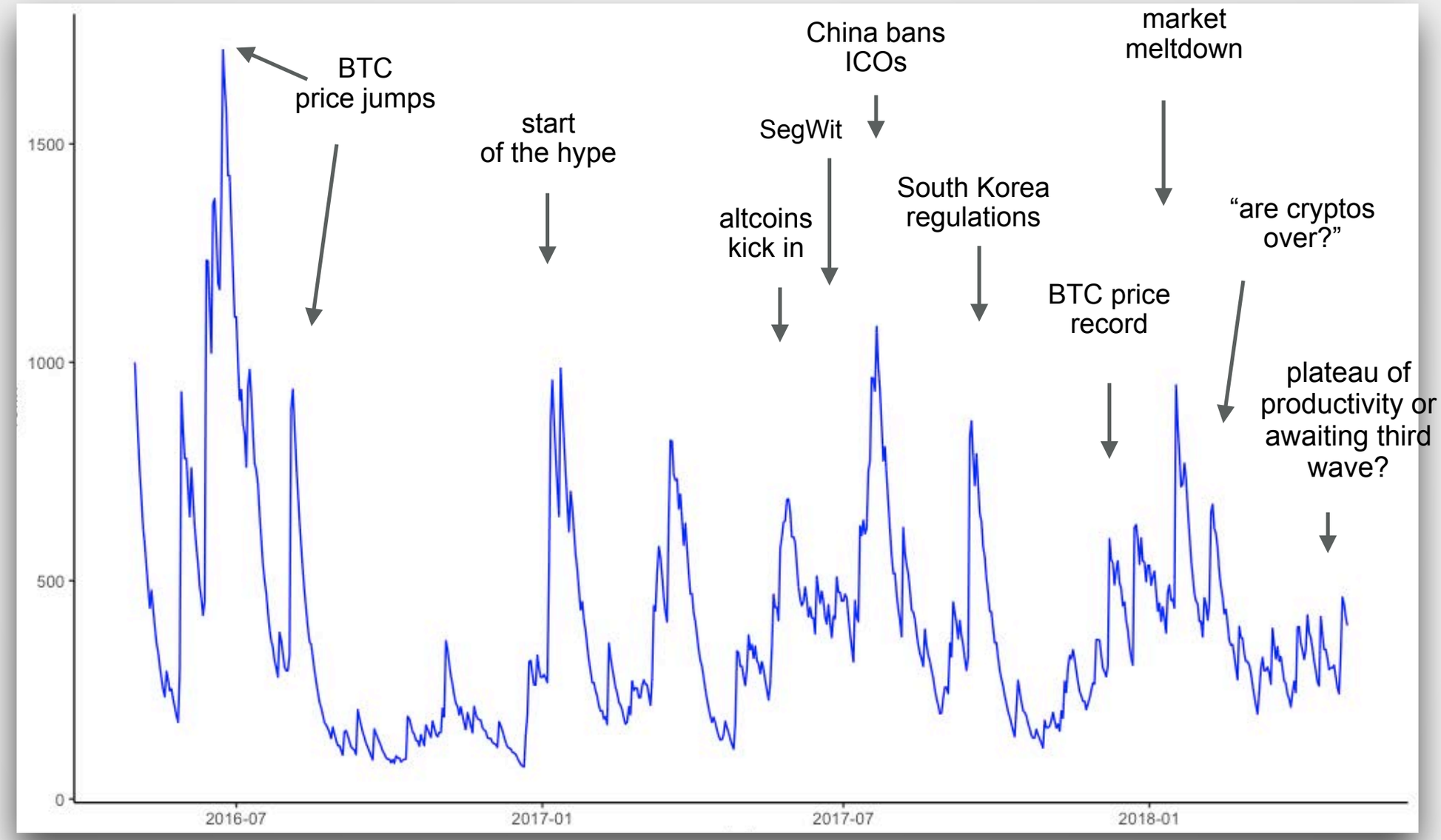
SHARE SELECTION

The selloff meant the cryptocurrency reversed most of the big gains it made over the past week.

The latest dip meant it was trading at $53,991, which is a whopping $12,000 below record highs of above $64,800 set on Wednesday.



VCRIX v.s. Events

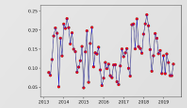


Source: Kolesnikova, 2018

VCRIXindex

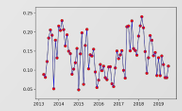
Regulatory Risk

- Risk of a change in laws and regulations
- Impact a security, business, sector, or market
- More than 1500 coins in cryptocurrency market
- High attention of the regulatory authorities: protect the investors, put a stop on money laundering, or prevent the fiat currency from being crowded out
- Result in volatile price movements: -16% on 14 Sep 2017, China's ban on cryptocurrency exchanges



Economic Index

- Monitor the market condition and provide information about future
- Methods of developing an index:
  - IFO Business Climate Index, Purchasing Managers' Index (PMI)
  - Real data: Fiscal expenditure, GDP, inflation
  - News data: News Analytics and MarketPsych Indices （sentiment）, Economic Policy Uncertainty Index

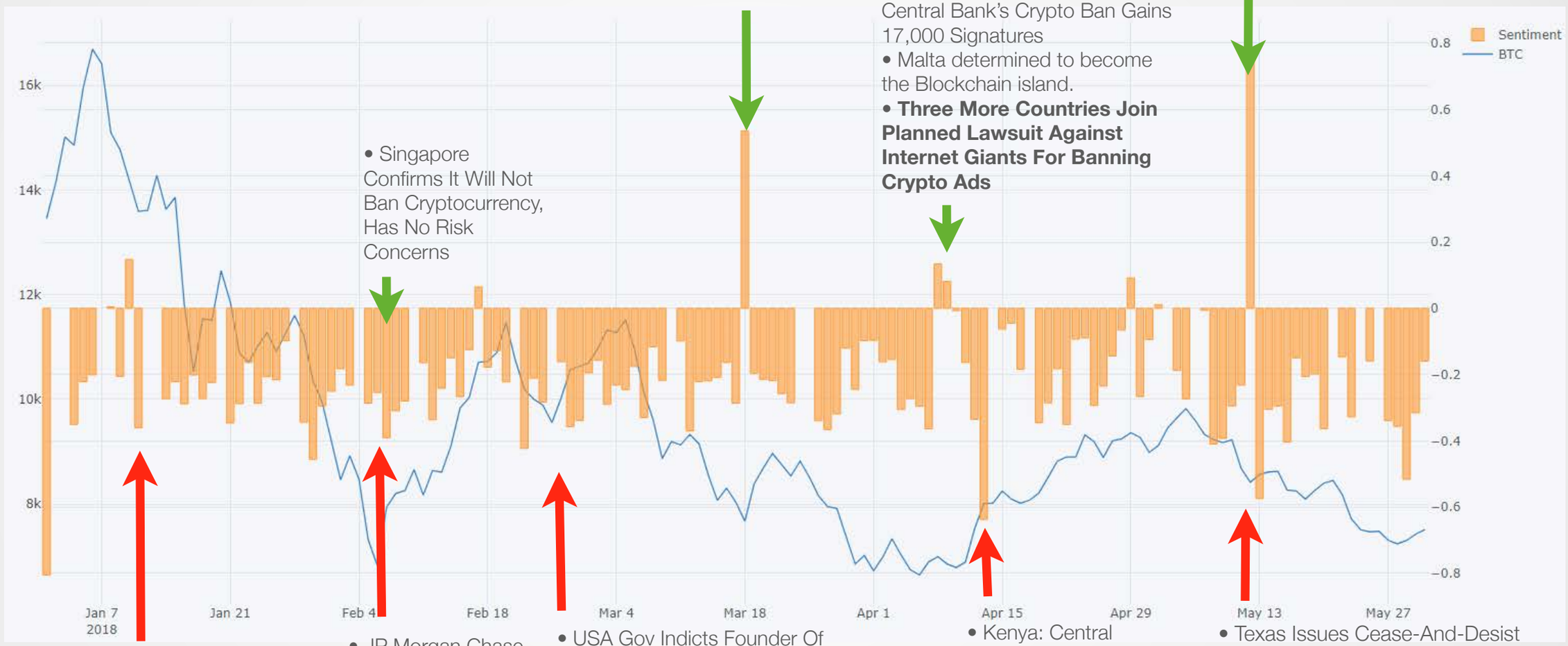


BTC v.s. Sentiment in 2018

• Israel: Steps towards cryptocurrency support.
• Blockchain takes hold in Israel: Expert take.

• India: Petition To Reverse Central Bank's Crypto Ban Gains 17,000 Signatures
• Malta determined to become the Blockchain island.
• **Three More Countries Join Planned Lawsuit Against Internet Giants For Banning Crypto Ads**

• French Crypto Regulation Ã la carte: Context, News, Perspectives



• Singapore Confirms It Will Not Ban Cryptocurrency, Has No Risk Concerns

• South Korea confirm strict regulation
• China move to close Exchange-like services
• Indonesia warn against Crypto

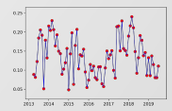
• JP Morgan Chase and Lloyds Bank ban credit card purchase of cryptos

• USA Gov Indicts Founder Of Long Defunct Crypto Exchange BitFunder
• South Africa: $50 mil scam
• USA: $25 mil seized

• Kenya: Central Bank Warns Banks Against Dealing In Crypto

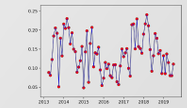
• Texas Issues Cease-And-Desist To BTC Investment Company-Promised 100% Profit In 21 Days
• India: Supreme Court Denies Interim Injunction Against Reserve Bank Restrictions

Source: Bingling Wang

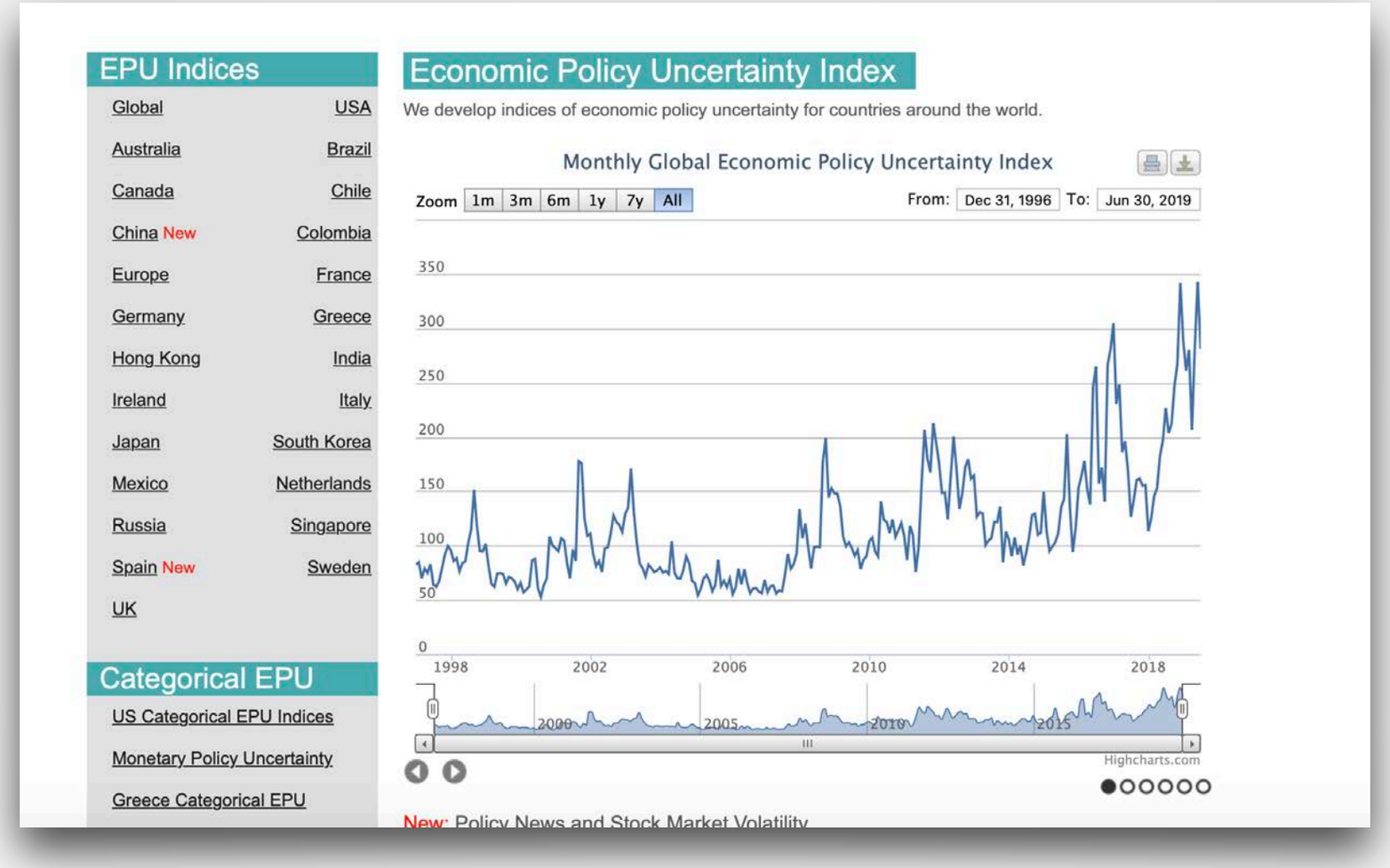


Economic Policy Uncertainty Index

- Baker, Bloom & Davis (2016, QJE)
- Three types of underlying components:
  - an index of search results from 10 large newspapers
  - reports by the Congressional Budget Office (CBO) that compile lists of temporary federal tax code provisions (expire over the next 10 years)
  - the dispersion between individual forecasters' predictions
- In the paper, policy-related economic uncertainty based on newspaper coverage frequency (only component 1)

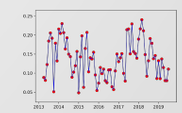


Regulatory Risk Index



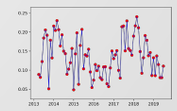
source: http://www.policyuncertainty.com

- frequency of articles contain “economic” or “economy”; “uncertain” or “uncertainty”; and one or more of “congress”, “deficit”, “Federal Reserve”, “legislation”, “regulation” or “White House”



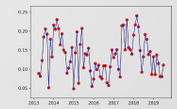
Research Questions

- How to identify the regulatory risk for cryptocurrencies?
- How to construct an index of regulatory risk for Cryptocurrencies based on news data?
- What is the impact of regulatory risk to the market?



Outline

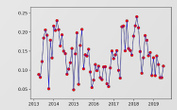
- ⊡ Motivation ✓
- ⊡ Data
- ⊡ Methodology
- ⊡ Results and CRRIX
- ⊡ Future plan and conclusion



News Data:

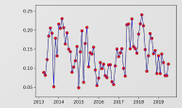
- Top Cryptocurrency or Blockchain news platforms

Websites	Data available from	Data Amount (accurate)	Updated to	Columns
CoinDesk	01 Apr 2013	24.113	15 Mar 2021	**Titles,authors,introduces,contents,releases,updates,imgs,tags,bars,links**
Bitcoin Magazine	28 Feb 2012	5.607	19 Feb 2021	**Titles,authors,introduces,contents,date,tags,types,links**
CoinTelegraph	05 Aug 2013	23.350	25 Feb 2021	**Author, title, content, url, views, shares, tags, dates**
News BTC	23 Oct 2013	18.040	15 Mar 2021	**titles, authors, dates, contents, category, tags, article_urls**
CryptoSlate	16 Sep 2017	5.184	19 Feb 2021	**titles, authors, dates, contents, category, article_urls**
Bitcoinist	22 Dec 2013	15.066	15 Mar 2021	**titles, authors, dates, contents, article_urls, category, tags**
Bitcoin.com	18 Jan 2015	13.497	15 Mar 2021	**titles, authors, dates, categories, tags, article_urls, contents**
Invest In Blockchain	07 Jul 2017	2.999	21 Dec 2019	**titles, authors, contents, dates, tags, article_urls**
Blockonomi	06 Dec 2017	3.173	9 Mar 2021	**titles, authors, dates, categories, abstracts, contents1, reads, article_urls**
Coinspeaker	16 Mar 2014	14.150	12 Mar 2021	**titles, dates, authors, contents, article_urls, tags**
Inside Bitcoins	18 Aug 2015	6.653	15 Mar 2021	**titles, update dates, authors, contents, article_urls,**
	Sum:	131.832		



Data: CRIX and VCRIX

- Represent the value (CRIX) and the volatility (VCRIX) of the entire cryptocurrency market
- CRIX (CRyptocurrency IndeX)
  - Trimborn et al (2018)
  - Track the entire cryptocurrency market performance (like S&P 500)
  - Data series starting from July 2014
- VCRIX (Volatility CRyptocurrency IndeX)
  - like VIX or VDAX
  - Kim et al (2020)
  - Available: thecrix.de



Article Timeline (Weekly Average per Day)

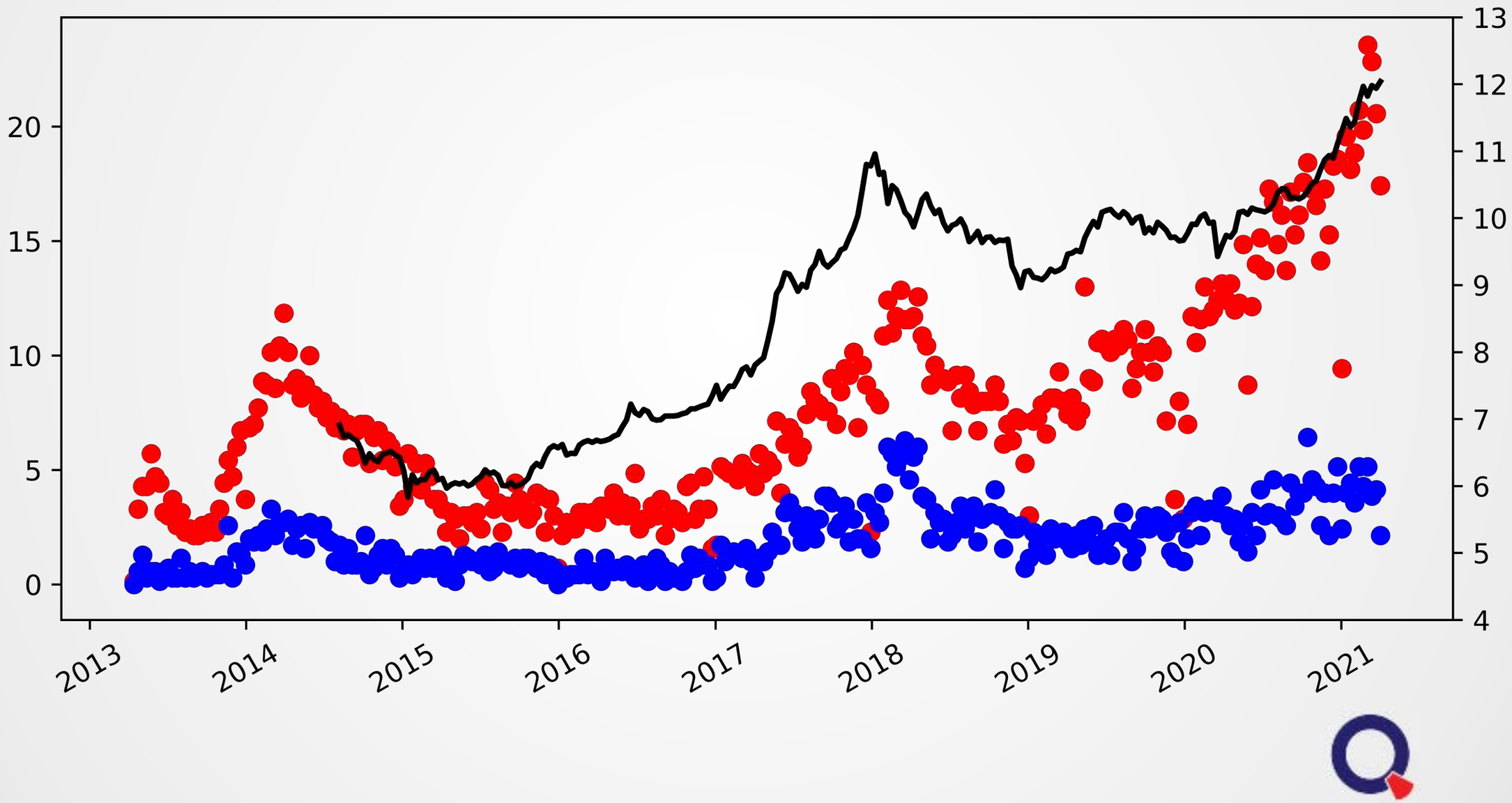


Figure: log CRIX and Number of Regulation News and News (CoinDesk)

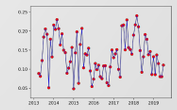
Cryptocurrency Regulatory Risk IndeX:

- CRRIX: frequency of policy-related news

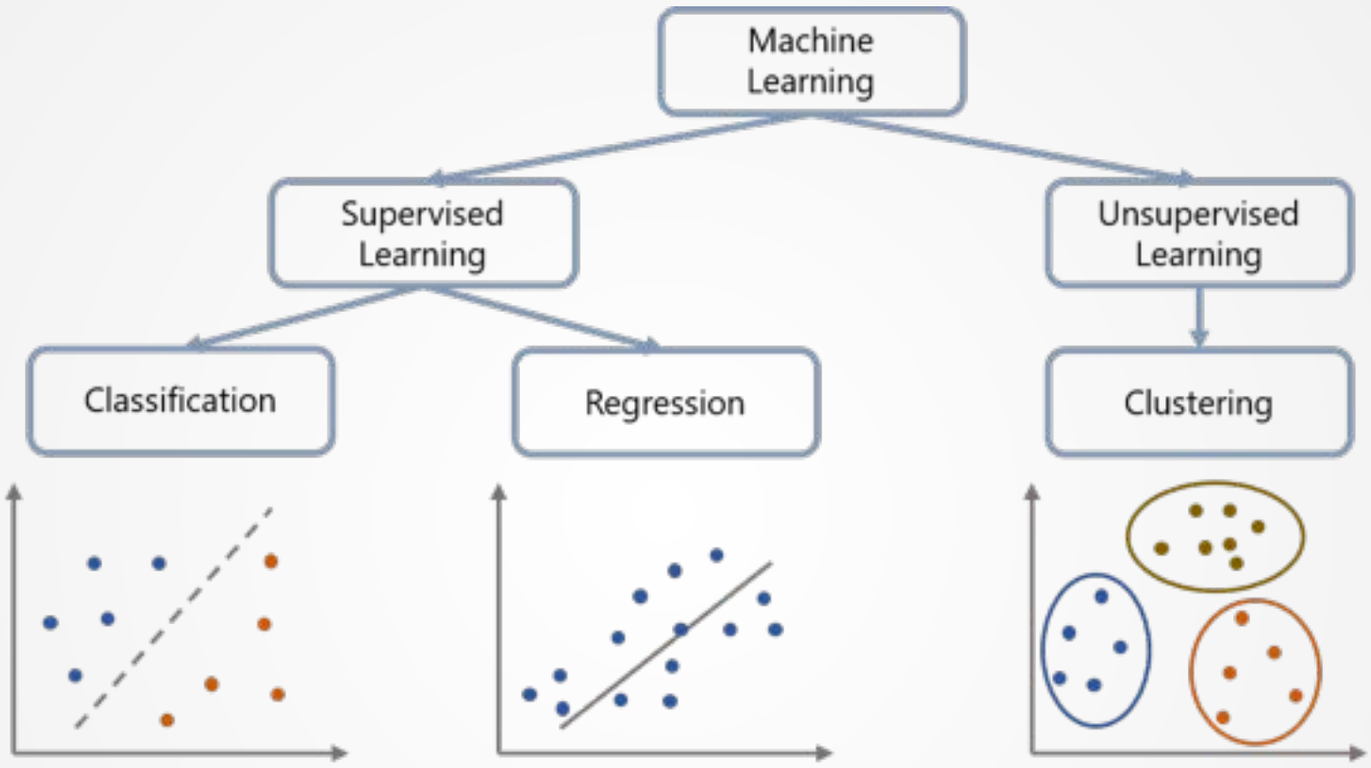
$$CRRIX_t^s = \frac{N_{t,reg}^s}{N_{t,all}^s}$$

where *s={daily, weekly, monthly}*, $N_{t,reg}$ and $N_{t,all}$ are the number of regulatory news and all news at time *t*

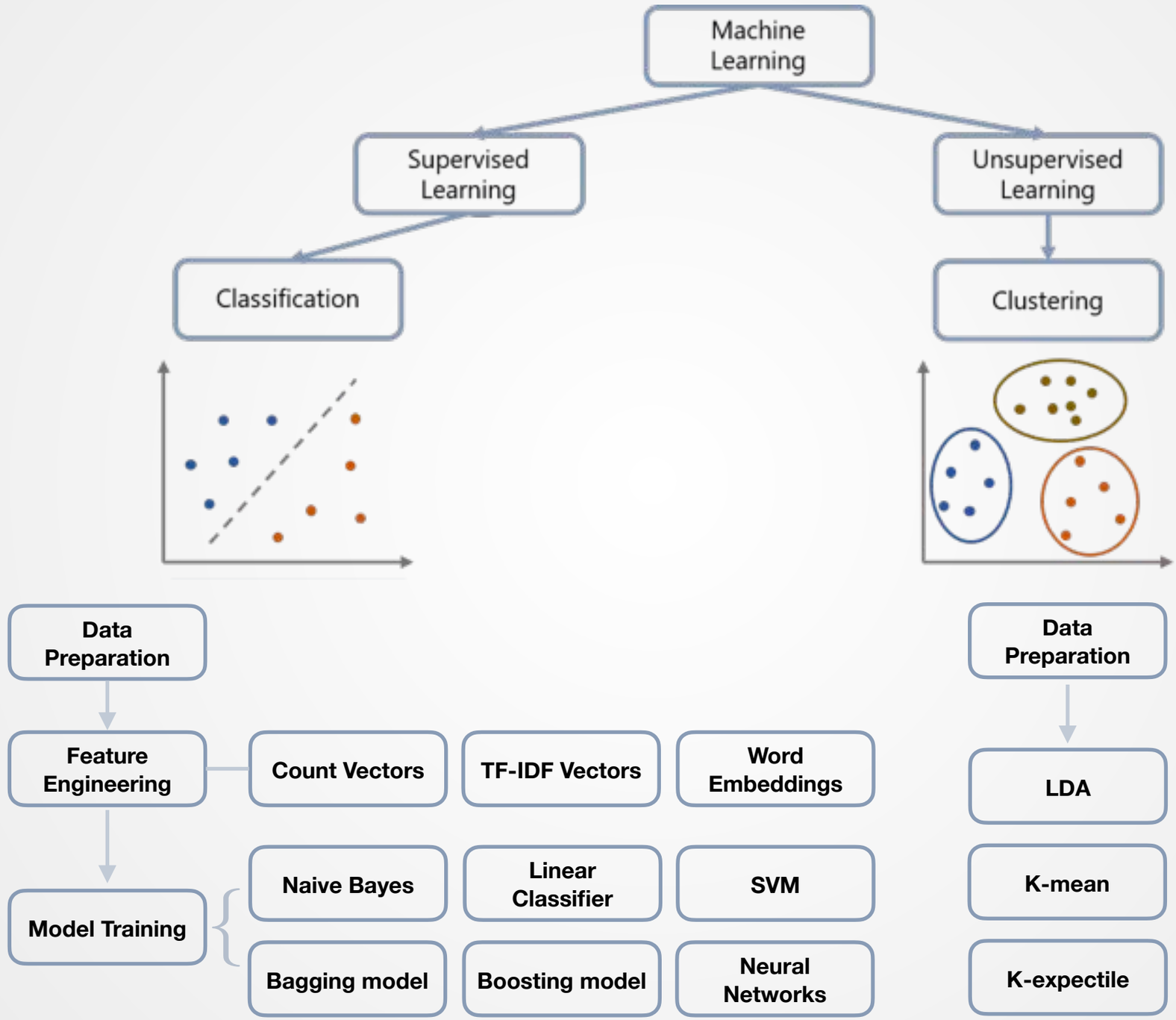
- Unknown $N_{t,reg}$ but big $N_{t,all}$
- Imbalance data: $N_{t,reg} << N_{t,all}$



Text Classification and Machine Learning

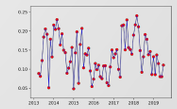


Text Classification and Machine Learning



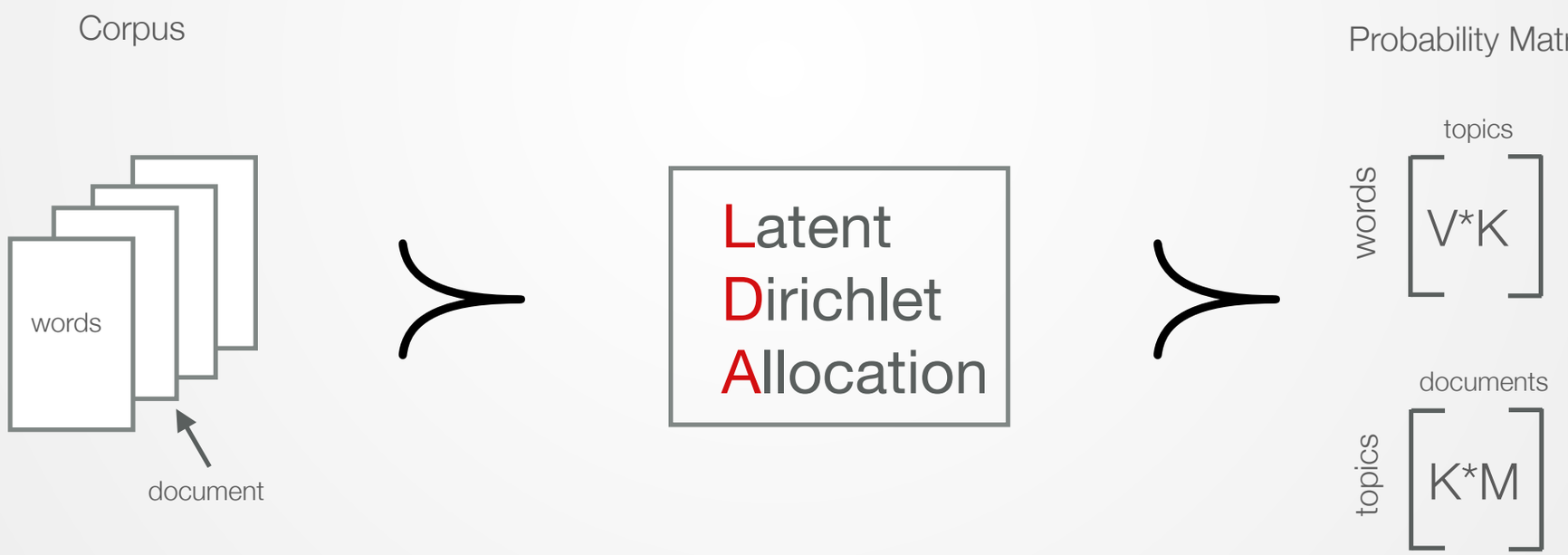
Classification for Cryptocurrency Corpus

- Supervised Method is possible:
  - Pre-classified data: CoinDesk data, Cointelegraph…
  - Unclassified data: Bitcoin Magazine,BTC News…
- Bad performance of Naive Bayes and SVM
  - Imbalanced data problem
  - Small training data
- Consider unsupervised method LDA
  - Capture policy-related topics
  - Calculate the similarity between unclassified documents and pre-classified regulatory news

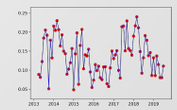


LDA: Basic Idea

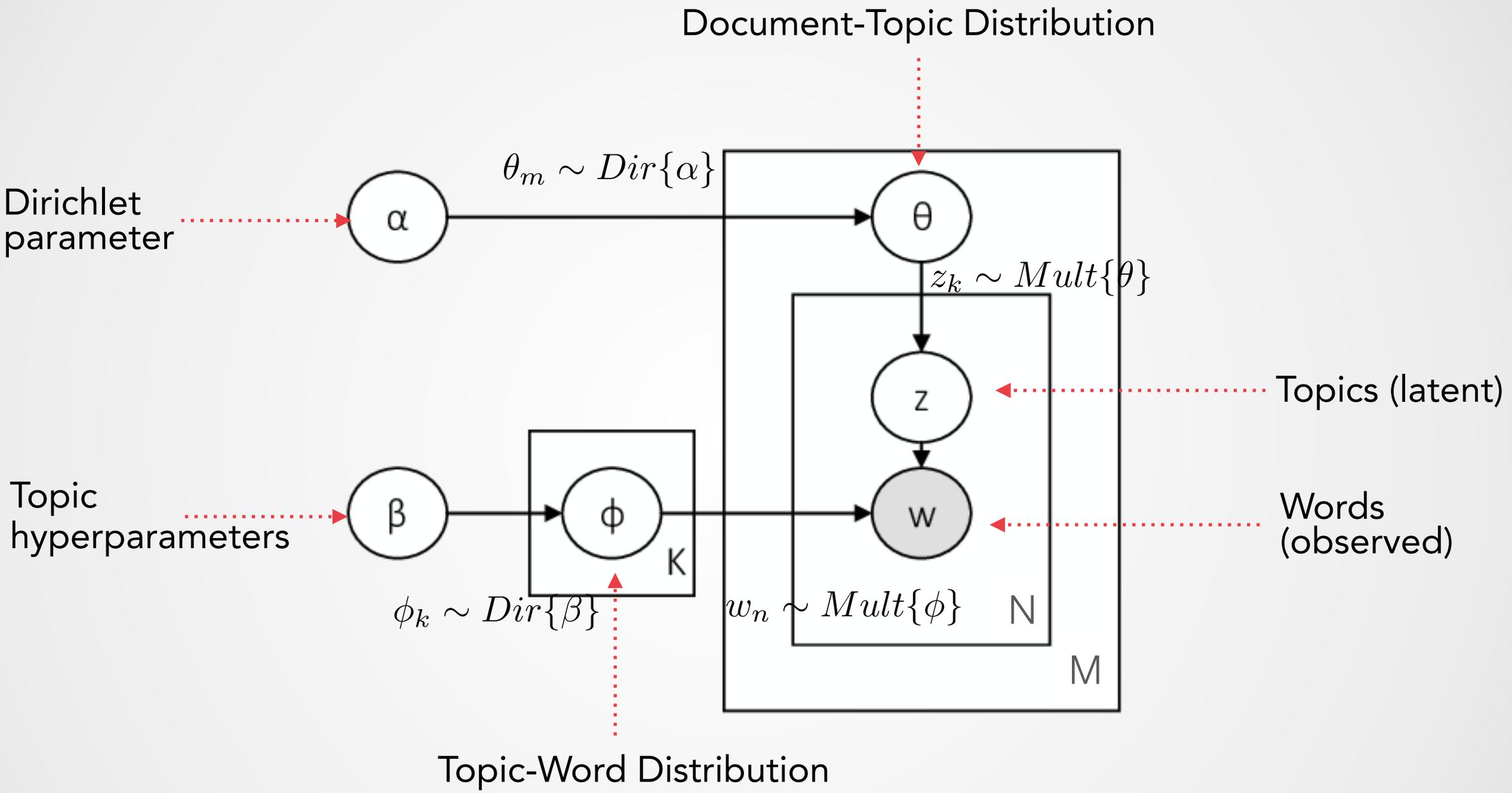
- Basic Idea:
  - Have a corpus of documents
  - LDA learns the topic representation of K topics and the word distribution of each topic
  - LDA identifies topics that are likely to generate the observed corpus



LDA

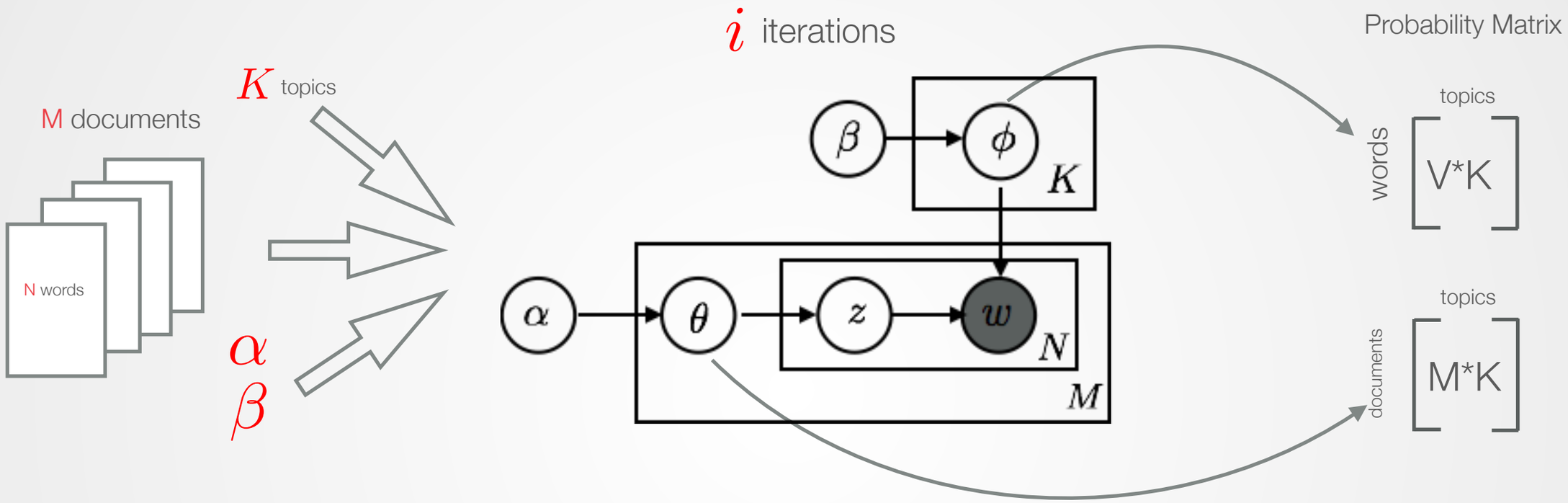


LDA Plate Notation



LDA

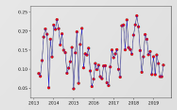
LDA: Plate Notation (Cont'd)



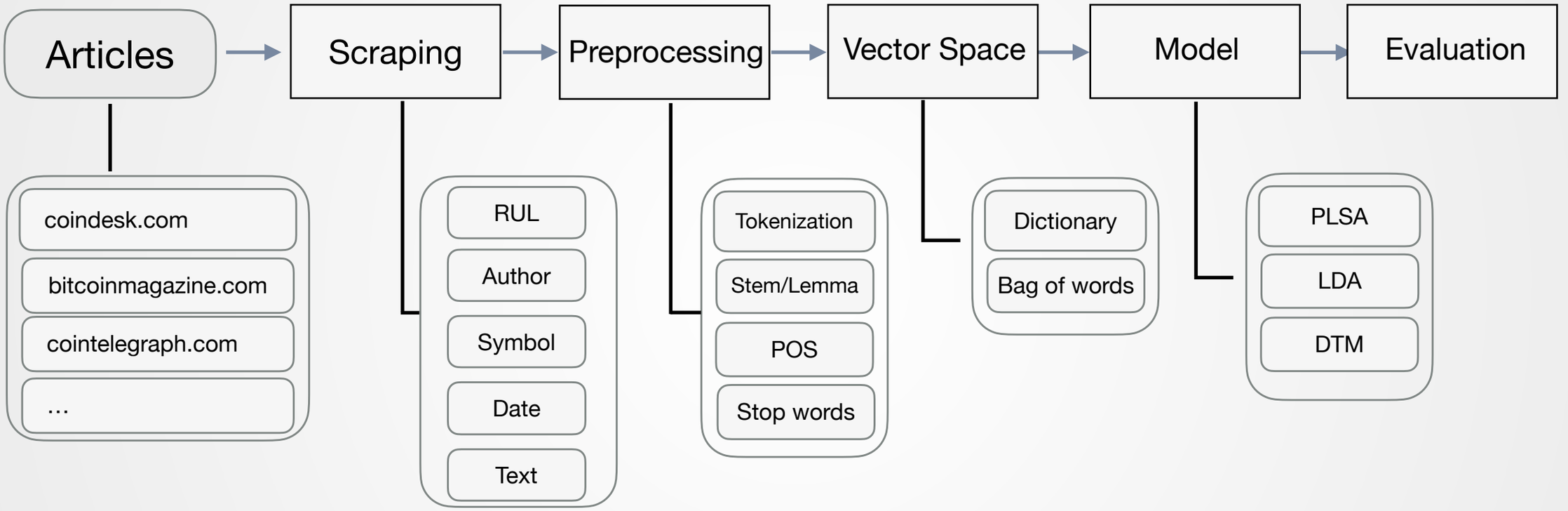
- The joint distribution of Words and topics:

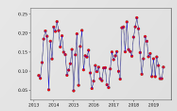
$$p(\boldsymbol{w}, \boldsymbol{z}, \boldsymbol{\theta}, \boldsymbol{\phi}; \alpha, \beta) = \prod_{k=1}^{K} p(\phi_k; \beta) \prod_{d=1}^{M} p(\theta_d; \alpha) \prod_{n=1}^{N} p(z_{d,k} \mid \theta_d) p(w_{d,n} \mid \phi_{z_{d,k}}),$$

LDA



Pipeline for Topic Modeling





Hellinger Distance

- P, Q two probability measures

$$H^2(P,Q) = \frac{1}{2}\int\left(\sqrt{\frac{dP}{d\lambda}} - \sqrt{\frac{dQ}{d\lambda}}\right)^2 d\lambda.$$

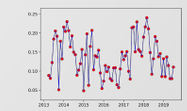
- denote the densities $dP/d\lambda$ and $dQ/d\lambda$ as f and g

$$H^2(f,g) = \frac{1}{2}\int\left(\sqrt{f(x)} - \sqrt{g(x)}\right)^2 dx = 1 - \int\sqrt{f(x)g(x)}\,dx,$$

- For discrete probability distributions: $P = (p_1,\ldots,p_k)$ and $Q = (q_1,\ldots,q_k)$

$$H(P,Q) = \frac{1}{\sqrt{2}}\sqrt{\sum_{i=1}^{k}(\sqrt{p_i} - \sqrt{q_i})^2},$$

Distances

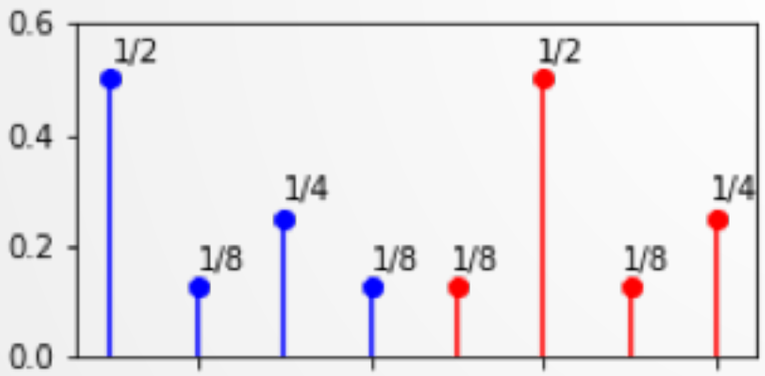


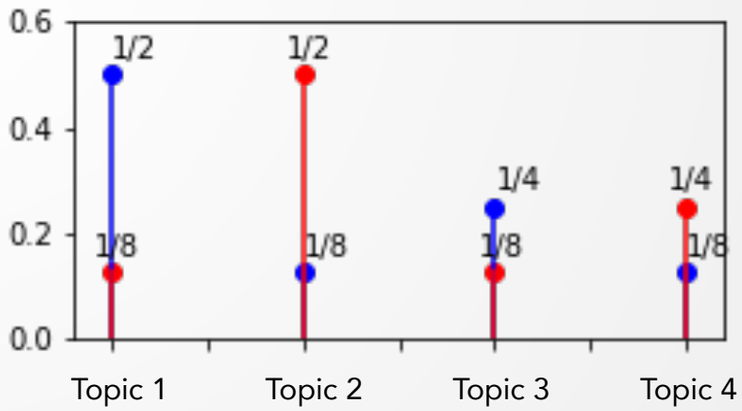
Hellinger Distance for LDA

- For $z_p = (z_{p,1}, \ldots, z_{p,k}, \ldots, z_{p,K})$ and $z_q = (z_{q,1}, \ldots, z_{q,k}, \ldots, z_{q,K})$

$$d_H(z_p, z_q, X) = \frac{1}{\sqrt{2}} \sqrt{\sum_{k=1}^{K} (\sqrt{z_{p,k}} - \sqrt{z_{q,k}})^2}$$

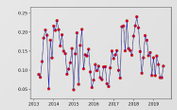
- Hellinger Distance between two News with K topics
- Example, given $K = 4$:





Topic distribution for document 1 and document 2

Wasserstein distance



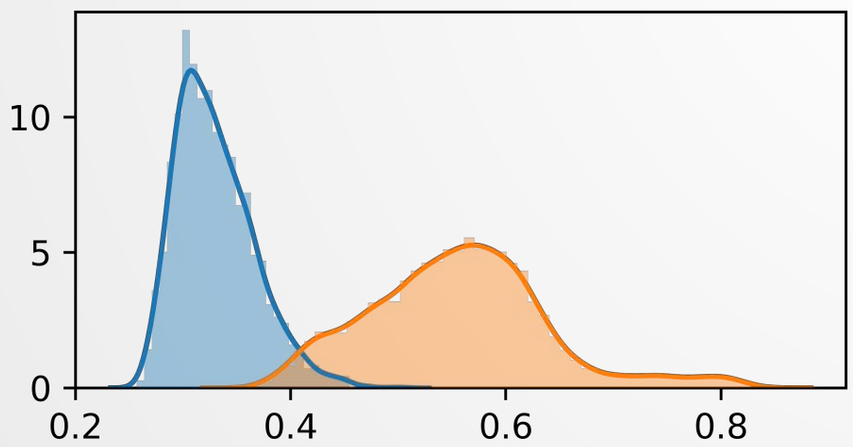
LDA Classification

- LDA model is trained with full pre-classified news data
- Calculate the average distance between policy-related news, for regulatory news i:

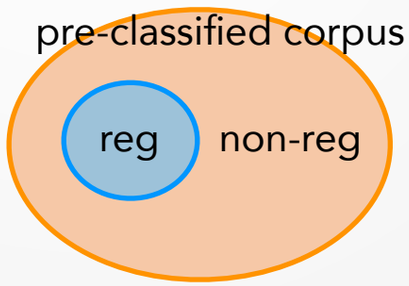
$$\bar{d}_{r,i} = \frac{1}{N_r} \sum_{j=1}^{N_r} d_H(z_{r,i}, z_{r,j})$$

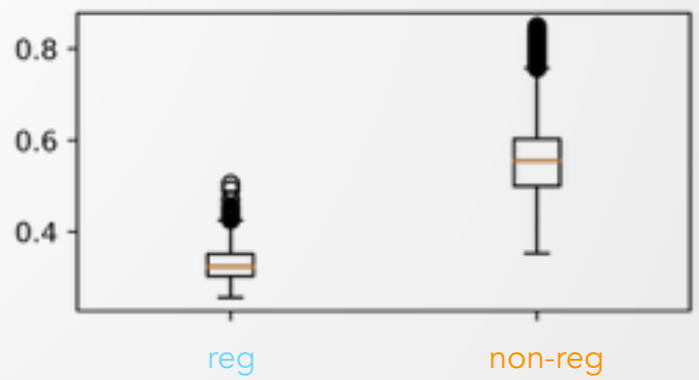
where Nr is the number of regulatory news, $z_{r,i}$ and $z_{r,j}$ represents topic distribution of regulatory new i and j

- Values closer to 0 indicate a smaller ‘distance’ and therefore a larger similarity

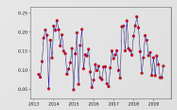


Figure： distribution of average distances (reg vs non-reg)



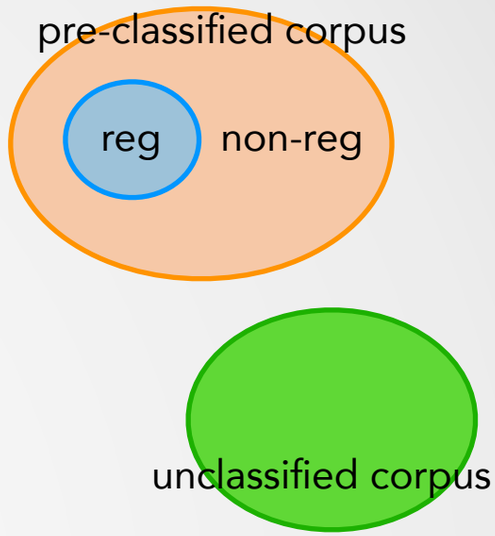


Figure： boxplot of average distances



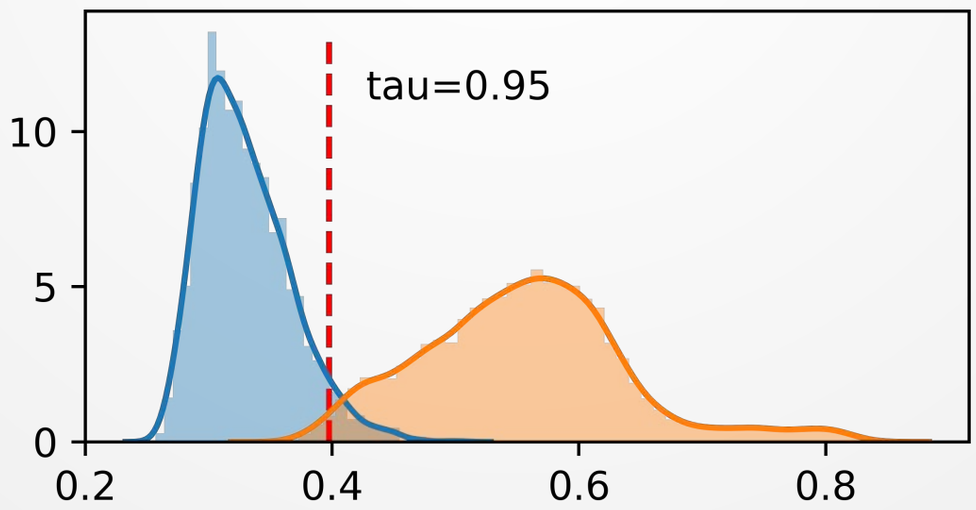
LDA Classification (Cont'd)

- Project unclassified document l to the trained LDA model
- Calculate its averages distance with policy-related news



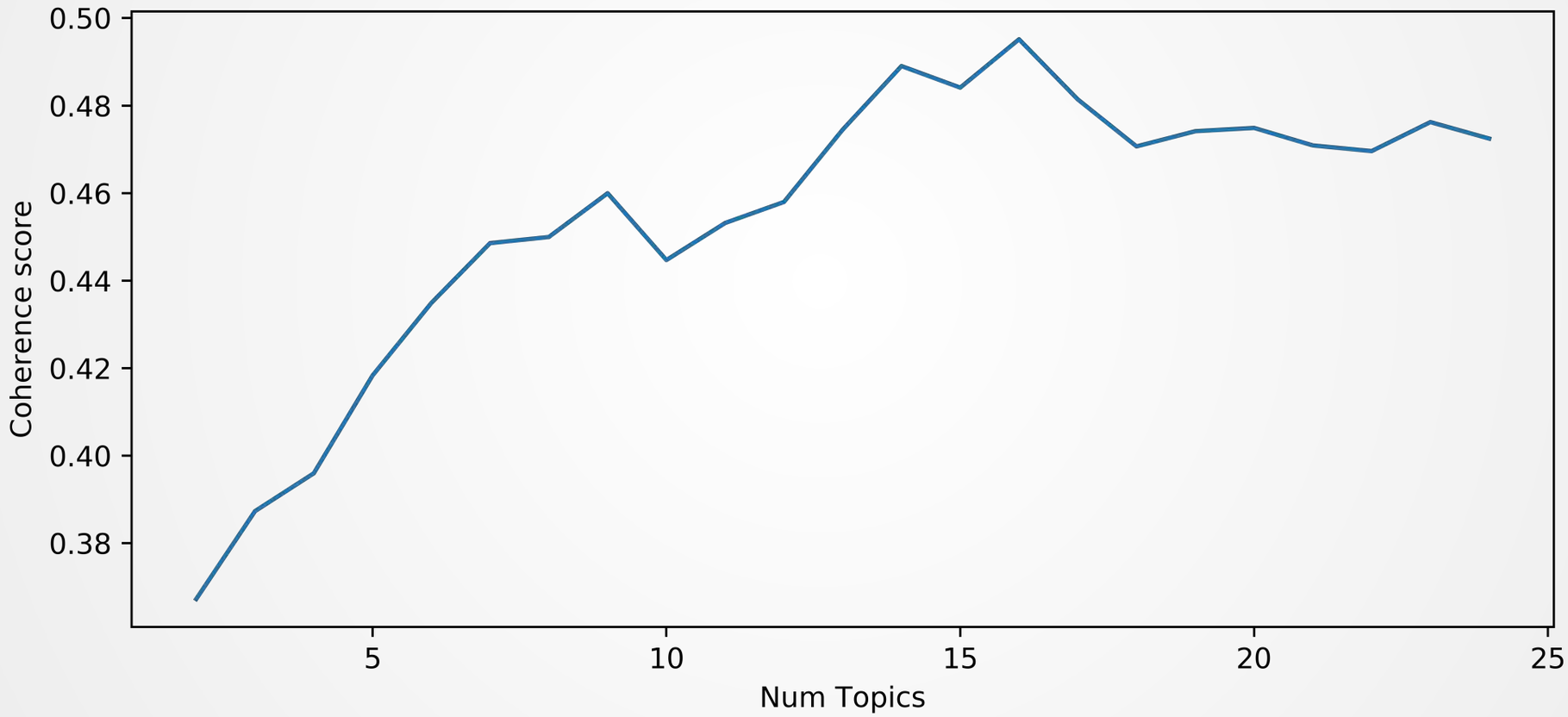
$$\bar{d}_{u,l} = \frac{1}{N_r} \sum_{j=1}^{N_r} d_H(z_{u,l}, z_{r,j})$$

- Set up a distance boundary $\bar{d}$
- If $\bar{d}_{u,l} < \bar{d}$, then label the unclassified news l as policy-related news



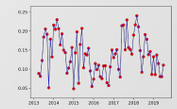
LDA Results: Number of Topics

- Coherence Value: based on word co-occurrence
- Comparable to the human rating of topics

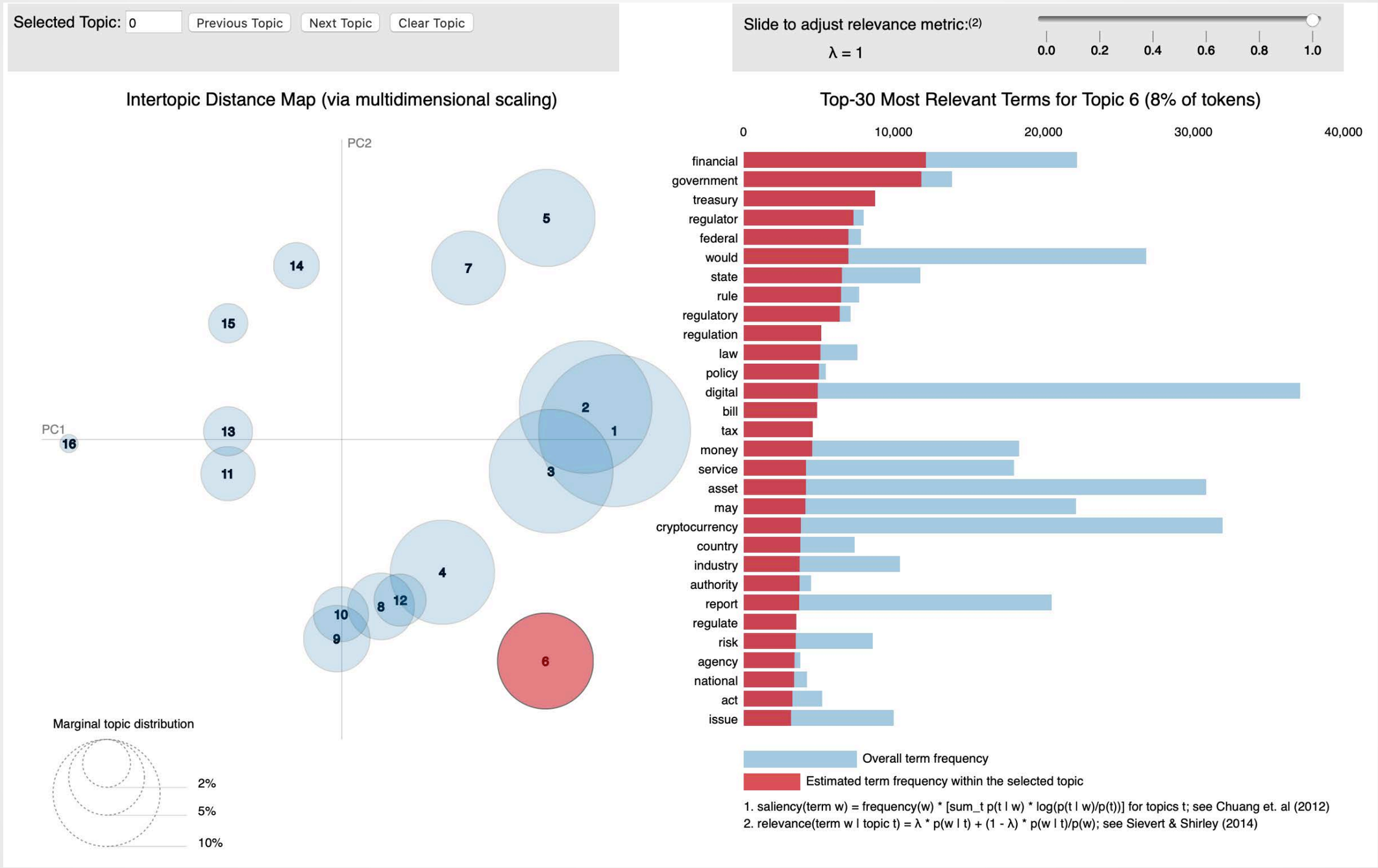


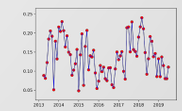
- K=16

Coherence

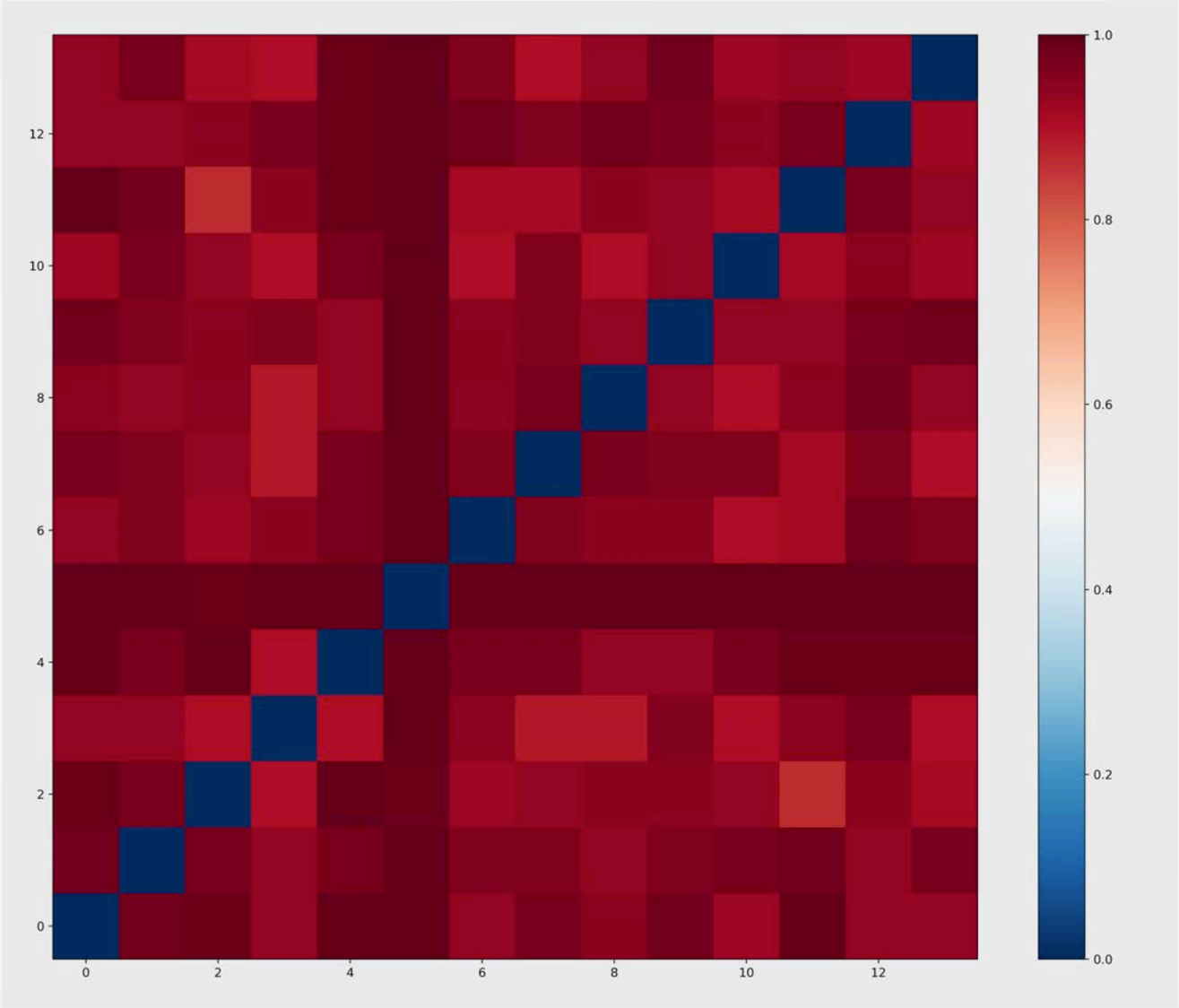


LDA Results: Topics





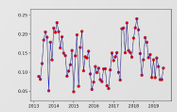
Distance Between topics



LDA Results: Topics

Coindesk Subcategory	LDA Topic	Top Keyword
Opinions	Topic 1: Opinions	Bitcoin, say, people, make, go get, take, would, could, way
Tech	Topic 3: Technology	System, blockchain, use, transaction, chain Technology, security, work, datum, network
Business	Topic 2: Business	Company, say, business, new, service base, startup, firm, founder, ceo
Policy & Regulation	Topic 6: Regulation	Financial, government, regulatory, treasury, federal State, rule, law, policy, digital
	Topic 9: Government	Bitcoin, foundation, member, government, event state, conference, donation, say, technology
	Topic 10: Crime	Mt Gox, site, silk road, claim, report Attack, withdraw, website, steal, time
Market	Topic 7: Investment	Bitcoin, market, currency, price, exchange value, investor, Litecoin, trade, investment
	Topic 13: Trade and exchange	Exchange, BTC, account, customer, trading User, deposit, page, trade, fund

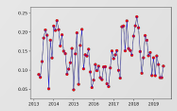
Figure: Categories (Coindesk.com) matched by LDA topics



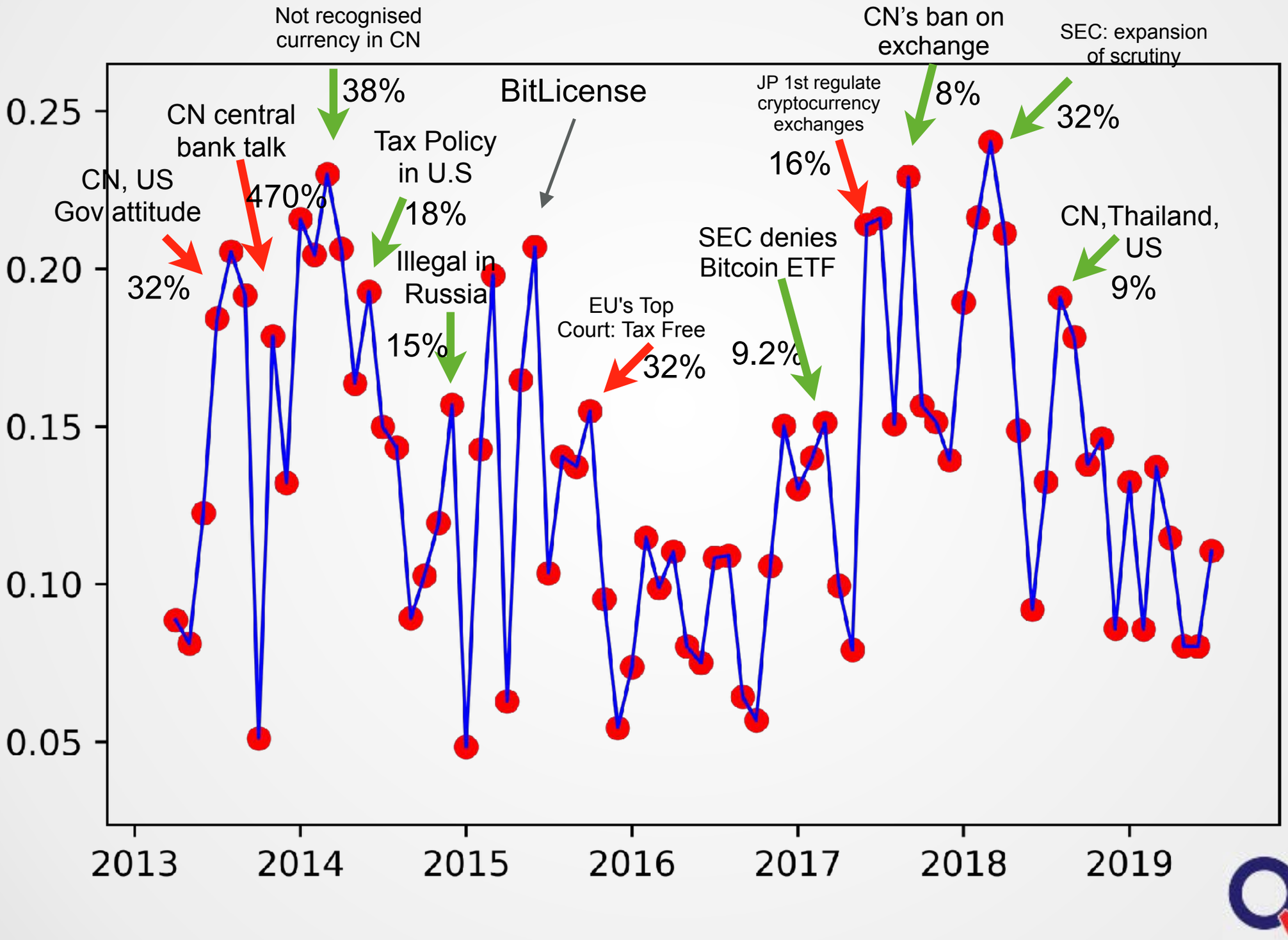
LDA: Classification

- Confusion matrices of classification based on LDA
- Accuracy: 0.907
- Relatively high type I error: 38%

True\Pred	NB		$SVM_{weighted}$		LDA		Total
	1	0	1	0	1	0	
1	0	582	0	582	361	221	582
0	0	4004	0	4004	188	3816	4004
Total	0	4586	0	4586	549	4073	4586
Accuracy	0.873		0.873		0.907		



Cryptocurrency Regulatory Risk IndeX (CRRIX)

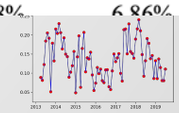


Daily- and Event- Based Study

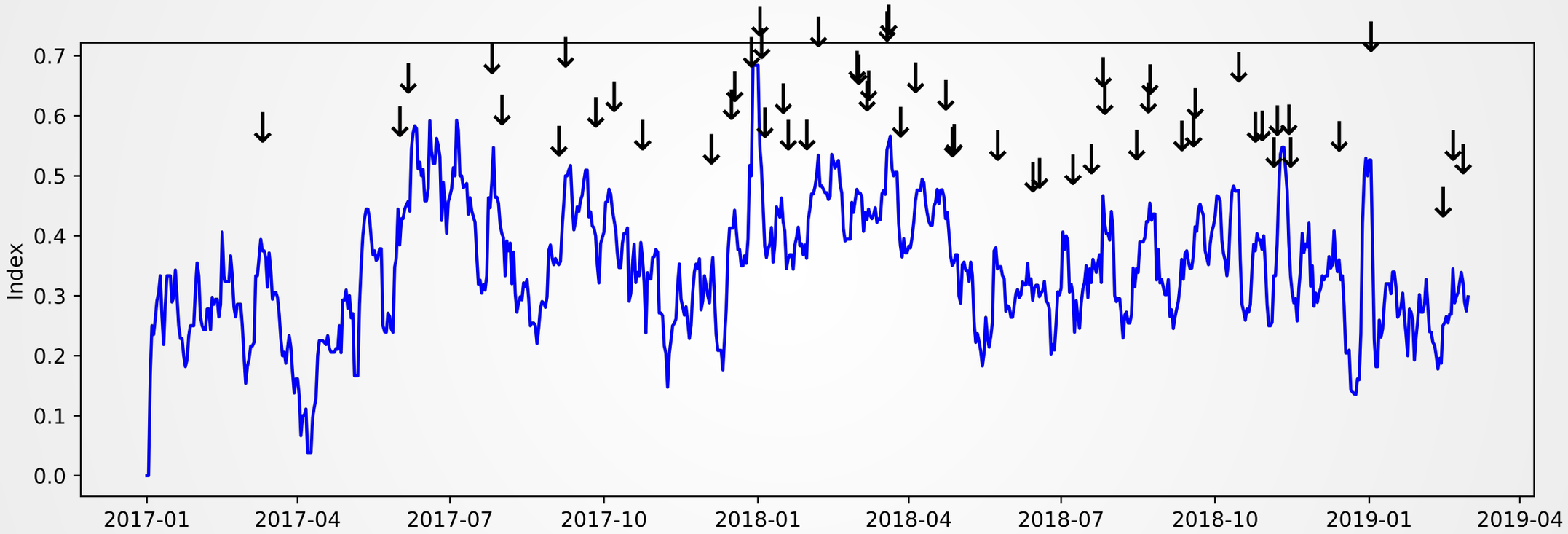
- Shanaev et. al, 2020 (Research in International Business and Finance)
- 5 regulatory classes:
  - anti-money laundering
  - exchange regulation
  - issuance regulation
  - risk concerns
  - state-backed issuance

Table 1. Full sample of regulatory events (01/01/2017 – 18/03/2019).

Date	Country	Brief event description	Regulatory event class	Dynamic	Market return	Volatility
11/03/2017	US	SEC: Bitcoin ETF proposal rejected	Issuance regulation	Negative	6.67%	15.12%
01/06/2017	China	Chinese exchange allows cryptocurrency withdrawals, moratorium ends	Exchange regulation	Positive	8.17%	15.22%
06/06/2017	Russia	Putin meets Buterin, wants to create state-backed coin based on Ethereum	State-backed issuance	Positive	5.17%	16.25%
26/07/2017	US	SEC: ICOs must abide by federal securities' laws	Issuance regulation	Negative	-1.28%	10.80%
01/08/2017	Russia	Banks start using Ethereum-based blockchain to process payments	State-backed issuance	Positive	1.68%	14.24%
08/08/2017	Russia	Russia plans and ICO of a state-backed cryptocurrency	State-backed issuance	Positive	4.47%	22.66%
17/08/2017	Australia	Australia plans to tighten AML crypto regulation	Anti-money laundering	Negative	1.90%	23.12%
22/08/2017	Estonia	State-backed "Estcoin" cryptocurrency proposed	State-backed issuance	Positive	3.61%	15.24%
04/09/2017	China	ICO ban	Issuance regulation	Negative	-11.34%	31.92%
07/09/2017	Estonia	Mario Draghi criticises Estcoin	State-backed issuance	Negative	0.70%	41.18%
08/09/2017	Russia	Russia will regulate cryptocurrencies as securities	Anti-money laundering	Negative	-9.04%	11.63%
14/09/2017	Russia	Central Bank of Russia: Cryptocurrencies are not a good idea	Risk concerns	Negative	-21.12%	46.62%
15/09/2017	China	Cryptocurrency exchanges ordered to wind down their operations	Exchange regulation	Negative	15.88%	15.68%
29/09/2017	South Korea	ICO ban	Issuance regulation	Negative	-1.34%	41.32%
29/09/2017	Japan	Japan: endorsement to 11 cryptocurrency exchanges	Exchange regulation	Positive	-1.34%	41.32%
10/10/2017	Russia	CBR: Websites selling cryptocurrencies will be blocked	Anti-money laundering	Negative	1.14%	28.20%
15/10/2017	UK	FCA warns retail traders against cryptocurrencies	Risk concerns	Negative	-1.68%	9.63%
24/10/2017	Singapore	Singapore is not planning to regulate cryptocurrencies	Risk concerns	Positive	-1.67%	44.57%
04/12/2017	UK	UK: money laundering concerns	Anti-money laundering	Negative	4.06%	19.69%
04/12/2017	UK	Treasury considers Bitcoin regulation due to money laundering concerns	Anti-money laundering	Negative	4.06%	19.69%
11/12/2017	US	Bitcoin futures start trading on CBOE	Issuance regulation	Positive	11.32%	16.90%
12/12/2017	US	US regulators warn cryptocurrency investors of potential risks	Risk concerns	Negative	10.85%	16.67%
13/12/2017	South Korea	Plans to tax capital gains from cryptocurrency trading	Risk concerns	Negative	-0.85%	17.29%
14/12/2017	UK	FCA sees no systematic risk in bitcoin and no need to regulate it	Risk concerns	Positive	9.44%	16.67%
16/12/2017	South Korea	North Korea hacking South Korean cryptocurrency exchanges	Anti-money laundering	Negative	9.34%	39.15%
18/12/2017	Germany	Germany joins France for G-20 based cryptocurrency regulation	Anti-money laundering	Negative	5.79%	33.92%
20/12/2017	UK	Carney is not worried about the systemic risk of cryptocurrencies	Issuance regulation	Positive	2.60%	25.24%
22/12/2017	Belarus	Belarus creates a tax-free cryptocurrency hub	Anti-money laundering	Positive	-14.33%	15.11%
28/12/2017	South Korea	South Korea: tightening of cryptocurrency regulation is expected	Anti-money laundering	Negative	-5.41%	25.61%
02/01/2018	Russia	Russia considers a state-backed "cryptorouble"	State-backed issuance	Positive	9.80%	17.64%
03/01/2018	China	China will restrict electricity access to miners	Risk concerns	Negative	11.09%	18.75%
05/01/2018	UK	BoE clarifies it is not launching its own cryptocurrency	State-backed issuance	Negative	5.10%	38.22%
09/01/2018	Venezuela	Venezuela's parliament outlaws the sovereign cryptocurrency	State-backed issuance	Negative	-0.79%	15.99%
11/01/2018	South Korea	South Korea considers ban on cryptocurrency trading	Anti-money laundering	Negative	-9.06%	13.01%
15/01/2018	China	China banning cryptocurrency exchange substitutes	Exchange regulation	Negative	-3.28%	9.78%
15/01/2018	South Korea	Regulators uncertain whether cryptocurrency exchanges should be banned	Exchange regulation	Positive	-3.28%	9.78%
16/01/2018	Russia	Russia plans to let Bitcoin trade on official exchanges	Exchange regulation	Positive	-20.82%	8.31%
19/01/2018	Switzerland	Switzerland encourages ICOs	Issuance regulation	Positive	0.76%	10.02%
22/01/2018	South Korea	Anonymous cryptocurrency trades banned	Anti-money laundering	Negative	-5.04%	8.02%
25/01/2018	Nigeria	Nigeria's central bank warns against cryptocurrency risks	Risk concerns	Negative	-0.42%	13.53%
25/01/2018	UK	May calls to look seriously into crypto due to money laundering concerns	Anti-money laundering	Negative	-0.42%	13.53%
29/01/2018	Japan	Cryptocurrency exchange investigations after massive fraud	Exchange regulation	Negative	-4.38%	12.70%
30/01/2018	US	Cryptocurrency advertisement ban on Facebook	Issuance regulation	Negative	-11.53%	8.93%
30/01/2018	US	Bitfinex and Tether subpoenaed	Exchange regulation	Negative	-11.53%	8.93%
31/01/2018	South Korea	Finance minister promises exchanges will not be banned	Exchange regulation	Positive	2.55%	7.88%
31/01/2018	Hong Kong	Hong Kong plans to regulate cryptocurrencies	Anti-money laundering	Negative	2.55%	7.88%
01/02/2018	India	India: money laundering concerns	Anti-money laundering	Negative	-11.59%	7.22%
05/02/2018	UK	Virgin Money bans cryptocurrency purchases with its credit cards	Anti-money laundering	Negative	-17.18%	6.86%
05/02/2018	UK	Lloyds bans cryptocurrency purchases	Anti-money laundering	Negative	-17.18%	6.86%

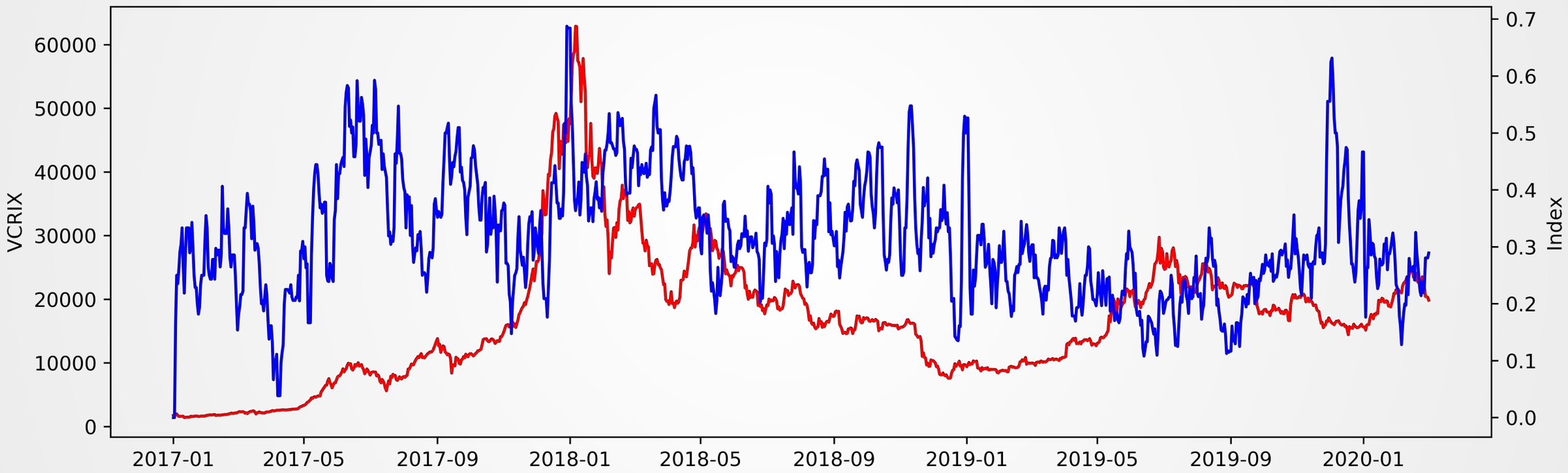


CRRIX and Events



Figure： Events and Cryptocurrency Regulatory Risk IndeX (CRRIX)

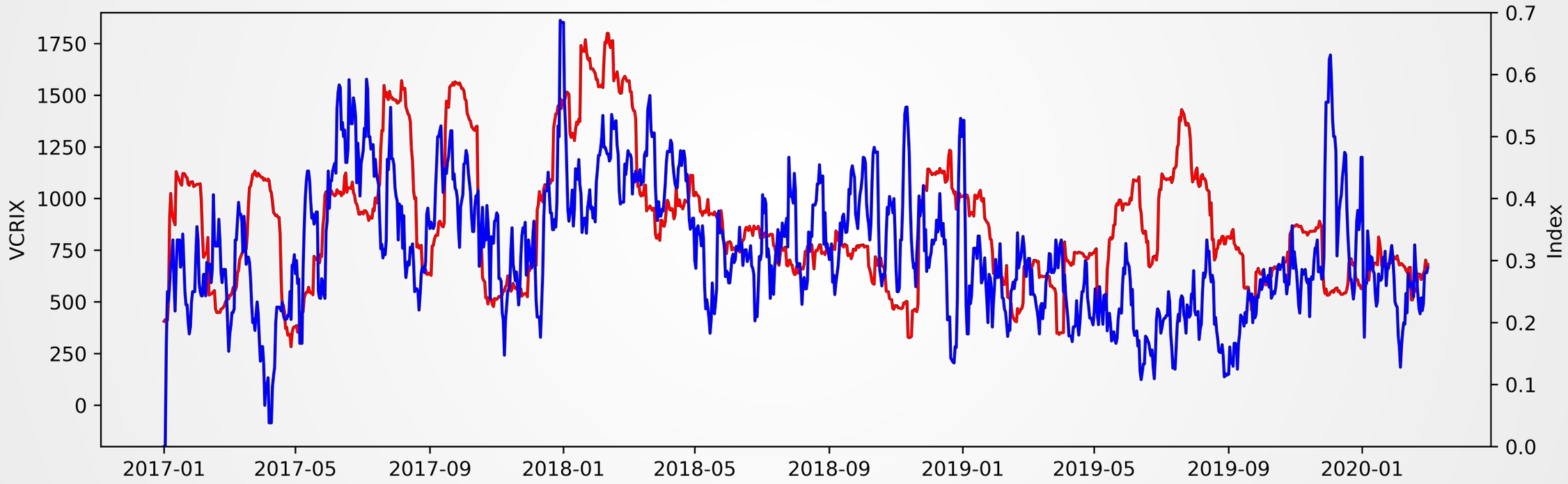
CRIX and CRRIX



Figure： CRIX and Cryptocurrency Regulatory Risk IndeX (CRRIX)

VCRIX and CRRIX

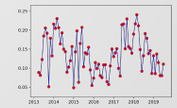
- VCRIX: stationary at 1% level (ADF test p=0.000067)
- CRRIX: stationary at 5% level (ADF test p=0.019296)



Figure： VCRIX and Cryptocurrency Regulatory Risk IndeX (CRRIX)

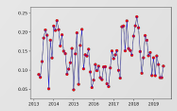
Granger Causality Test: VCRIX and CRRIX

```
In [21]: gc_res = grangercausalitytests(test_data, maxlag=4, verbose=True)

Granger Causality
number of lags (no zero) 1
ssr based F test:         F=23.1736 , p=0.0000  , df_denom=825, df_num=1
ssr based chi2 test:   chi2=23.2579 , p=0.0000  , df=1
likelihood ratio test: chi2=22.9372 , p=0.0000  , df=1
parameter F test:         F=23.1736 , p=0.0000  , df_denom=825, df_num=1

Granger Causality
number of lags (no zero) 2
ssr based F test:         F=8.0382  , p=0.0003  , df_denom=822, df_num=2
ssr based chi2 test:   chi2=16.1743 , p=0.0003  , df=2
likelihood ratio test: chi2=16.0181 , p=0.0003  , df=2
parameter F test:         F=8.0382  , p=0.0003  , df_denom=822, df_num=2

Granger Causality
number of lags (no zero) 3
ssr based F test:         F=8.4604  , p=0.0000  , df_denom=819, df_num=3
ssr based chi2 test:   chi2=25.5981 , p=0.0000  , df=3
likelihood ratio test: chi2=25.2094 , p=0.0000  , df=3
parameter F test:         F=8.4604  , p=0.0000  , df_denom=819, df_num=3

Granger Causality
number of lags (no zero) 4
ssr based F test:         F=6.4008  , p=0.0000  , df_denom=816, df_num=4
ssr based chi2 test:   chi2=25.8855 , p=0.0000  , df=4
likelihood ratio test: chi2=25.4877 , p=0.0000  , df=4
parameter F test:         F=6.4008  , p=0.0000  , df_denom=816, df_num=4
```



Conclusion

- CRRIX: an Regulatory Index for CCs market using news data
- Applying LDA method to classify policy-related and non-policy-related news
- The index successfully capture the big historical policy changes
- Tested causality between CRRIX and market volatility index VCRIX
- Helpful for the market forecasting

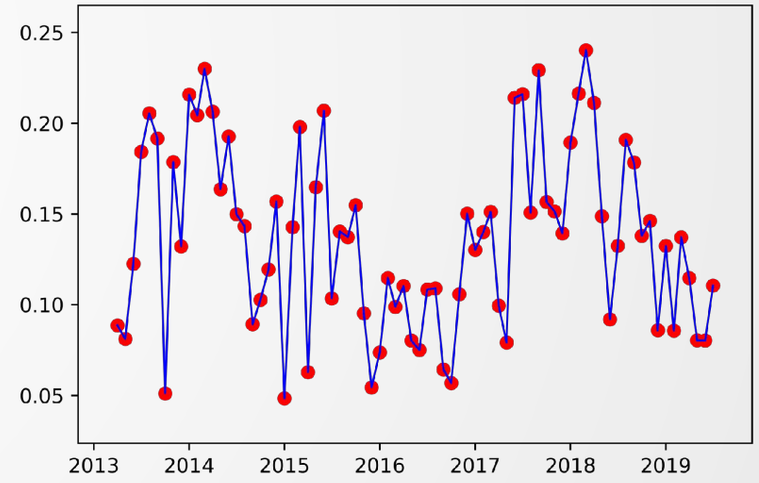


CRRIX: A Machine Learning Based Regulatory Risk Index for Cryptocurrencies

Xinwen Ni

Wolfgang Karl Härdle

Taojun Xie

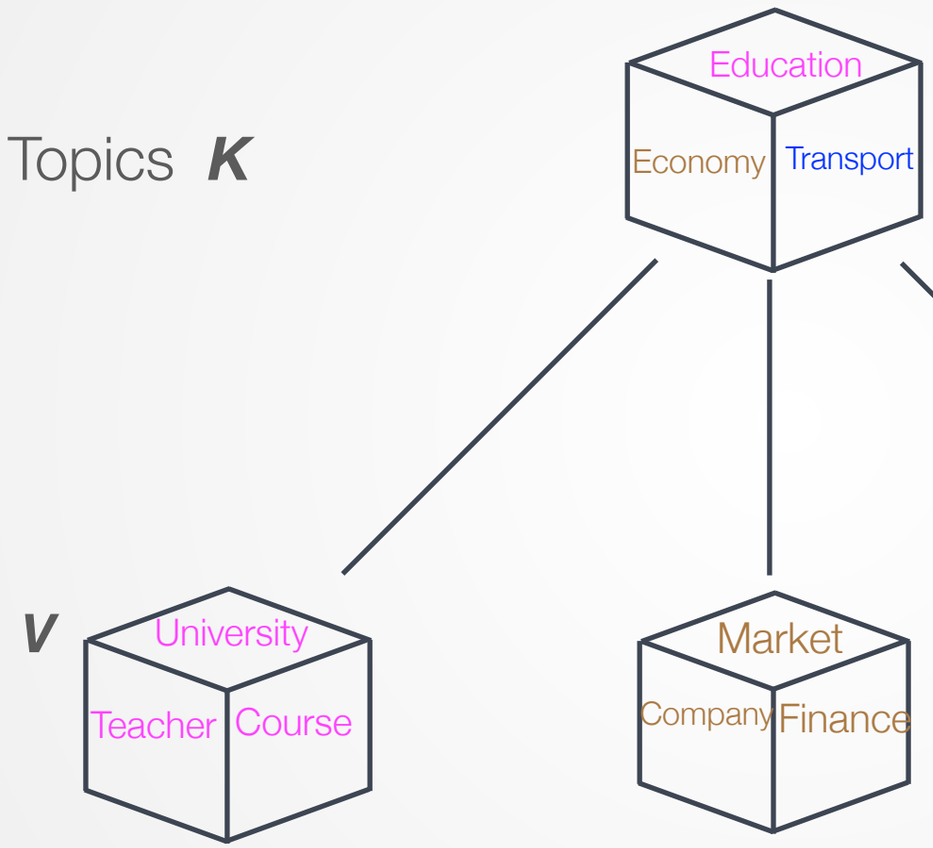


Ladislaus von Bortkiewicz Professor of Statistics
Humboldt-Universität zu Berlin
BRC Blockchain Research Center
lvb.wiwi.hu-berlin.de
Charles University, WISE XMU, NCTU 玉山学者

Probabilistic Latent Semantic Analysis (PLSA)

Documents ***M*** $p(d_m)$

Topics ***K***

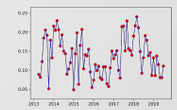


$\theta_{m,k} = p(z_k|d_m)$

$\phi_{k,n} = p(w_n|z_k)$

Vocabulary ***V***

$$p(d_m, w_n) = p(d_m)p(w_n|d_m) = p(d_m)\sum_{k=1}^{K} \underbrace{p(w_n|z_k)}_{\phi} \underbrace{p(z_k|d_m)}_{\theta}$$

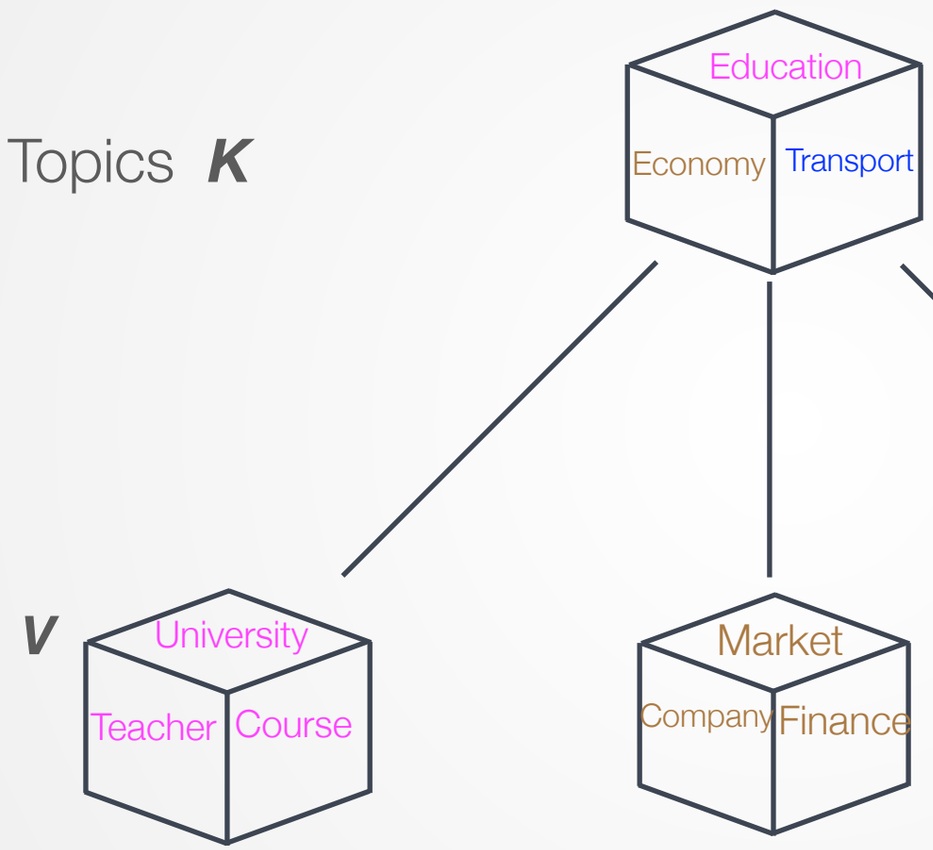


LDA vs PLSA

Documents ***M*** $p(d_m)$

$$\theta_m \sim Dir\{\alpha\}$$

Topics ***K***



$$\theta_{m,k} = p(z_k|d_m)$$

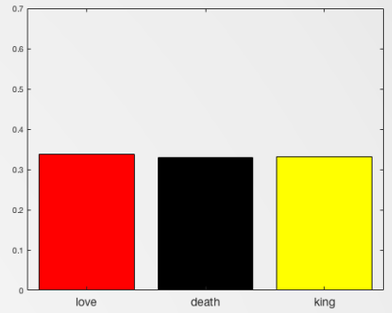
$$\phi_k \sim Dir\{\beta\}$$

$$\phi_{k,n} = p(w_n|z_k)$$

Vocabulary ***V***

Back

Dirichlet Distribution and LDA

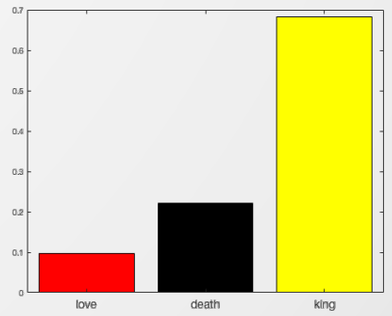


- Text I: Dirichlet parameter $\beta = (1, 1, 1)$

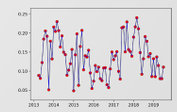
love love death king love love king king love love king king death king death king love king death king love love death king king king death death love love death king king love death love king king death death death king love king love love death king king love love death death death death love king king death king king love king king love king king king king love death king love king death king death king love king death love king king death love death love death death love king death king love death love king king death

- Text II： Dirichlet parameter $\beta = (2, 5, 15)$

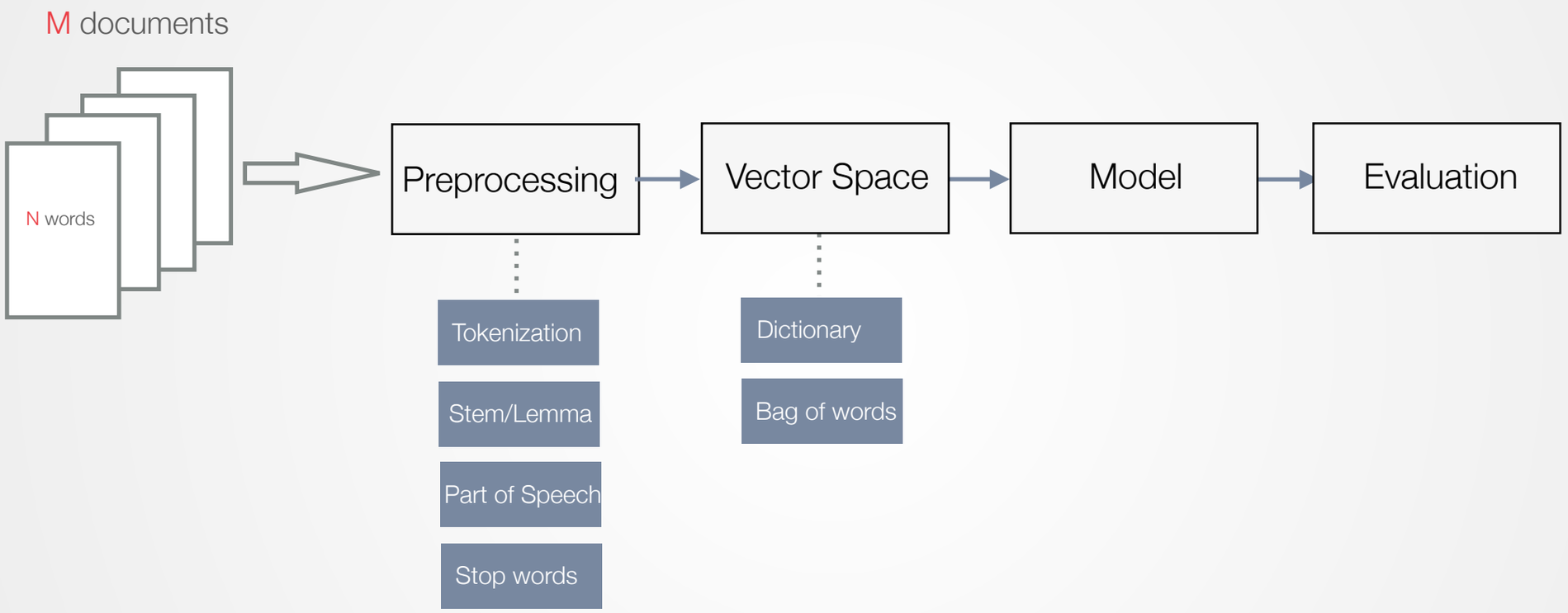
death king king king king king king death king king king love love death king king king king death king love king king king death king death death king king king king love king king king king king king death king death king king king king death king king death king king king king king king death king king king king king king death king king king death death king king king king king king death king king love king death king death death king king king death

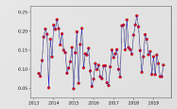


Back



Pipeline for Topic Modeling





Hellinger Distance: examples

- For two normal distributions $P \sim \mathcal{N}\left(\mu_1, \sigma_1^2\right)$ and $Q \sim \mathcal{N}\left(\mu_2, \sigma_2^2\right)$

$$H^2(P,Q) = 1 - \sqrt{\frac{2\sigma_1\sigma_2}{\sigma_1^2+\sigma_2^2}}\, e^{-\frac{1}{4}\frac{(\mu_1-\mu_2)^2}{\sigma_1^2+\sigma_2^2}}.$$

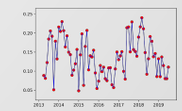
- For two multivariate normal distributions $P \sim \mathcal{N}\left(\mu_1, \Sigma_1\right)$ and $Q \sim \mathcal{N}\left(\mu_2, \Sigma_2\right)$

$$H^2(P,Q) = 1 - \frac{\det(\sum_1)^{1/4}\det(\sum_2)^{1/4}}{\det\left(\frac{\sum_1+\sum_2}{2}\right)^{1/2}} \exp\left\{-\frac{1}{8}(\mu_1-\mu_2)^T\left(\frac{\sum_1+\sum_2}{2}\right)^{-1}(\mu_1-\mu_2)\right\}$$

- For two exponential distributions $P \sim \mathrm{Exp}(\alpha)$ and $Q \sim \mathrm{Exp}(\beta)$

$$H^2(P,Q) = 1 - \frac{2\sqrt{\alpha\beta}}{\alpha+\beta}.$$

Back



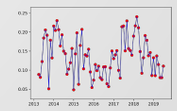
Hellinger Distance: examples (Cont'd)

- For two Poisson distributions $P \sim \text{ Poisson } (\alpha)$ and $Q \sim \text{ Poisson } (\beta)$

$$H^2(P,Q) = 1 - e^{-\frac{1}{2}(\sqrt{\alpha}-\sqrt{\beta})^2}$$

- For two Beta distributions $P \sim \text{Beta}\,(a_1, b_1)$ and $Q \sim \text{Beta}\,(a_2, b_2)$

$$H^2(P,Q) = 1 - \frac{B\left(\frac{a_1+a_2}{2}, \frac{b_1+b_2}{2}\right)}{\sqrt{B(a_1,b_1)B(a_2,b_2)}}$$



Distances

- Let X~P and Y~Q and let the densities be p and q, then distance between P and Q

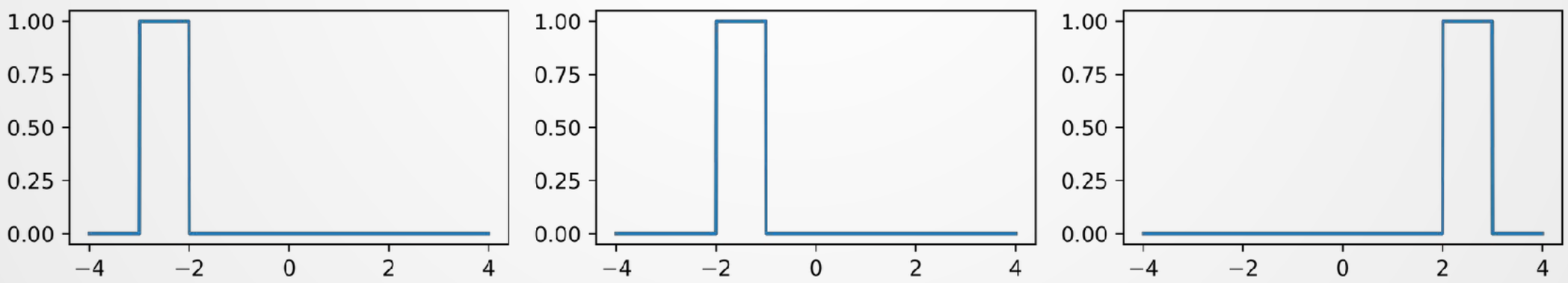
$$\text{Total Variation}: \sup_A |P(A) - Q(A)| = \tfrac{1}{2}\int |p - q|$$

$$\text{Hellinger}: \sqrt{\tfrac{1}{2}\int (\sqrt{p} - \sqrt{q})^2}$$

$$L_2 : \int (p - q)^2$$

$$\chi^2 : \int \frac{(p-q)^2}{q}$$

- These distances ignore the underlying geometry of the space



$H_{12}=H_{13}=H_{23}=1$ $W_{12}= 1 < W_{23}= 4 < W_{13}= 5$

Back

Asymptotic distribution of $\sqrt{nh}[\hat{f}(x) - f(x)]$ depends on $f(x)$

- Denote $\hat{f}(x)$ is a kernel density estimator,

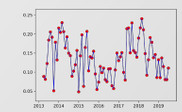
$$\sqrt{nh}[\hat{f}(x) - f(x)] = \sqrt{nh}\{\hat{f}(x) - \mathrm{E}[\hat{f}(x)]\} + \sqrt{nh}\{\mathrm{E}[\hat{f}(x)] - f(x)\}$$

For the second part

$$\begin{aligned}
\mathrm{Bias}\left\{\hat{f}_h(x)\right\} &= \mathrm{E}\left\{\hat{f}_h(x)\right\} - f(x) \\
&= \frac{1}{n}\sum_{i=1}^{n} \mathrm{E}\left\{K_h\left(x - X_i\right)\right\} - f(x) \\
&= \mathrm{E}\left\{K_h(x - X)\right\} - f(x) \\
&= \int \frac{1}{h} K\left(\frac{x-u}{h}\right) f(u)du - f(x)
\end{aligned}$$

The transformation $s = \frac{u-x}{h}$, i.e. $u = hs + x, \left|\frac{ds}{du}\right| = \frac{1}{h}$. A second-order Taylor expansion of $f(u)$ around x is given by

$$f(x + hs) = f(x) + f(x)'hs + \frac{1}{2}f''(x)h^2s^2 + o\left(h^2\right)$$

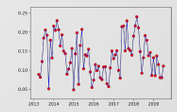


Asymptotic distribution of $\sqrt{nh}[\hat{f}(x) - f(x)]$ depends on $f(x)$

- Then

$$\begin{aligned}
\text{Bias}\left\{\hat{f}_h(x)\right\} &= \int \frac{1}{h} K(-s) f(x+hs) h ds - f(x) \\
&= \int K(s) \left[f(x) + f(x)' hs + \frac{1}{2} f''(x) h^2 s^2 + o\left(h^2\right) \right] ds - f(x) \\
&= f(x) \int K(s) ds + f(x)' h \int s K(s) ds + \frac{1}{2} f''(x) h^2 \int s^2 K(s) ds - f(x) + o\left(h^2\right) \\
&= \frac{h^2}{2} f''(x) \mu_2(K) + o\left(h^2\right), \quad \text{as } h \to 0
\end{aligned}$$

where $\int K(s)ds = 1, \int sK(s)ds = 0$ and $\int s^2 K(s)ds = \mu_2(K)$

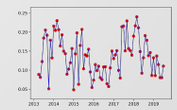


Asymptotic distribution of $\sqrt{nh}[\hat{f}(x) - f(x)]$ depends on $f(x)$

$$\sqrt{nh}[\hat{f}(x) - f(x)] = \sqrt{nh}\{\hat{f}(x) - \mathrm{E}[\hat{f}(x)]\} + \sqrt{nh}\{\mathrm{E}[\hat{f}(x)] - f(x)\}$$

- For the first part,

$$\begin{aligned}
\mathrm{Var}\left\{\hat{f}_h(x)\right\} &= \mathrm{Var}\left\{\frac{1}{n}\sum_{i=1}^{n} K_h\left(x - X_i\right)\right\} \\
&= \frac{1}{n^2}\sum_{i=1}^{n} \mathrm{Var}\left\{K_h\left(x - X_i\right)\right\} \\
&= \frac{1}{n}\mathrm{Var}\left\{K_h(x - X)\right\} \\
&= \frac{1}{n}\left\{\mathrm{E}\left[K_h^2(x - X)\right] - \left\{\mathrm{E}\left[K_h(x - X)\right]\right\}^2\right\} \\
&= \frac{1}{n}\int \frac{1}{h^2} K\left(\frac{x - t}{h}\right)^2 f(t)dt - \frac{1}{n}\left(\frac{1}{h}\int K\left(\frac{x - t}{h}\right) f(t)dt\right)^2 \\
&= \frac{1}{n}\int \frac{1}{h^2} K\left(\frac{x - t}{h}\right)^2 f(t)dt - \frac{1}{n}(f(x) + \mathrm{Bias}(\hat{f}(x)))^2
\end{aligned}$$



Asymptotic distribution of $\sqrt{nh}[\hat{f}(x) - f(x)]$ depends on $f(x)$

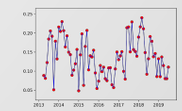
$$\text{Substituting } s = \tfrac{u-x}{h}$$
$$\text{Var}\left\{\hat{f}_h(x)\right\} = \tfrac{1}{nh}\int K(s)^2 f(x+hs)ds - \tfrac{1}{n}(f(x)+o(h))^2$$

- Applying a Taylor approximation yields

$$\begin{aligned}\text{Var}\left\{\hat{f}_h(x)\right\} &= \frac{1}{nh}\int K(z)^2\left(f(x)+hsf'(x)+\frac{1}{2}f''(x)h^2s^2+o\left(h^2\right)\right)ds - \frac{1}{n}(f(x)+o(h))^2 \\ &= \frac{1}{nh}\|K\|_2^2 f(x) + o\left(\frac{1}{nh}\right), \qquad \text{as } nh\to\infty\end{aligned}$$

where $\int K^2(s)ds = \|K\|_2^2$. With $\text{Var}(\hat{f}(x)) \to 0$ as $nh \to \infty$

$$\frac{\hat{f}(x) - \text{E}[\hat{f}(x)]}{\sqrt{\text{Var}(\hat{f}(x))}} \xrightarrow{d} \mathcal{N}(0,1)$$



Asymptotic distribution of $\sqrt{nh}[\hat{f}(x) - f(x)]$ depends on $f(x)$

- Substituting the expression for $\mathrm{Var}(\hat{f}(x))$

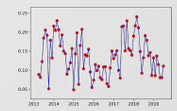
$$\sqrt{nh}\{\hat{f}(x) - \mathrm{E}[\hat{f}(x)]\} \xrightarrow{d} \mathcal{N}\left(0, f(x)\|K\|_2^2\right)$$

If the bandwidth tends to zero faster than the optimal rate, then

$$\sqrt{nh}\{\mathrm{E}[\hat{f}(x)] - f(x)\} \to 0$$

and the bias term vanishes from the asymptotic distribution,

$$\sqrt{nh}[\hat{f}(x) - f(x)] \xrightarrow{d} \mathcal{N}\left(0, f(x)\|K\|_2^2\right)$$



Asymptotic distribution of $\sqrt{nh}[\sqrt{\hat{f}(x)} - \sqrt{f(x)}]$ NOT depends on $f(x)$

- From transform theorems, we know that if $\sqrt{n}(t-\mu) \xrightarrow{\mathcal{L}} N_p(0, \Sigma)$

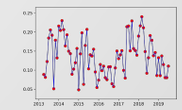
$$\sqrt{n}[f(t) - f(\mu)] \xrightarrow{\mathcal{L}} N_q\left(0, \mathcal{D}^\top \Sigma \mathcal{D}\right) \qquad \text{for } n \longrightarrow \infty$$

Denote $g(x) = x^{1/2}$, then $\frac{dg}{dx} = \frac{1}{2}x^{-1/2}$. With $\sqrt{nh}(\hat{f}(x) - f(x)) \xrightarrow{d} \mathcal{N}\left(0, f(x)\|K\|_2^2\right)$ then

$$\sqrt{nh}[\sqrt{\hat{f}(x)} - \sqrt{f(x)}] \xrightarrow{d} \mathcal{N}\left(0, \tfrac{1}{4}\|K\|_2^2\right)$$

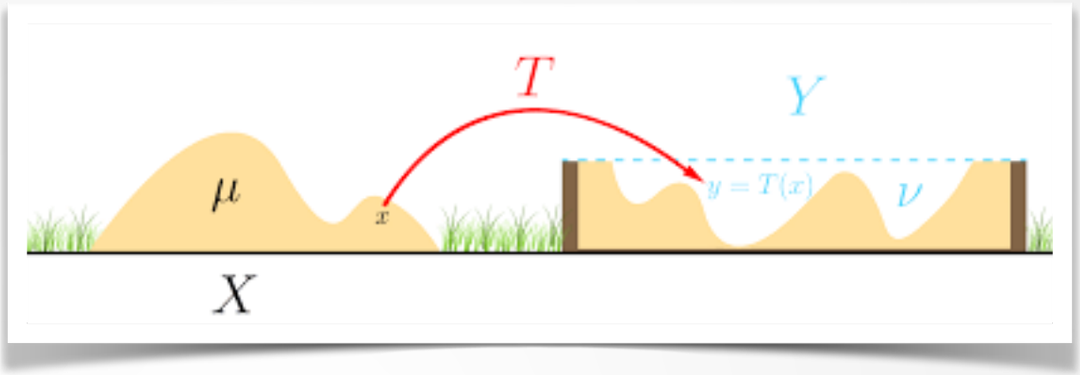
The asymptotic distribution of $\sqrt{nh}[\sqrt{\hat{f}(x)} - \sqrt{f(x)}]$ does not depend on $f(x)$

Back

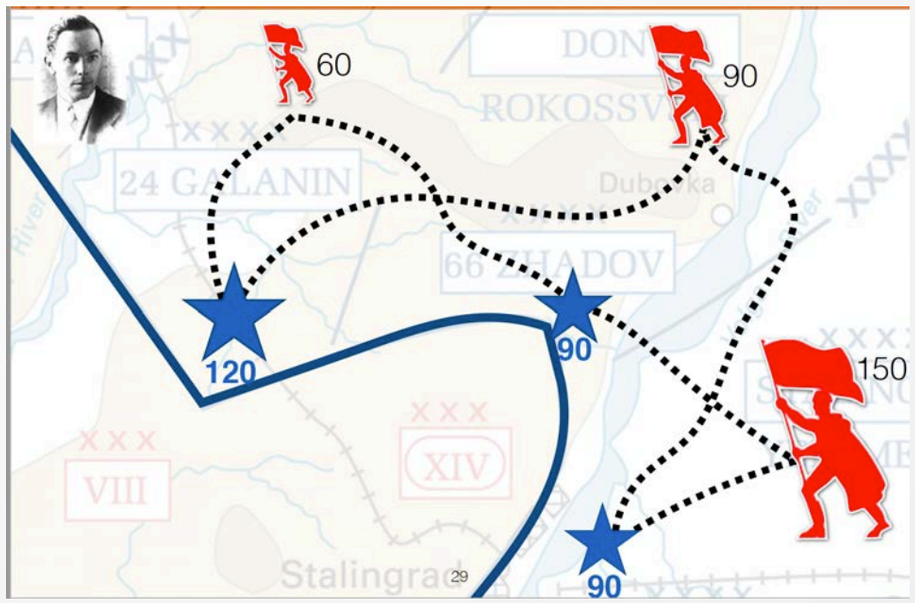


Optimal Transport and Wasserstein Distance

- Monge's Problem



- Kantorovich Problem

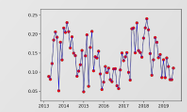


Wasserstein Distance

- Distance between probability distributions on a given metric space M
- Intuitively, if each distribution is viewed as a unit amount of "dirt" piled on M, the metric is the minimum "cost" of turning one pile into the other
- The amount of dirt that needs to be moved times the mean distance it has to be moved
- The p^{th} Wasserstein distance between two probability measures μ and ν in $P_p(M)$ is defined as

$$W_p(\mu,\nu) := \left(\inf_{\gamma \in \Gamma(\mu,\nu)} \int_{M\times M} d(x,y)^p \, \mathrm{d}\gamma(x,y) \right)^{1/p},$$

where Γ(μ,ν) denotes the collection of all measures on M×M with marginals μ and ν on the first and second factors respectively



Wasserstein Distance

- 1th Wasserstein distance between probability distribution μ and ν :

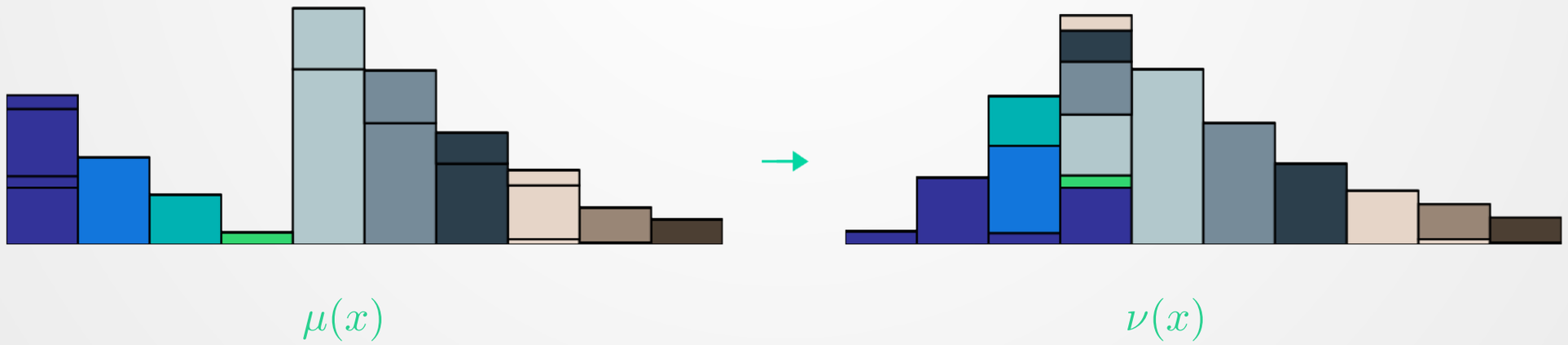
$$W(\mu,\nu) = \inf_{\gamma\in\Gamma(\mu,\nu)} \int_{M\times M} d(x,y)\,\mathrm{d}\gamma(x,y)$$

- γ>0 is the joint distribution, $\gamma \in \Gamma(\mu,\nu)$

$$\int \gamma(x,y)\mathrm{d}y = \mu(x)$$

$$\int \gamma(x,y)\mathrm{d}x = \nu(y)$$

$$W(\mu,\nu) = \inf_{\gamma\in\Gamma(\mu,\nu)} \iint d(x,y)\gamma(x,y)\,\mathrm{d}x\,\mathrm{d}y$$



d(x,y) is the moving cost

Wasserstein Distance (Cont'd)

- γ>0 is the joint distribution, $\gamma \in \Gamma(\mu, \nu)$

$$W(\mu,\nu) = \inf_{\gamma \in \Gamma(\mu,\nu)} \iint d(x,y)\gamma(x,y)\,\mathrm{d}x\,\mathrm{d}y$$

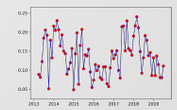
$$\text{s.t.} \quad \int \gamma(x,y)\mathrm{d}y = \mu(x)$$

$$\int \gamma(x,y)\mathrm{d}x = \nu(y)$$

- Let

$$\mathbf{D} = \begin{pmatrix} d(\boldsymbol{x}_1,\boldsymbol{y}_1) & d(\boldsymbol{x}_1,\boldsymbol{y}_2) & \cdots & d(\boldsymbol{x}_1,\boldsymbol{y}_n) \\ d(\boldsymbol{x}_2,\boldsymbol{y}_1) & d(\boldsymbol{x}_2,\boldsymbol{y}_2) & \cdots & d(\boldsymbol{x}_2,\boldsymbol{y}_n) \\ \vdots & \vdots & \vdots & \vdots \\ d(\boldsymbol{x}_m,\boldsymbol{y}_1) & d(\boldsymbol{x}_m,\boldsymbol{y}_2) & \cdots & d(\boldsymbol{x}_m,\boldsymbol{y}_n) \end{pmatrix} \qquad \boldsymbol{\Gamma} = \begin{pmatrix} \gamma(\boldsymbol{x}_1,\boldsymbol{y}_1) & \gamma(\boldsymbol{x}_1,\boldsymbol{y}_2) & \cdots & \gamma(\boldsymbol{x}_1,\boldsymbol{y}_n) \\ \gamma(\boldsymbol{x}_2,\boldsymbol{y}_1) & \gamma(\boldsymbol{x}_2,\boldsymbol{y}_2) & \cdots & \gamma(\boldsymbol{x}_2,\boldsymbol{y}_n) \\ \vdots & \vdots & \vdots & \vdots \\ \gamma(\boldsymbol{x}_m,\boldsymbol{y}_1) & \gamma(\boldsymbol{x}_m,\boldsymbol{y}_2) & \cdots & \gamma(\boldsymbol{x}_m,\boldsymbol{y}_n) \end{pmatrix}$$

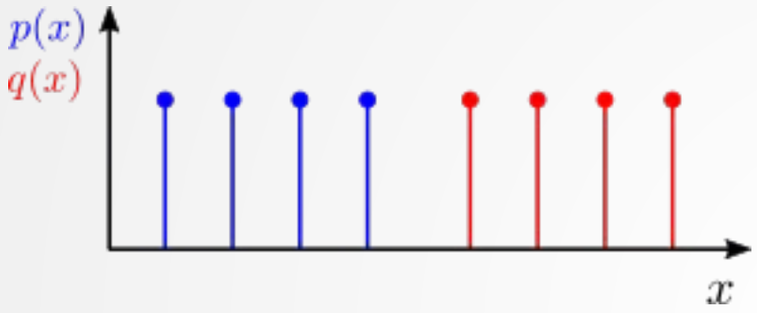
$$\iint d(x,y)\gamma(x,y)\,\mathrm{d}x\,\mathrm{d}y \quad \rightarrow \quad \langle \mathbf{D}, \boldsymbol{\Gamma} \rangle$$



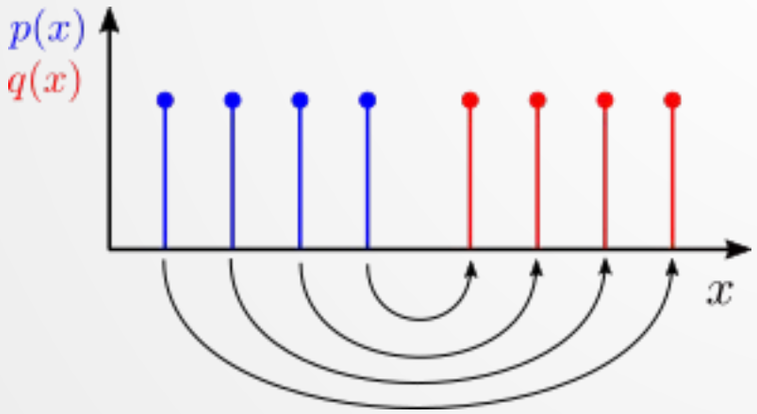
Wasserstein Distance: Discrete Example

- D is the distance (or cost), Γ is the transport plan

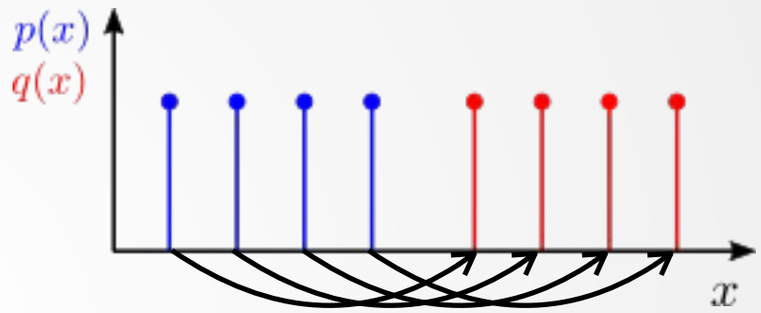
$$W(p,q) = \min\langle \mathbf{D}, \mathbf{\Gamma} \rangle = \min \sum_{ij} \mathbf{D}_{ij} \mathbf{\Gamma}_{ij}$$



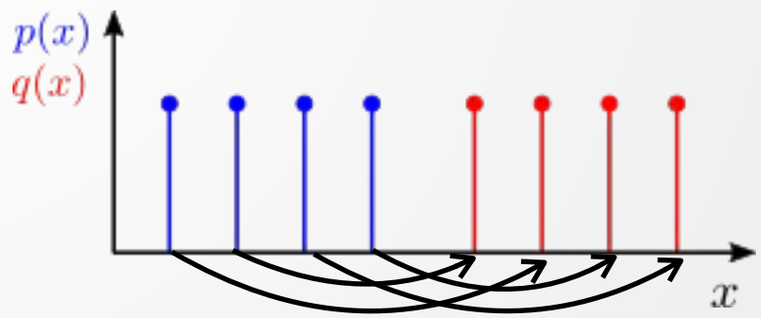
$$\mathbf{D} = \begin{pmatrix} 4 & 5 & 6 & 7 \\ 3 & 4 & 5 & 6 \\ 2 & 3 & 4 & 5 \\ 1 & 2 & 3 & 4 \end{pmatrix}$$



$$\mathbf{\Gamma} = \begin{pmatrix} 0 & 0 & 0 & \frac{1}{4} \\ 0 & 0 & \frac{1}{4} & 0 \\ 0 & \frac{1}{4} & 0 & 0 \\ \frac{1}{4} & 0 & 0 & 0 \end{pmatrix}$$

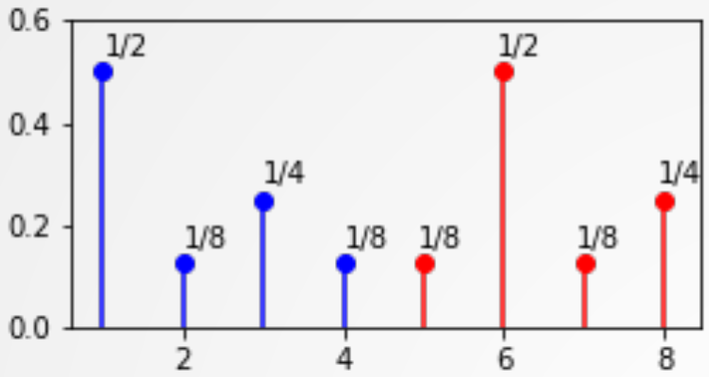


$$\mathbf{\Gamma} = \begin{pmatrix} \frac{1}{4} & 0 & 0 & 0 \\ 0 & \frac{1}{4} & 0 & 0 \\ 0 & 0 & \frac{1}{4} & 0 \\ 0 & 0 & 0 & \frac{1}{4} \end{pmatrix}$$

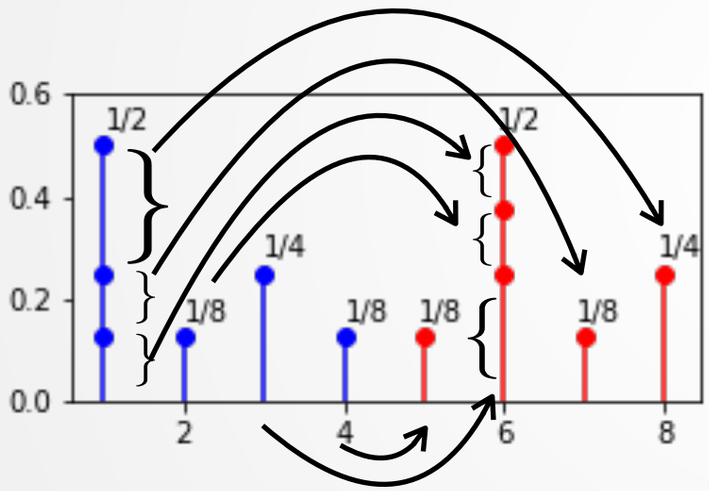


$$\mathbf{\Gamma} = \begin{pmatrix} \frac{1}{4} & 0 & 0 & 0 \\ 0 & \frac{1}{4} & 0 & 0 \\ 0 & 0 & \frac{1}{4} & 0 \\ 0 & 0 & 0 & \frac{1}{4} \end{pmatrix}$$

Wasserstein Distance: Discrete Example

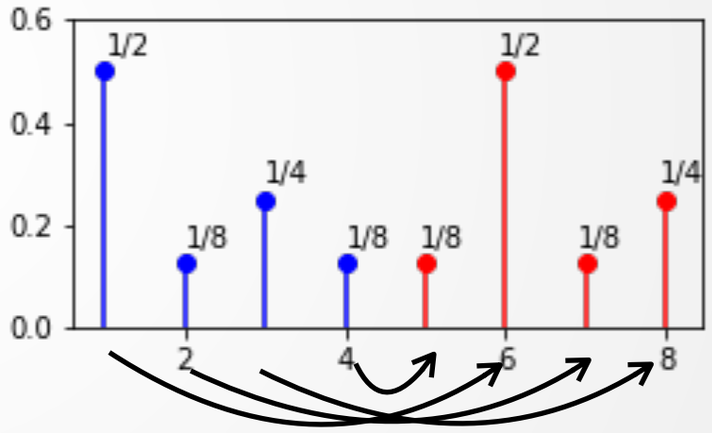


$$\mathbf{D} = \begin{pmatrix} 4 & 5 & 6 & 7 \\ 3 & 4 & 5 & 6 \\ 2 & 3 & 4 & 5 \\ 1 & 2 & 3 & 4 \end{pmatrix}$$



$$\mathbf{\Gamma} = \begin{pmatrix} 0 & 1/8 & 1/8 & 1/4 \\ 0 & 1/8 & 0 & 0 \\ 0 & 1/4 & 0 & 0 \\ 1/8 & 0 & 0 & 0 \end{pmatrix}$$

$$\langle \mathbf{D}, \mathbf{\Gamma} \rangle = \sum_{ij} \mathbf{D}_{ij} \mathbf{\Gamma}_{ij} = 3.125$$



$$\mathbf{\Gamma} = \begin{pmatrix} 0 & 1/2 & 0 & 0 \\ 0 & 0 & 1/8 & 0 \\ 0 & 0 & 0 & 1/4 \\ 1/8 & 0 & 0 & 0 \end{pmatrix}$$

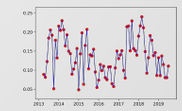
$$\langle \mathbf{D}, \mathbf{\Gamma} \rangle = \sum_{ij} \mathbf{D}_{ij} \mathbf{\Gamma}_{ij} = 4.5$$

Back

Hellinger Distance v.s. Wasserstein Distance

- Less than 10% true policy-related news was selected

True\Pred	LDA_W		LDA_H		Total
	1	0	1	0	
1	52	530	361	221	582
0	519	3485	188	3816	4004
Total	571	4015	549	4073	4586
Accuracy	0.771		0.907		



Coherence Value

- C_{UCI} (*Newman et al. , 2010*): take the set of the top J words (w_1,… , w_J) for a given topic and sum a confirmation measure over all word pairs

$$C_{\mathrm{UCI}} = \frac{2}{J \cdot (J-1)} \sum_{i=1}^{J-1} \sum_{j=i+1}^{J} \log \frac{P\left(w_i, w_j\right) + \epsilon}{P\left(w_i\right) \cdot P\left(w_j\right)}$$

  where the probabilities are estimated on Wikipedia outperform

- C_{UMass} (*Mimno et al. , 2011*): the probabilities estimated from the original corpus

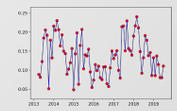
$$C_{\mathrm{UMass}} = \frac{2}{J \cdot (J-1)} \sum_{i=2}^{J} \sum_{j=1}^{i-1} \log \frac{P\left(w_i, w_j\right) + \epsilon}{P\left(w_j\right)}$$

- C_v (*Röder et al., 2015*): confirmation using normalised point-wise mutual information (NPMI)

$$v_{ij} = \mathrm{NPMI}\left(w_i, w_j\right)^{\gamma} = \left(\frac{\log \frac{P(w_i, w_j) + \epsilon}{P(w_i) \cdot P(w_j)}}{-\log\left(P\left(w_i, w_j\right) + \epsilon\right)} \right)^{\gamma}$$

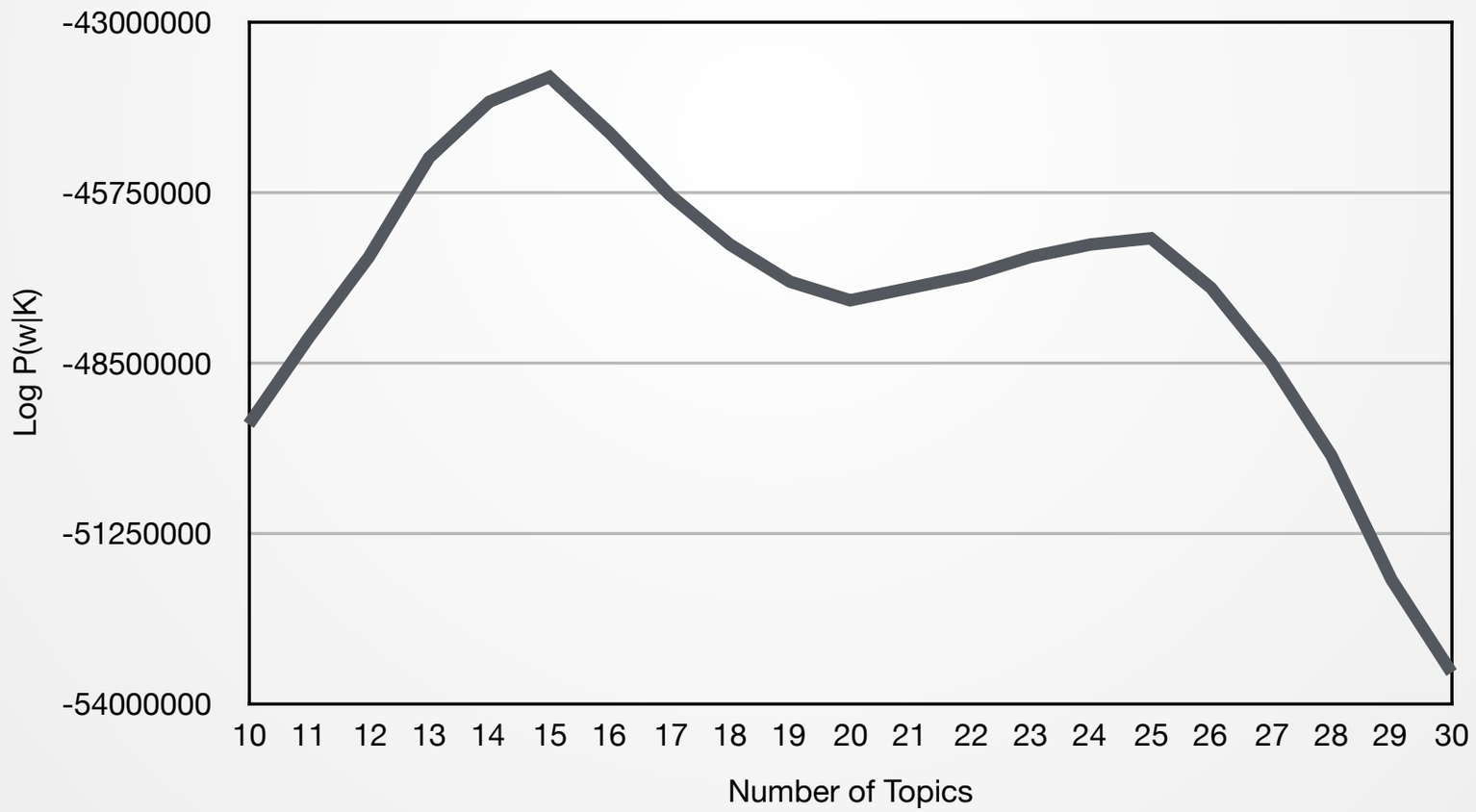
  where γ controls the weight of NPMI

Back



LDA Results: Number of Topics

- Likelihood method: Grifths and Steyvers (2004)
- likelihood of the probability of words for a different number of topics
- K=15



Back