

算力集中如何影 响共识竞争

基于比特币网络的分析

报告人:**徐家画** 伦敦大学学院, exponential science基金会





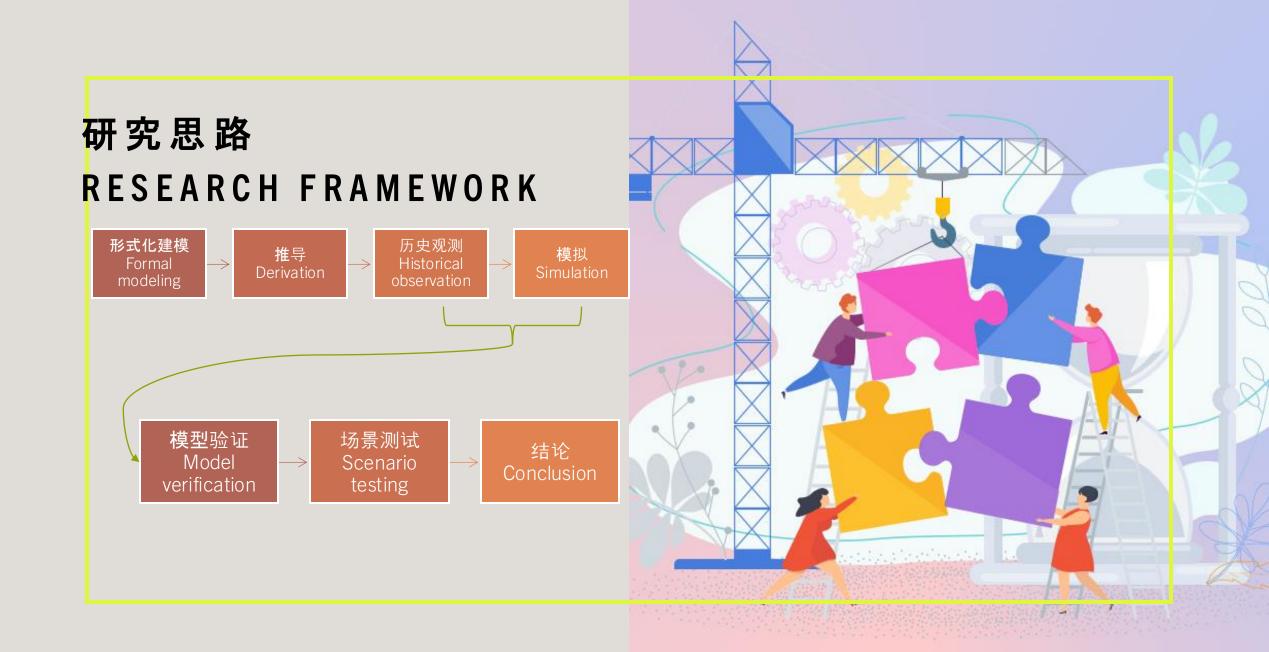
背景知识 BACKGROUND

比特币区块链

- 分布式记账系统
 - 记账节点称为"矿工"
- 工作量证明共识机制
 - 矿工争先计算一个数学难题
 - 计算速度大体和算力("哈希率")成正比
 - 先算好的人可以得到区块链奖励 (一些比特币)
 - 先算好的人要把"挖到"的区块广播给网络中其他矿工

PROOF OF WORK

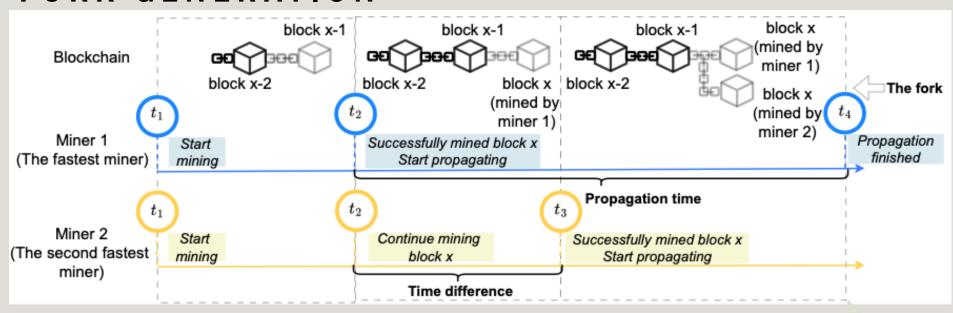




形式化建模 FORMAL MODELING



分叉的形成 FORK GENERATION



- ❖将区块链中挖矿最快的两位标为矿工1和2
- ❖同时动工
- ❖矿工1**挖到一个区**块后立即向整个网络传播该信息
- ❖信息传递有一定时间
- ▶若矿工2在收到信息前还未挖好一个区块,矿工2停止当前区块挖掘,并直接在矿工1挖的区块基础上挖掘下一个区块,没有分叉 形成
- ▶否则,**分叉生成**

模型 简化

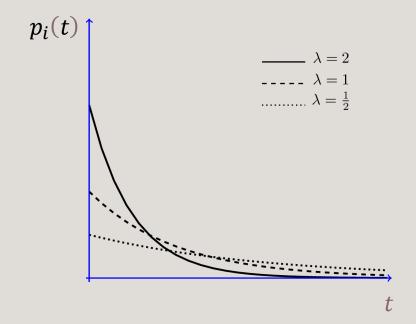
区块挖成—伯努利過程 ARRIVAL OF MINED BLOCK – A BERNOULLI PROCESS

每一个矿工 i对应的算力为 λ_i

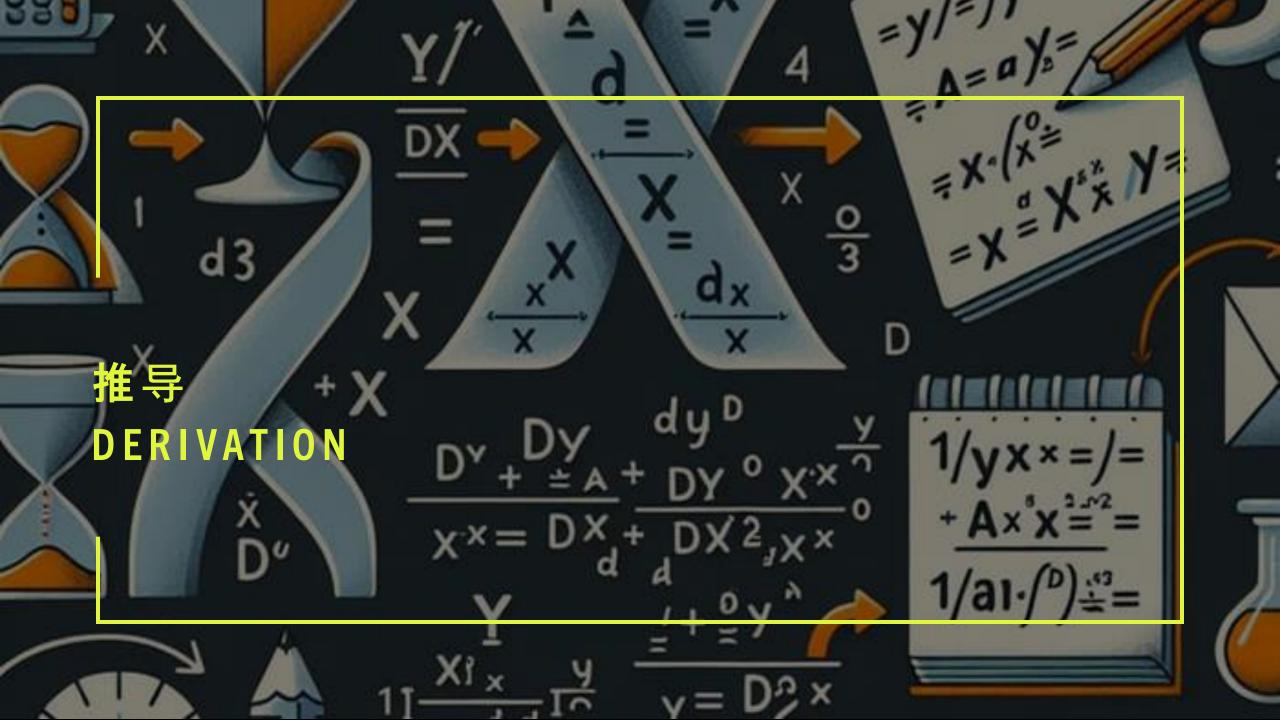
- **❖ 算力越高,区**块挖成用时越短
- **❖ 第一个区**块挖成的时间分布符合**指数分布**, **其概率密度函数及累**积密度函数分别为

$$p_i(t) = rac{\lambda_i}{e^{\lambda_i t}}$$
 $S_i(t) = \int_t^\infty p_i(t) \, dt = rac{1}{e^{\lambda_i t}}$ $S_i(t)$: 矿工 i 未能在时间 t 内挖到下一个区块

区块时间期望为 $\frac{1}{\Sigma\lambda}$ 。



令 $\Lambda := \Sigma \lambda$ 表示比特币区块链网络内总算力,那么,因为区块时间稳定在60秒/min * 10 min = 600秒,所以理论框架下算力总和总是约为1/600。



分叉率计算 DERIVATION OF FORK RATE



给定矿工们的算力分布 $\{\lambda_i\}_i$ 两个最快矿工挖矿时间分别为 t和 t'的概率为:

$$p(t,t'|\{\lambda_i\}) = \sum_{i \neq j} \theta(t'-t)p_i(t)p_j(t') \prod_{k \neq i,j}^N S_k(t')$$

其中,单位阶跃函数 $\theta(t'-t)=egin{cases} 1,\ t'\geq t \ 0,\ t'< t \end{cases}$ 那么两个最快矿工挖矿时间相差正好为 Δ 的概率是:

$$p(\Delta = t' - t | \{\lambda_i\}) = \iint dt dt' \, \delta(\Delta - t' + t) p(t, t' | \{\lambda_i\})$$

分叉生成条件是,第二快矿工在得到第一块矿工的区块前已经挖好自己的区块,即,最快两位矿工挖矿时间差要**小于区**块广播时间。那么给定区块广播时间 Δ_0 ,**分叉率可表示**为:

$$C(\Delta_0) = \int_0^{\Delta_0} dt \int_D p(\Delta|\{\lambda_i\}) p(\{\lambda_i\}) = 1 - \int_D p(\{\lambda_i\}) \frac{\sum_i^N \frac{\lambda_i}{e^{\Delta_0 \sum_{j \neq i} \lambda_j}}}{\sum_i^N \lambda_i}$$

分叉率计算—特殊情况 DERIVATION OF FORK RATE – SPECIAL CONDITIONS

• 给定任意算力联合概率分布 $\{\lambda_i\}\sim D$, 两个最快矿工挖矿时间为 Δ_0 的函数:

$$C(\Delta_0) = \int_0^{\Delta_0} dt \int_D p(\Delta|\{\lambda_i\}) p(\{\lambda_i\}) = 1 - \int_D p(\{\lambda_i\}) \frac{\sum_i^N \frac{\lambda_i}{e^{\Delta_0 \sum_{j \neq i} \lambda_j}}}{\sum_i^N \lambda_i}$$

一般情况下,该表达式运算复杂度数量级为 $O(m^N)$, $m \gg N$ 为一维积分计算量, N为矿工数。

• **当**矿工们的算力两两独立,**即** $\forall i \neq j, \lambda_i \perp \lambda_i$ 时,该表达式可化简为:

$$C(\Delta_0) = 1 - \int_0^\infty \left[\sum_i^N \left(\int_0^\infty \frac{\lambda_i p(\lambda_i) d\lambda_i}{e^{x\lambda_i}} \prod_{j \neq i} \int_0^\infty \frac{p(\lambda_j) d\lambda_j}{e^{(\Delta + x)\lambda_j}} \right) \right] dx$$

此时运算复杂度**数**量级为0(m³)。

• 当矿工们的算力独立同分布(iid)时,表达式可进一步化简为:

$$C(\Delta_0) = 1 - N \int_0^\infty \left[\left(\int_0^\infty \frac{\lambda p(\lambda) d\lambda}{e^{x\lambda}} \right) \left(\int_0^\infty \frac{p(\lambda) d\lambda}{e^{(\Delta_0 + x)\lambda}} \right)^{N-1} \right] dx$$

此时运算复杂度**数**量级降为 $\mathrm{O}(m^2)$ 。

算力分布 HASH RATE DISTRIBUTION

• 半经验分布:以一定时间区间内挖出区块数b_i和该区间内区块总数 B 比率为基准计算出的分布:

$$p(\lambda|b) = \frac{e^{\frac{-(b-\gamma\cdot\lambda)^2}{2\gamma\cdot\lambda}}}{\sqrt{2\pi\lambda/\gamma}},$$
 其中 $\gamma = \frac{B}{\Lambda}$

• 指数分布:简单,只有一个参数。 $\lambda_i^{iid} \sim \text{Exp}(r)$:

$$p(\lambda) = \frac{r}{e^{r\lambda}}$$

该情况下分叉率表达式进一步简化为

$$C(\Delta_0) = 1 - Nr^N \int_0^\infty \frac{dx}{(r+x)^2 (\Delta_0 + r + x)^{N-1}}$$

• 对数正态分布: 常用于有马太效应的社会经济系统中。 $\lambda_i \stackrel{\text{iid}}{\sim} \text{LN}(\mu, \sigma^2)$:

$$p(\lambda) = \frac{e^{\frac{-(\ln \lambda - \mu)^2}{2\sigma^2}}}{\lambda \sigma \sqrt{2\pi}}$$

• 指数截尾的幂律分布:适用于有资源(如能源)限制的系统中。 $\lambda_i^{iid}_{i}$ $\mathrm{TPL}(lpha,eta)$:

$$p(\lambda) = \frac{\beta^{1-\alpha}}{\Gamma(1-\alpha) \cdot \lambda^{\alpha} \cdot e^{\beta\lambda}}$$

在半经验分布, **指数分布**, **及**指数截尾的幂律分布的情况下, 分叉率 $C(\Delta_0)$ 运算复杂度数量级都为 O(m)。



数据

DATA

period	propagation time								empirical miner hash rate					
period of blocks		——————												
start #	start time	50%	90%	99%	block time	In(difficulty)	fork rate	miners	\sum (hash rate)	mean	std	skewness	kurtosis	max share
"		[s]	[s]	[s]	$\frac{1}{\Lambda}$ [s]	(,	[%]	N	$\Lambda [s^{-1}]$	$m [s^{-1}]$	$s [s^{-1}]$			[%]
			.,											
	2015-06-08		16.51	26.61	586.1	24.71	0.620	90	0.00171	0.000019	0.000054	4.01	17.42	19.34
			18.03		546.2	25.24	0.495	86	0.00183	0.000021	\ /	4.63	25.98	22.88
400000	2016-02-25	5.87	15.68	27.24	583.3	25 <mark>.</mark> 93	0.200	135	0.00171	0.000013	0.00 048	5.88	41.04	24.44
	2016-07-09				584.8	26.17	0.240	94	0.00171	0.000018		4.06	18.28	17.80
440000	2016-11-22	3.11	10.49	24.24	566.5	26.64	0.185	63	0.00177	0.000028		2.59	8.22	14.12
	2017-04-02				565.0	27.19	0.185	67	0.00177	0.000026		2.10	3.38	10.53
	2017-08-10			19.98	560.7	27.79	0.125	71	0.00178	0.000025		2.74	6.84	14.16
	2017-12-18			16.25	555.9	28.65	0.085	65	0.00180	0.000028		3.65	14.27	22.42
	2018-04-26			13.33	569.1	29.26	0.035	71	0.00176	0.000025		3.26	10.59	17.90
	2018-09-05			14.77	613.4	29.50	0.025	88	0.00163	0.000019		3.45	11.86	16.75
	2019-01-25			18.10	586.2	29.48	0.030	75	0.00171	0.000023		3.37	11.86	17.21
	2019-06-09			17.83	566.5	29.92	0.045	66	0.00177	0.000027		3.31	11.66	18.39
	2019-10-19			14.41	590.2	30.26	0.030	54	0.00169	0.000031		2.62	7.35	16.89
	2020-03-03			15.78	599.2	30.36	0.050	45	0.00167	0.000037	0.000064	2.58	7.26	18.34
	2020-07-20	The state of the s		17.15	595.1	30.52	0.075	41	0.00168	0.000041		1.87	2.75	16.30
660000	2020-12-05	0.75	4.44	19.22	593.3	30.67	0.095	49	0.00169	0.000034		2.25	4.15	16.81
680000	2021-04-21	0.46	2.32	13.70	616.7	30.52	0.040	40	0.00162	0.000041		1.92	2.48	15.92
	2021-09-11			14.13	579.3	30.71	0.045	35	0.00173	0.000049	0.000084	1.78	1.98	16.16
	2022-01-23	- 1		14.32	591.6	30.98	0.020	33	0.00169	0.000051	0.000090	1.77	2.08	19.38
	2022-06-09			11.30	589.8	31.04	0.0 <mark>05</mark>	26	0.00170	0.000065		1.90	3.10	24.06
	2022-10-23			13.28	590.7	31.25	0.0 <mark>3</mark> 0	37	0.00169	0.000046		3.05	9.62	29.75
	2023-03-09			20.23	590.3	31.52	0.100	26 37 38 33 33	0.00169	0.000045		3.31	11.43	31.38
	2023-07-24	1		21.92	584.6	31.69	0.070	33	0.00171	0.000052		2.91	8.25	28.84
820000	2023-12-06	0.89	4.39	19.81	585.5	31.97	0.065		0.00171	0.000052	0.000111	2.91	8.79	29.19
840000	2024-04-20	0.88	3.86	21.14	598.4	32.07	0.065	3 <mark>3</mark>	0.00167	0.000051	0.000135	3.28	10.46	35.38

参数估计 PARAMETER ESTIMATION

用样本算力平均值 m 和标准差 s 通过矩估(MoM)估计计来三种情况下—— $\lambda_i^{iid} \exp(r)$, $\lambda_i^{iid} \sim \text{LN}(\mu, \sigma^2)$, $\lambda_i^{iid} \sim \text{TPL}(\alpha, \beta)$ ——算力分布参数:

$$r = \frac{1}{m}$$

$$\sigma = \sqrt{\ln\left[1 + \left(\frac{s}{m}\right)^2\right]} \quad \mu = \ln m - \frac{s^2}{2}$$

$$\alpha = 1 - \left(\frac{m}{s}\right)^2 \quad \beta = \frac{m}{s^2}$$

period of blocks

start # start time

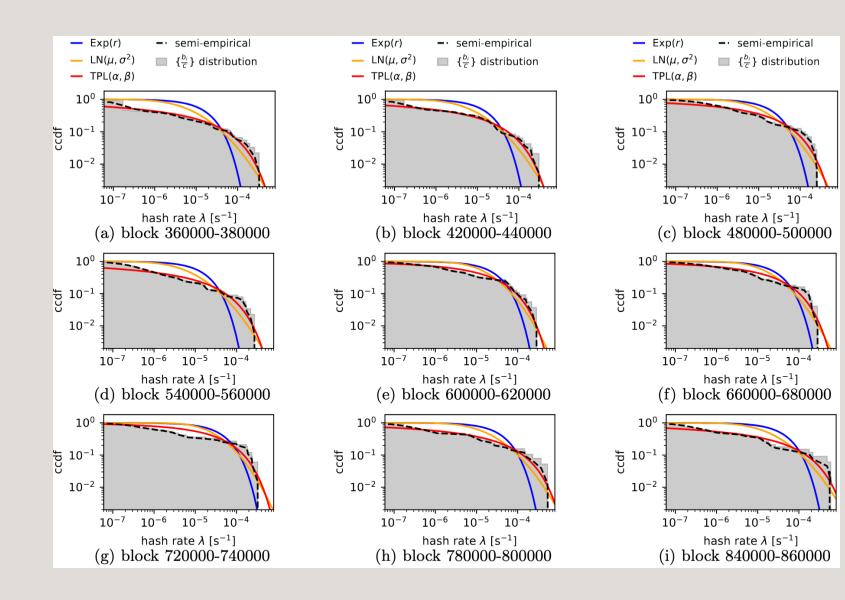
360000 2015-06-08 380000 2015-10-22 400000 2016-02-25 420000 2016-07-09 440000 2016-11-22 460000 2017-04-02 480000 2017-08-10 500000 2017-12-18 520000 2018-04-26 540000 2018-09-05 560000 2019-01-25 580000 2019-06-09 600000 2019-10-19 620000 2020-03-03 640000 2020-07-20 660000 2020-12-05 680000 2021-04-21 700000 2021-09-11 720000 2022-01-23 740000 2022-06-09 760000 2022-10-23 780000 2023-03-09 800000 2023-07-24 820000 2023-12-06 840000 2024-04-20

fitted distributions

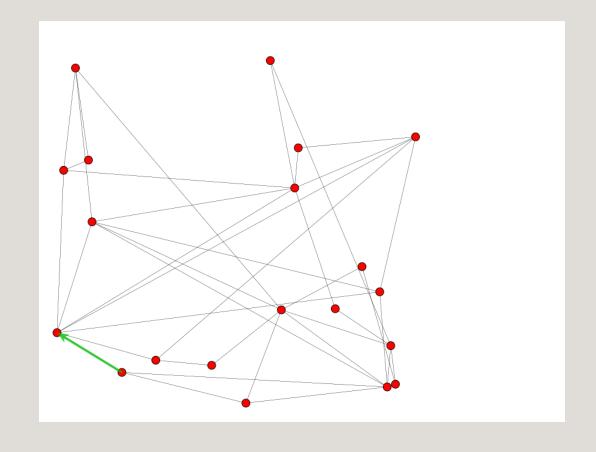
```
\operatorname{Exp}(r) \operatorname{LN}(\mu, \sigma^2)
                      TPL(\alpha, \beta)
52,747 -11.97 1.48 0.88
                            6,556
46,971 - 11.84 1.47 0.87
                            6,080
78,741 - 12.65 \ 1.66 \ 0.93
                            5,407
54,967 - 11.94 1.43 0.85
                            8,072
35,690 - 11.14 1.15 0.63 13,167
37,853 - 11.23 1.18 0.67 12,630
39,812 -11.49 1.34 0.80
                            8,015
36,131 - 11.51 1.43 0.85
                            5,422
40,405 - 11.60 \ 1.41 \ 0.84
                            6,360
53,983 - 11.95 1.45 0.86
                            7,397
43,965 -11.63 1.37 0.82
                            7,992
37,389 - 11.43 1.34 0.80
                            7,359
31,868 - 11.13 1.23 0.72
                            8,920
26,965 - 10.90 \ 1.18 \ 0.67
                            8,934
24,399 - 10.78 1.16 0.65
                            8,486
29,072 - 11.07 \ 1.26 \ 0.74
                            7,441
24,669 - 10.82 \ 1.19 \ 0.68
                            7,853
20,274 - 10.59 1.16 0.65
                            7,043
19,524 - 10.59 \ 1.19 \ 0.68
                            6,296
15,335 - 10.30 \ 1.15 \ 0.64
                            5,520
21,855 - 10.93 \ 1.37 \ 0.82
                            3,932
22,433 - 11.00 1.40 0.84
                            3,648
19,292 -10.78 \ 1.35 \ 0.81
                            3,719
19,323 - 10.73 1.31 0.78
                            4,177
19,747 - 10.94 \ 1.45 \ 0.86
                            2,769
```

历史算力分布 HISTORICAL HASH RATE DISTRIBUTION

- 最强矿工的算力逐年增加
- 除半经验分布外(黑虚线),指数截尾的幂律分布 (红实线)拟合程度最好, 指数分布(蓝实线)拟合程 度最弱。

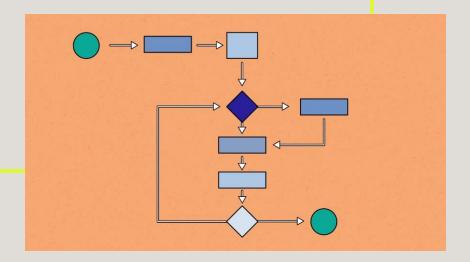


模拟 SIMULATION



环境编译 ENVIRONMENT ENCODING

- 假定算力独立同分布,每一个矿工算力都服从一个特定分布
- 根据该分布随机生成N个数,以表示N个矿工的算力
- 每一个矿工随机生成一个数作为挖矿时间,该时间分布服从指数分布、速率为该矿工算力
- 计算两个最低挖矿时间的差值
- 若差值低于区块广播时间,则有分叉;否则无
- 重复实验足够多次(如, 100万)
- 有分叉数和实验次数比为分叉率

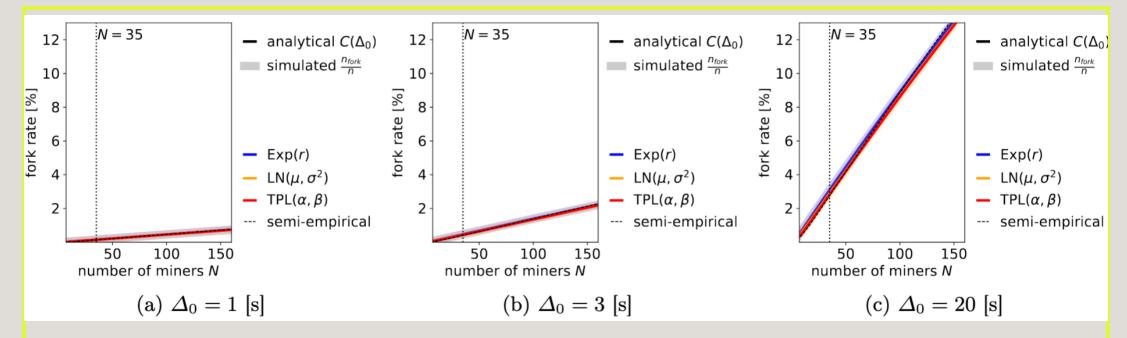


模型验证 MODEL VERIFICATION



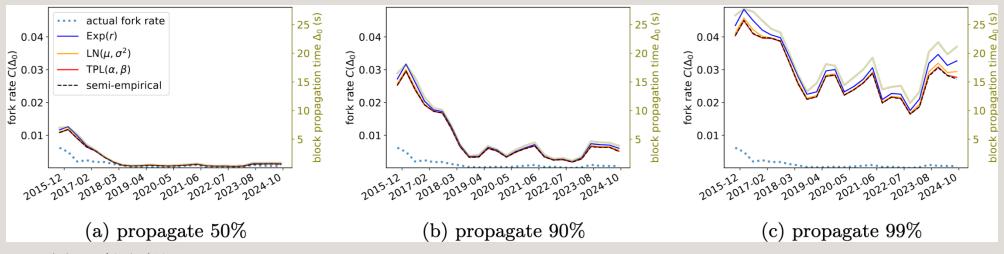
解析解验证

VERIFICATION OF ANALYTICAL SOLUTION

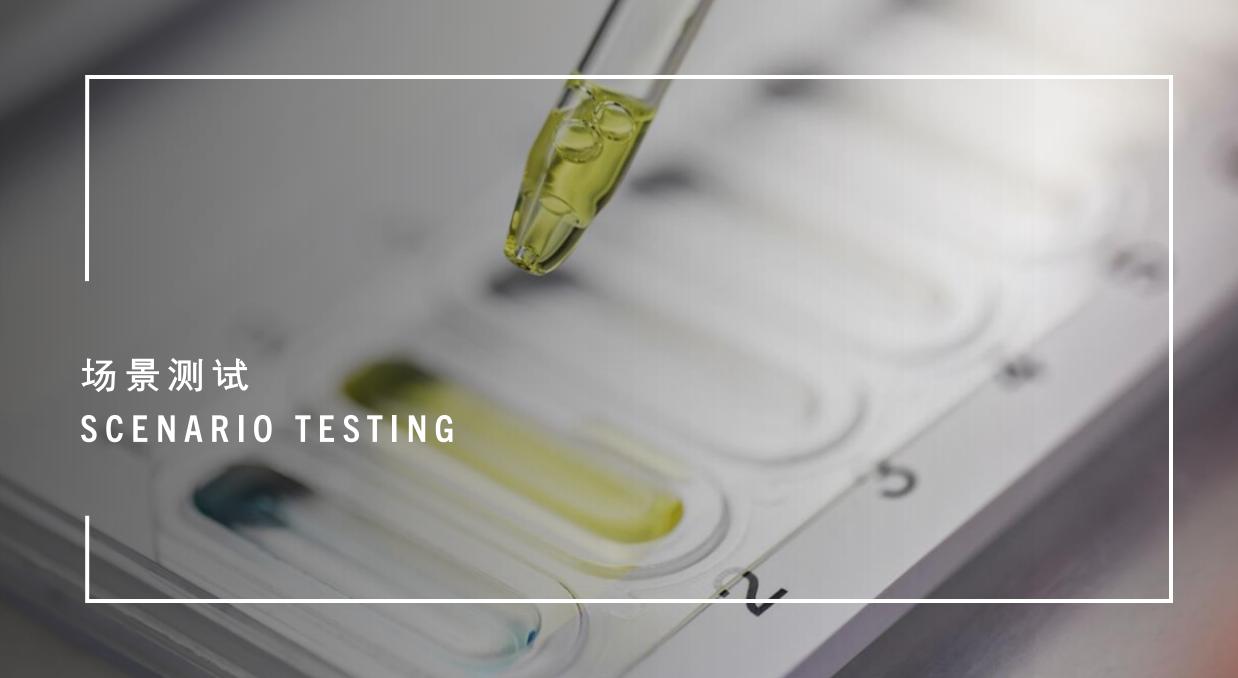


- 用分叉率 $C(\Delta_0)$ 终极解析式运算出的值和对应的模拟环境中运行得到的值高度统一
- 在算力均值m,方差s,矿工人数N,区块广播时间 Δ_0 一定的情况下,独立同分布算力的具体分布形态似乎对分叉率影响不大
- 给定算力均值m, **方差s**, 和区块广播时间 Δ_0 , **分叉率** $C(\Delta_0)$ 随矿工人数 N 增加
- 给定算力均值m,**方差s**,**和** 矿工人数 N,**分叉率** $C(\Delta_0)$ 随区块广播时间 Δ_0 增加

模型估测和历史观测分叉率值比较 COMPARISON BETWEEN MODEL-ESTIMATED AND HISTORICALLY OBSERVED FORK RATES

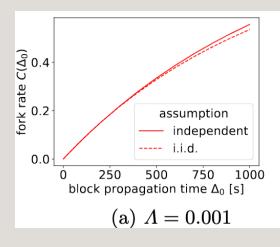


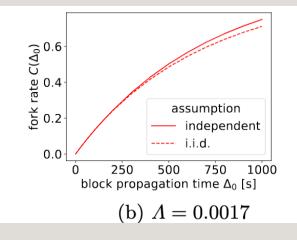
- 历史上分叉率逐年降低
- 趋势上和数值上,用 \mathbf{C} 块广播覆盖50%网络的时间作为最快两位矿工之间区块广播时间 Δ_0 估测出的值和历史观测数值比略高,但最接近
- 符合预期:理论上,最快两位矿工之间传播时间的无条件中值应该接近任意两位矿工之间区块传播时间的中值;但是实际上,最快矿工(通常为大矿池)可能会聚集 在地理政策有利的地段,因此预期他们之间传播时间会比所有传播时间中值稍短
- 带入历史观测的**算力均**值m 和方差s得出的模型估测分叉率几乎和历史观测的区块广播时间 Δ_0 走势重合!(为什么?!)

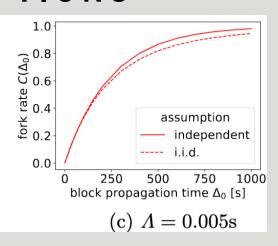


分布假设影响

DIFFERENT DISTRIBUTION ASSUMPTIONS

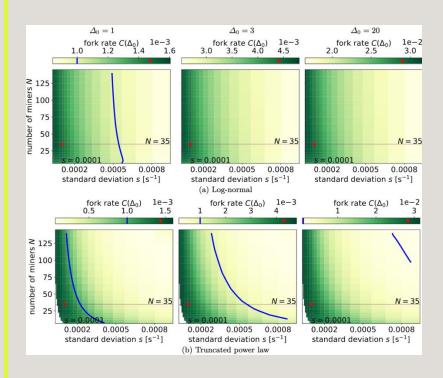






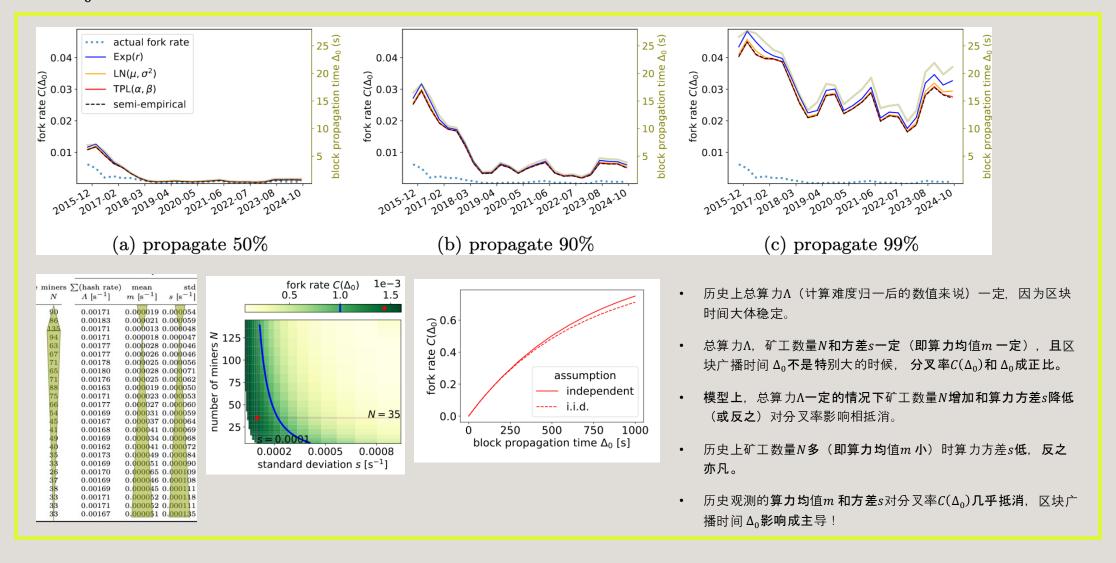
- 用半经验分布为基础, 在其他条件一定的情况下
 - 算力的分布假设是独立同分布还是仅独立分布对分叉率几乎无影响
 - 分叉率 $C(\Delta_0)$ 随区块广播时间 Δ_0 增加
 - 在区块广播时间 Δ_0 较低的区间,增加几乎呈线性
 - 分叉率 $C(\Delta_0)$ 随总算力 Λ 增加

算力异质性影响 FORK RATE HETEROGENEITY



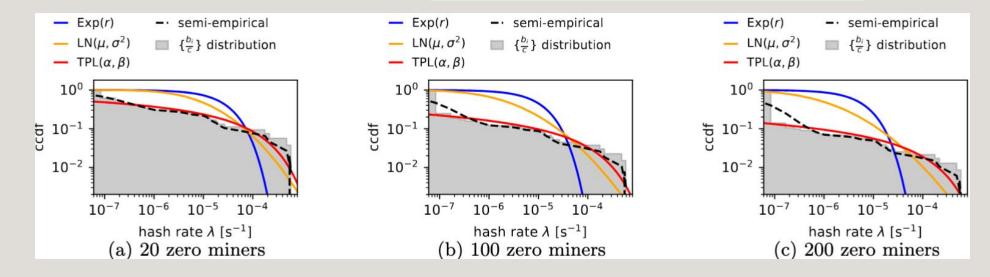
- **特定一种算力分布**(对数正态,或指数截尾的幂律分 布),总算力**Λ一定的情况下**
 - **分叉率** $C(\Delta_0)$ 随区块广播时间 Δ_0 增加(数值轴上数值增大)
 - 算力方差s一定,分叉率 $C(\Delta_0)$ 随矿工人数 N增加(算力均值m 减少)而降低。
 - 矿工人数 N一定,分叉率 $C(\Delta_0)$ 随算力方差s 增加(异质性增加)而降低
 - 在分叉率 $C(\Delta_0)$ 等值线上,高矿工人数 N对应低算力方差s, 反之亦然

带入历史观测的**算力均**值 *m* 和方差 *s* 得出的模型估测分叉率几乎和历史观测的区块广播时间 Δ ₀ 走势重合!(为什么?!)



"零矿工"影响 "ZERO-MINERS"

"零矿工":实际参与到挖矿,但是在 (可能由于算力太低或运气太差)特定 观测区间内没有挖到任何区块。

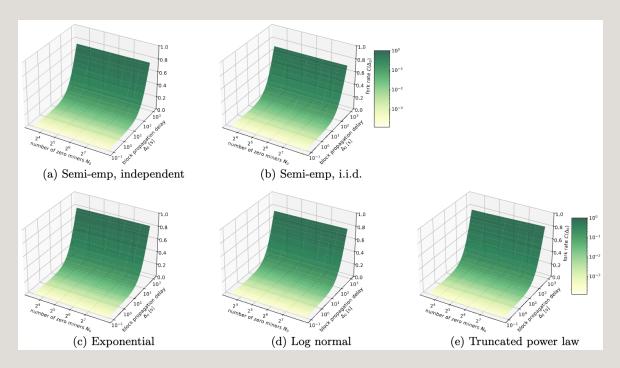


考虑"零矿工"的存在会增加矿工数量N,

降低算力均值m,

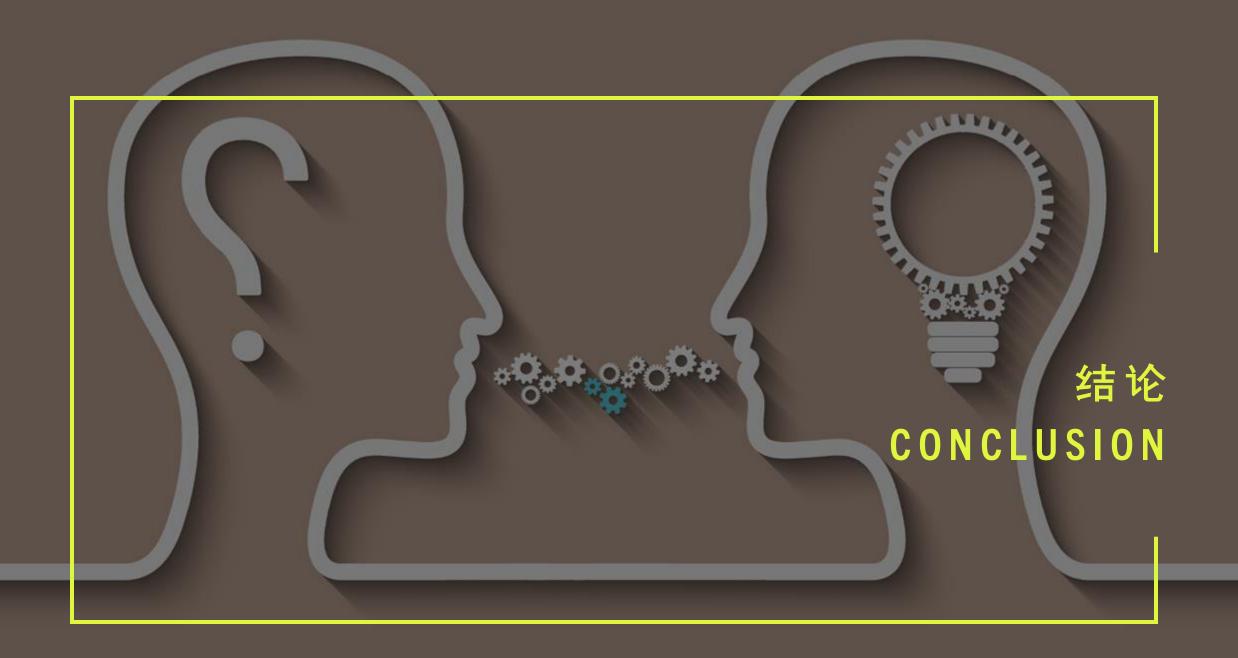
也会影响算力方差s: "零矿工"数量较大时, 算力方差s会降低。

"零矿工"影响 "ZERO-MINERS"



总算力 Λ , 非"零矿工"数量N和方 差s一定时:

- "零矿工"数量 N_0 对分叉率 $C(\Delta_0)$ 的影响微乎其微。
- 算力分布具体形状对分叉率 $C(\Delta_0)$ 的影响微乎其微。
- **分叉率** $C(\Delta_0)$ 随区块广播时间 Δ_0 增加。



- **区**块链网络**分叉率可以通**过解析式**利用**网络拓扑结构(距离,广播速度)和节点矿工算力分布来迅速 预测出,**形式化地**阐释了并量化了分叉率和这些因素的关系:
 - 和总算力,广播速度,算力均值一定下的矿工人数呈正相关,
 - 和总算力一定下的矿工人数, 算力异质性呈负相关。
- 分叉率解析式解释了为什么历史分叉率似乎只和区块广播时间相关
 - 原因:和矿工数量和他们的算力异质性也相关, 但是两者影响相抵消。
- 给出了减少分叉的方案:
 - 增加矿工人数
 - 增加算力异质性(最好算力均匀地分布在可行区间内,而不是大部分人拥有差不多的算力)
 - 提高广播速度
- 在实际分叉率和模型预测值严重相悖时, 可以推断出网络拓扑结构中隐藏信息:
 - 比如:实际分叉率显著低于模型预测值时,可以推断出大矿工之间可能在地理上聚集,或者在挖矿上协同(而不是竞争)。

还有呢? SO WHAT?

- 自由而无用的灵魂
- 物理学家们的自嗨
- 跨学科研究的乐趣







Geoffrey E. Hinton University of Toronto, Canada

"för grundläggande upptäckter och uppfinningar som möjliggör maskininlärning med artificiella neuronnätverk"

for foundational discoveries and inventions that enable machine learning with artificial neural networks" THE NOBEL PRIZE #NobelPrize

联系我 CONTACT

- jiahua.xu@ucl.ac.uk
- https://www.linkedin.com/in/jiahuaxu/
- https://scholar.google.co.uk/citations?u
 ser=20GXpooAAAAJ

