Operation Dumbo Drop: To Airdrop or Not to Airdrop for Initial Coin Offering Success?

Online Appendix

Jian Li,[a] Xiang (Shawn) Wan,[b] Hsing Kenneth Cheng,[c] Xi Zhao [a]

[a] School of Management, Xi'an Jiaotong University, Xi'an, China 710049;

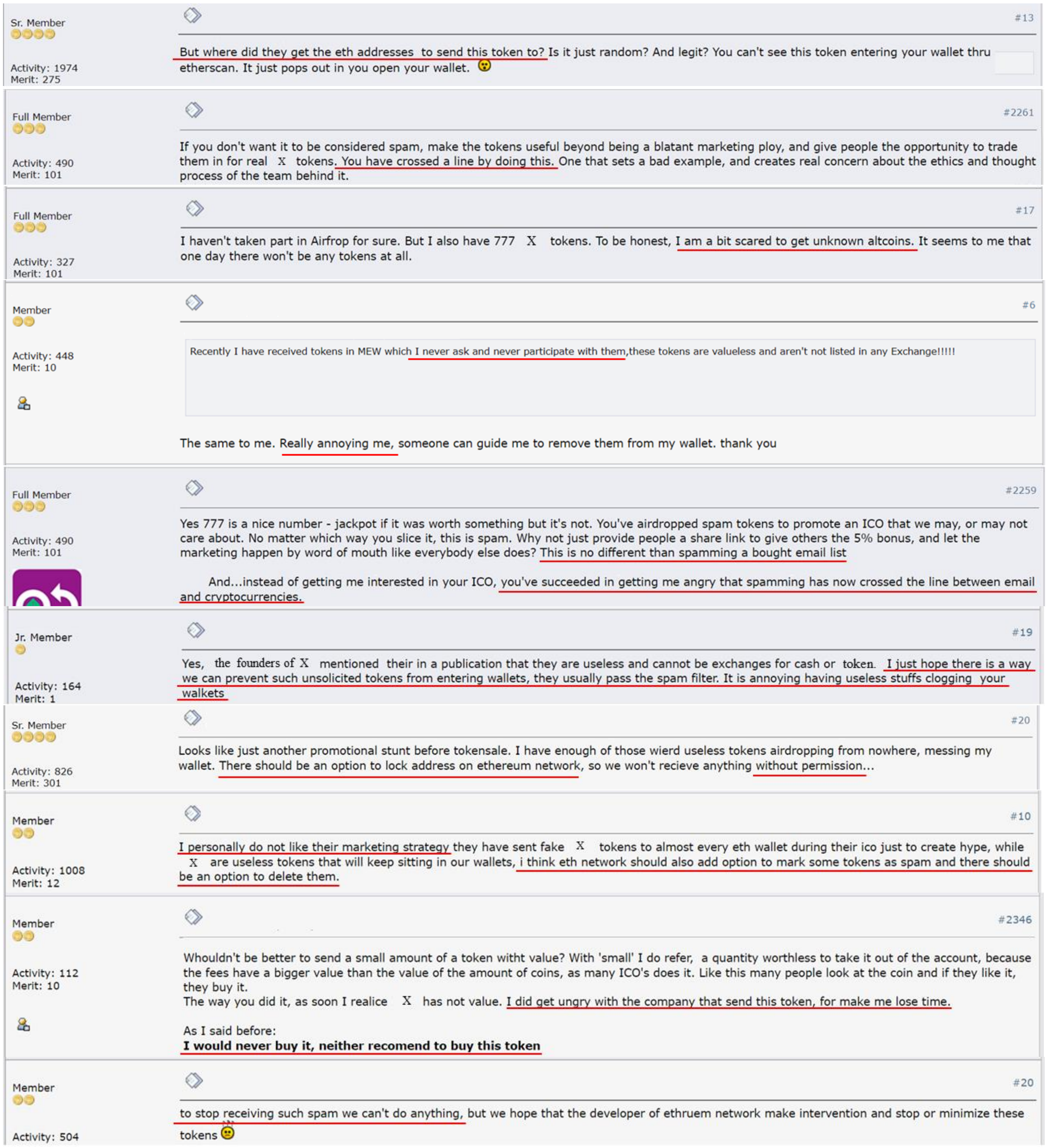
[b] Leavey School of Business, Santa Clara University, Santa Clara, California 95053;

[c] Warrington College of Business, University of Florida, Gainesville, Florida 32611

Contact: lijian_som_phd@stu.xjtu.edu.cn; xwan@scu.edu; hkcheng@ufl.edu; zhaoxi1@mail.xjtu.edu.cn

Appendix A. Complaints about the Promotional Token Airdrop

We show representative discussions around the promotional airdrop of the focal project on the blockchain-related forums in Figure A1.



Sr. Member
Activity: 1974
Merit: 275

#13

But where did they get the eth addresses to send this token to? Is it just random? And legit? You can't see this token entering your wallet thru etherscan. It just pops out in you open your wallet.

Full Member
Activity: 490
Merit: 101

#2261

If you don't want it to be considered spam, make the tokens useful beyond being a blatant marketing ploy, and give people the opportunity to trade them in for real X tokens. You have crossed a line by doing this. One that sets a bad example, and creates real concern about the ethics and thought process of the team behind it.

Full Member
Activity: 327
Merit: 101

#17

I haven't taken part in Airfrop for sure. But I also have 777 X tokens. To be honest, I am a bit scared to get unknown altcoins. It seems to me that one day there won't be any tokens at all.

Member
Activity: 448
Merit: 10

#6

Recently I have received tokens in MEW which I never ask and never participate with them,these tokens are valueless and aren't not listed in any Exchange!!!!!

The same to me. Really annoying me, someone can guide me to remove them from my wallet. thank you

Full Member
Activity: 490
Merit: 101

#2259

Yes 777 is a nice number - jackpot if it was worth something but it's not. You've airdropped spam tokens to promote an ICO that we may, or may not care about. No matter which way you slice it, this is spam. Why not just provide people a share link to give others the 5% bonus, and let the marketing happen by word of mouth like everybody else does? This is no different than spamming a bought email list

And...instead of getting me interested in your ICO, you've succeeded in getting me angry that spamming has now crossed the line between email and cryptocurrencies.

Jr. Member
Activity: 164
Merit: 1

#19

Yes, the founders of X mentioned their in a publication that they are useless and cannot be exchanges for cash or token. I just hope there is a way we can prevent such unsolicited tokens from entering wallets, they usually pass the spam filter. It is annoying having useless stuffs clogging your walkets

Sr. Member
Activity: 826
Merit: 301

#20

Looks like just another promotional stunt before tokensale. I have enough of those wierd useless tokens airdropping from nowhere, messing my wallet. There should be an option to lock address on ethereum network, so we won't recieve anything without permission...

Member
Activity: 1008
Merit: 12

#10

I personally do not like their marketing strategy they have sent fake X tokens to almost every eth wallet during their ico just to create hype, while X are useless tokens that will keep sitting in our wallets, i think eth network should also add option to mark some tokens as spam and there should be an option to delete them.

Member
Activity: 112
Merit: 10

#2346

Whouldn't be better to send a small amount of a token witht value? With 'small' I do refer, a quantity worthless to take it out of the account, because the fees have a bigger value than the value of the amount of coins, as many ICO's does it. Like this many people look at the coin and if they like it, they buy it.
The way you did it, as soon I realice X has not value. I did get ungry with the company that send this token, for make me lose time.

As I said before:
I would never buy it, neither recomend to buy this token

Member
Activity: 504

#20

to stop receiving such spam we can't do anything, but we hope that the developer of ethruem network make intervention and stop or minimize these tokens

Source: Bitcointalk

Figure A1. Discussions around Promotional Airdrops on Blockchain-related Forums

Appendix B. Sample Whitepaper of Blockchain Projects

Whitepapers include a rich set of information about the blockchain project. We show a few examples of the content outline of the whitepapers of blockchain projects in Figure B1.

Contents

a. Streamr

Table of Contents

b. EthBet

TABLE OF CONTENTS

c. Blackmoon

Table of contents

d. AdEx

Figure B1. Contents of Whitepapers

Appendix C. Topic Modeling Method

We utilize the Word Embedding along with the K-means Clustering approach to extract topics from whitepapers in our dataset. Word Embedding is based on neural networks and is widely used to map text words to numerical vectors (Mikolov et al. 2013). Based on the consensus that words in similar contexts should share similar meanings (Li et al. 2021), this approach captures the relationships between words automatically and generate vectors close in vector space for words with similar meanings. For example, ‘protocol’ and ‘algorithm’ should be close in vector space for they both relate to technique, but ‘algorithm’ and ‘market’ should be relatively far. Word embedding effectively reduces the number of dimensions in vector spaces and achieves the differences in meaning by measuring the distances between vectors.

Specifically, we first conduct some basic preprocessing steps (e.g., tokenization, lemmatization) and then leverage the Skip-gram model to obtain each word’s word embedding vector. After acquiring the word-to-vector model, we utilize the K-means clustering algorithm to divide these words into different topics based on their distances in vector space.

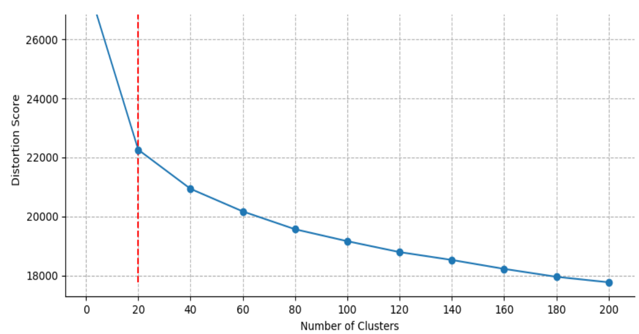


Figure C1. Elbow Method: Variation Trend of Distortion Scores

We apply a data-driven approach, the Elbow method, to locate the optimal number of topics. Given the specific number of clusters, the Elbow method calculates the sum of squared distances between each vector and its assigned centroid. The “Elbow” point denotes the dot where the results would not gain much improvement by adding more clusters. Figure C1 presents the results of the ‘Elbow Method’. The blue solid line represents the variation trend of the distortion score (the sums of squared distances) by increasing the number of clusters. We can observe that the optimal number of topics is 20 (highlighted by the red dashed line). We extract the top 15 most frequent terms of each topic and label each cluster by the most appropriate term. We further group these topics into different categories, including 'sector-related' and 'technology-related'. Table C1 presents the specific categories, labels, and top 15 terms of each topic.

Table C1. Key Terms of Different Topics Derived from Wording Embedding

Category	Topic Label	Most Frequent Terms (Top 15)
Sector-related (6 topics)	Exchange (3.14%)	'use', 'exchange', 'trade', 'asset', 'payment', 'fee', 'currency', 'account', 'digital', 'wallet', 'pay', 'cryptocurrency', 'access', 'investment', 'bitcoin'
	Finance (3.08%)	'include', 'security', 'party', 'management', 'financial', 'operation', 'state', 'legal', 'control', 'third', 'standard', 'activity', 'report', 'relate', 'law'
	Game (3.95%)	'user', 'game', 'content', 'event', 'customer', 'online', 'like', 'provider', 'player', 'social', 'medium', 'website', 'revenue', 'channel', 'advertise'
	Market (7.58%)	'market', 'ha', 'business', 'company', 'work', 'product', 'world', 'wa', 'industry', 'experience', 'global', 'people', 'lead', 'internet', 'currently'
	Platform (3.99%)	'platform', 'ethereum', 'project', 'fund', 'development', 'team', 'offer', 'first', 'follow', 'future', 'available', 'term', 'right', 'part', 'plan'
	DAO (2.35%)	'community', 'share', 'distribute', 'investor', 'reward', 'member', 'vote', 'participant', 'individual', 'group', 'manage', 'pool', 'participate', 'stake', 'structure'
Technology-related (6 topics)	Algorithm (5.55%)	'network', 'data', 'node', 'protocol', 'chain', 'store', 'function', 'storage', 'proof', 'mechanism', 'type', 'code', 'algorithm', 'secure', 'peer'
	Blockchain (6.24%)	'blockchain', 'provide', 'technology', 'new', 'create', 'allow', 'decentralize', 'solution', 'model', 'build', 'develop', 'open', 'ecosystem', 'design', 'well'
	Contract (6.72%)	'contract', 'transaction', 'order', 'require', 'able', 'get', 'send', 'request', 'owner', 'call', 'record', 'must', 'complete', 'agent', 'return'
	Address (2.96%)	'block', 'key', 'address', 'hash', 'signature', 'parameter', 'id', 'let', 'random', 'tree', 'di', 'input', 'map', 'ed', 'output'
	Software (3.66%)	'service', 'smart', 'application', 'support', 'developer', 'app', 'client', 'mobile', 'software', 'feature', 'run', 'implement', 'via', 'device', 'program'
	System (13.72%)	'system', 'base', 'also', 'one', 'process', 'set', 'public', 'example', 'give', 'within', 'two', 'source', 'different', 'every', 'current'
Others (8 topics)	ICO (1.65%)	'ico', 'end', 'start', 'launch', 'day', 'stage', 'release', 'month', 'pre', 'phase', 'date', 'early', 'begin', 'crowdsale', 'week'
	Token (4.86%)	'token', 'sale', 'price', 'amount', 'receive', 'purchase', 'coin', 'issue', 'sell', 'eth', 'limit', 'hold', 'total', 'holder', 'rate'
	Team (6.85%)	'founder', 'university', 'international', 'co', 'advisor', 'finance', 'expert', 'manager', 'engineer', 'ceo', 'director', 'china', 'linkedin', 'degree', 'ltd'
	Disclaimer (2.93%)	'may', 'information', 'risk', 'whitepaper', 'paper', 'doe', 'white', 'condition', 'document', 'purpose', 'person', 'action', 'loss', 'contain', 'subject'
	Roadmap (4.33%)	'time', 'value', 'cost', 'increase', 'high', 'mine', 'power', 'level', 'large', 'grow', 'growth', 'resource', 'demand', 'long', 'expect'
	URL (4.29%)	'br', 'http', 'com', 'www', 'io', 'page', 'org', 'introduction', 'reference', 'pdf', 'net', 'de', 'github', 'index', 'en'
	Verb (7.16%)	'make', 'need', 'result', 'take', 'would', 'case', 'change', 'way', 'trust', 'many', 'could', 'problem', 'without', 'mean', 'possible'
	Numerical (4.99%)	'number', 'year', 'million', 'per', 'bet', 'billion', 'reach', 'accord', 'second', 'next', 'top', 'spend', 'win', 'average', 'last'

Notes. We identified 20 topics from the whitepapers and presented them in Column (2). We listed the 15 most frequent words for each topic in Column (3). We further aggregated these 20 topics into 3 categories and presented them in Column (1).

Appendix D. Definitions of Three Aggregate Similarity Metrics

We leverage the most frequently used similarity metrics between two vectors in our study: Euclidean Distance, Cosine Similarity, and Jaccard Similarity. We describe these three metrics below.

Euclidean Distance measures the straight-line distance of two points in a multi-dimensional space. The larger the number, the higher dissimilarity of the two vectors. For two vectors $A_i = (a_1, a_2, \ldots, a_n)$ and $B_i = (b_1, b_2, \ldots, b_n)$, the Euclidean Distance is defined in formula (D1):

$$\text{Euclidean Distance} = \sqrt{\sum_{i=1}^{n} (a_i - b_i)^2} \qquad \text{(D1)}$$

Cosine Similarity refers to the cosine of the angle between two non-zero vectors in a multi-dimensional space and displays how close the directions that the two vectors point are. The larger the number, the higher similarity of the two vectors. The specific formula is defined in (L2)

$$\text{Cosine Similarity} = \cos(\theta) = \frac{\sum_{i=1}^{n} a_i \times b_i}{\sqrt{\sum_{i=1}^{n} (a_i)^2} \times \sqrt{\sum_{i=1}^{n} (b_i)^2}} \qquad \text{(D2)}$$

Jaccard Similarity, also known as the Jaccard index or Jaccard similarity coefficient, is a statistic to measure which part of two sample groups is consistent and which is distinct. The higher the percentage, the more similar the two groups are. For two *n*-dimensional vectors, the specific formula is defined in (D3):

$$\text{Jaccard Similarity} = \frac{\sum_{i=1}^{n} \min(a, b_i)}{\sum_{i=1}^{n} \max(a_i, b_i)} \qquad \text{(D3)}$$

In our research, the whitepaper texts of the focal project or any historically interacted project can be treated as a 20-dimension topic vector. Hence, we employ the aforementioned three measures to gauge the project similarity between an individual's historically interacted projects and the focal project. To ensure that a larger number represents higher similarity for all three measures, we employ the *inverse of Euclidean Distance* in the main analysis and *Cosine Similarity* and *Jaccard Similarity* in robustness checks.

Appendix E. Falsification Tests on Placebo Cutoffs

We present the estimated results based on the placebo cutoffs of 0.05, 0.2, and 0.5 Ether around the true cutoff of 0.1 Ether in Table E1. We find an insignificant effect of token airdrop around these placebo cutoffs. These results strengthen the causal claim around the cutoff of 0.1 Ether.

Table E1. Estimations with Placebo Cutoffs

	Prob.			log (Amt. + 1)		
	(1)	(2)	(3)	(4)	(5)	(6)
I(Balance>C)	0.0007	0.0011	-0.0004	0.0039	0.0078	-0.0017
	(0.0004)	(0.0008)	(0.0014)	(0.0029)	(0.0053)	(0.0094)
Balance-C	-0.9128***	-0.6866**	-0.7649**	-6.2199***	-4.8354**	-5.4160**
	(0.1117)	(0.2338)	(0.2600)	(0.7476)	(1.5952)	(1.7970)
I(Balance>C) × (Balance-C)	1.5297***	0.6379	1.3806**	10.1780***	4.6409	9.2482*
	(0.1867)	(0.3558)	(0.5287)	(1.2493)	(2.4272)	(3.6537)
Placebo cutoff	**0.05**	**0.20**	**0.50**	**0.05**	**0.20**	**0.50**
Bandwidth	0.008	0.008	0.008	0.008	0.008	0.008
Kernel function	Triangular	Triangular	Triangular	Triangular	Triangular	Triangular
Eff. observations	85,385	29,221	17,529	85,385	29,221	17,529
Adj R^2	0.0010	0.0002	0.0004	0.0010	0.0002	0.0004

*Notes. *p<0.05; **p<0.01; ***p<0.001. Standard errors in parentheses.*

Appendix F. Estimations with Different Bandwidths and Kernel Functions

We report the estimated results based on the bandwidths of 0.004 (i.e., shrunk to half of the optimal bandwidth) and 0.016 (i.e., broadened to double the optimal bandwidth), as well as the uniform kernel in Table F1. Our findings are robust with the selection of different bandwidths and kernel functions.

Table F1. Estimations with Different Bandwidths and Kernel Functions

Panel A: Investment Probability as the DV

	Shrunken Bandwidth		Broadened Bandwidth		Uniform Kernel	
	(1)	(2)	(3)	(4)	(5)	(6)
I(Balance>0.1)	0.0035^{***}	0.0065^{***}	0.0025^{***}	0.0042^{***}	0.0026^{***}	0.0047^{***}
	(0.0005)	(0.0008)	(0.0004)	(0.0006)	(0.0005)	(0.0008)
I(Balance>0.1) × Project similarity		-0.0061^{***}		-0.0047^{***}		-0.0047^{***}
		(0.0007)		(0.0005)		(0.0006)
Project similarity		-0.0018^{***}		-0.0021^{***}		-0.0021^{***}
		(0.0002)		(0.0002)		(0.0003)
Balance-0.1	-1.0566^{***}	-1.3200^{***}	-0.3599^{***}	-0.4249^{***}	-0.4803^{***}	-0.4065^{**}
	(0.2018)	(0.3415)	(0.0455)	(0.0847)	(0.0722)	(0.1323)
I(Balance>0.1) × (Balance-0.1)	0.6489	0.2127	0.3988^{***}	0.3317^{*}	0.4781^{***}	0.0674
	(0.4457)	(0.6793)	(0.0844)	(0.1342)	(0.1391)	(0.2211)
Bandwidth	**0.004**	**0.004**	**0.016**	**0.016**	**0.008**	**0.008**
Kernel function	**Triangular**	**Triangular**	**Triangular**	**Triangular**	**Rectangular**	**Rectangular**
Eff. observations	99,289	47,311	147,445	76,806	117,630	58,837
Adj R^2	0.0008	0.0060	0.0008	0.0052	0.0007	0.0050

*Notes. *p<0.05; **p<0.01; ***p<0.001. Standard errors in parentheses.*

Panel B: log (Investment Amount ($) + 1) as the DV

	Shrunken Bandwidth		Broadened Bandwidth		Uniform Kernel	
	(1)	(2)	(3)	(4)	(5)	(6)
I(Balance>0.1)	0.0237^{***}	0.0429^{***}	0.0165^{***}	0.0275^{***}	0.0174^{***}	0.0306^{***}
	(0.0037)	(0.0060)	(0.0029)	(0.0045)	(0.0035)	(0.0056)
I(Balance>0.1) × Project similarity		-0.0400^{***}		-0.0304^{***}		-0.0308^{***}
		(0.0047)		(0.0035)		(0.0040)
Project similarity		-0.0127^{***}		-0.0147^{***}		-0.0145^{***}
		(0.0016)		(0.0016)		(0.0018)
Balance-0.1	-7.1671^{***}	-8.8453^{***}	-2.4580^{***}	-2.8038^{***}	-3.3349^{***}	-2.7022^{**}
	(1.4144)	(2.4357)	(0.3169)	(0.5955)	(0.5061)	(0.9362)
I(Balance>0.1) × (Balance-0.1)	4.0738	1.0318	2.6735^{***}	2.1510^{*}	3.3969^{***}	0.5655
	(3.1232)	(4.8456)	(0.5887)	(0.9442)	(0.9749)	(1.5649)
Bandwidth	**0.004**	**0.004**	**0.016**	**0.016**	**0.008**	**0.008**
Kernel function	**Triangular**	**Triangular**	**Triangular**	**Triangular**	**Rectangular**	**Rectangular**
Eff. observations	99,289	47,311	147,445	76,806	117,630	58,837
Adj R^2	0.0007	0.0054	0.0007	0.0047	0.0007	0.0046

*Notes. *p<0.05; **p<0.01; ***p<0.001. Standard errors in parentheses.*

Appendix G. Estimations with High-Order Polynomials

We estimate the data with alternative specifications by varying the highest polynomial order of the term (Balance-0.1) from 1 (i.e., linear) to 2 and 3 and present the estimations in Table G1. We find qualitatively similar results across different specifications, indicating that pseudo-discontinuity arising from over-fitting is not a concern in our context.

Table G1. Estimations with High-Order Polynomials

	Prob.				log (Amt. + 1)			
	Polynomial order 2		Polynomial order 3		Polynomial order 2		Polynomial order 3	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
I(Balance>0.1)	**0.0036*****	0.0063***	**0.0035*****	0.0065***	**0.0243*****	0.0418***	**0.0240*****	0.0441***
	(0.0006)	(0.0009)	**(0.0007)**	(0.0011)	**(0.0041)**	(0.0065)	**(0.0050)**	(0.0077)
I(Balance>0.1) × Project similarity		**-0.0055*****		**-0.0055*****		**-0.0359*****		**-0.0359*****
		(0.0006)		**(0.0006)**		**(0.0041)**		**(0.0041)**
Project similarity		-0.0019***		-0.0019***		-0.0137***		-0.0137***
		(0.0002)		(0.0002)		(0.0016)		(0.0016)
Balance-0.1	-1.3765***	-1.7190***	-0.9586*	-2.1267**	-9.4163***	-11.7309***	-6.6933*	-14.5564*
	(0.2731)	(0.4584)	(0.4775)	(0.8151)	(1.9134)	(3.2580)	(3.3454)	(5.7929)
I(Balance>0.1) × (Balance-0.1)	0.8439	0.5340	0.5525	0.6296	5.5512	3.4362	1.9390	1.7489
	(0.5938)	(0.9006)	(1.2231)	(1.8582)	(4.1606)	(6.4007)	(8.5695)	(13.2060)
(Balance-0.1)^2	-1.6e+02**	-2.4e+02**	72.2951	-4.7e+02	-1.1e+03**	-1.7e+03**	420.6630	-3.2e+03
	(53.1028)	(88.9099)	(224.1740)	(386.3179)	(372.0661)	(631.8755)	(1.6e+03)	(2.7e+03)
I(Balance>0.1) × (Balance-0.1)^2	223.9698*	344.2369*	-62.0971	703.1040	1.6e+03*	2.4e+03*	448.7594	5.9e+03
	(102.3214)	(155.5693)	(486.1543)	(744.1044)	(716.9176)	(1.1e+03)	(3.4e+03)	(5.3e+03)
(Balance-0.1)^3			2.8e+04	-2.7e+04			1.8e+05	-1.9e+05
			(2.6e+04)	(4.5e+04)			(1.8e+05)	(3.2e+05)
I(Balance>0.1) × (Balance-0.1)^3			-2.3e+04	1.4e+04			-2.2e+05	-8.3e+03
			(5.1e+04)	(7.9e+04)			(3.6e+05)	(5.6e+05)
Bandwidth	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.008
Kernel function	Triangular	Triangular	Triangular	Triangular	Triangular	Triangular	Triangular	Triangular
Eff. observations	117,630	58,837	117,630	58,837	117,630	58,837	117,630	58,837
Adj R^2	0.0008	0.0057	0.0008	0.0056	0.0007	0.0051	0.0007	0.0051

*Notes. *p<0.05; **p<0.01; ***p<0.001. Standard errors in parentheses.*

Appendix H. Estimations Controlling for Investor Characteristics

We conduct additional analysis by including investor characteristics as control variables and report the estimations in Table H1. We find qualitatively similar results, indicating the robustness of our findings.

Table H1. Estimations with Control Variables

	Prob.		log (Amt. + 1)	
	(1)	(2)	(3)	(4)
I(Balance>0.1)	**0.0024***** **(0.0005)**	**0.0050***** **(0.0007)**	**0.0157***** **(0.0033)**	**0.0318***** **(0.0052)**
Balance-0.1	-0.5663*** (0.1027)	-0.5751** (0.1811)	-3.8771*** (0.7196)	-3.8923** (1.2870)
I(Balance>0.1) × (Balance-0.1)	0.3711 (0.1947)	-0.0017 (0.3034)	2.6341 (1.3639)	0.2021 (2.1562)
The moderating role of project similarity				
I(Balance>0.1) × Project similarity		**-0.0050***** **(0.0006)**		**-0.0324***** **(0.0041)**
Project similarity		-0.0013*** (0.0002)		-0.0088*** (0.0017)
Other investor characteristics				
Account age (day)	-0.0004** (0.0001)	-0.0007* (0.0003)	-0.0016 (0.0010)	-0.0038 (0.0020)
Amount of total ether flows (ether)	0.0125*** (0.0011)	0.0171*** (0.0021)	0.0875*** (0.0080)	0.1210*** (0.0148)
Amount of ether outflows (ether)	-0.0066*** (0.0009)	-0.0096*** (0.0017)	-0.0469*** (0.0065)	-0.0729*** (0.0120)
Amount of ether inflows (ether)	-0.0027*** (0.0006)	-0.0024* (0.0010)	-0.0165*** (0.0044)	-0.0120 (0.0074)
Maximum transaction value (ether)	-0.0001 (0.0012)	-0.0048* (0.0019)	-0.0017 (0.0086)	-0.0330* (0.0138)
Average value per transaction (ether)	-0.0062*** (0.0008)	-0.0052*** (0.0013)	-0.0426*** (0.0054)	-0.0361*** (0.0091)
Transaction fees spent (ether)	0.0246 (0.0166)	-0.0000 (0.0233)	0.1600 (0.1165)	-0.0130 (0.1653)
Maximum historical balance (ether)	-0.0014** (0.0005)	-0.0017 (0.0009)	-0.0100** (0.0034)	-0.0115 (0.0063)
Average daily historical balance (ether)	0.0022* (0.0009)	0.0048** (0.0016)	0.0186** (0.0065)	0.0384*** (0.0112)
Bandwidth	0.008	0.008	0.008	0.008
Kernel function	Triangular	Triangular	Triangular	Triangular
Eff. observations	117,630	58,837	117,630	58,837
Adj. R^2	0.0043	0.0087	0.0044	0.0084

*Notes. *p<0.05; **p<0.01; ***p<0.001. Standard errors in parentheses.*

Appendix I. Mapping Multiple Addresses to the Same User

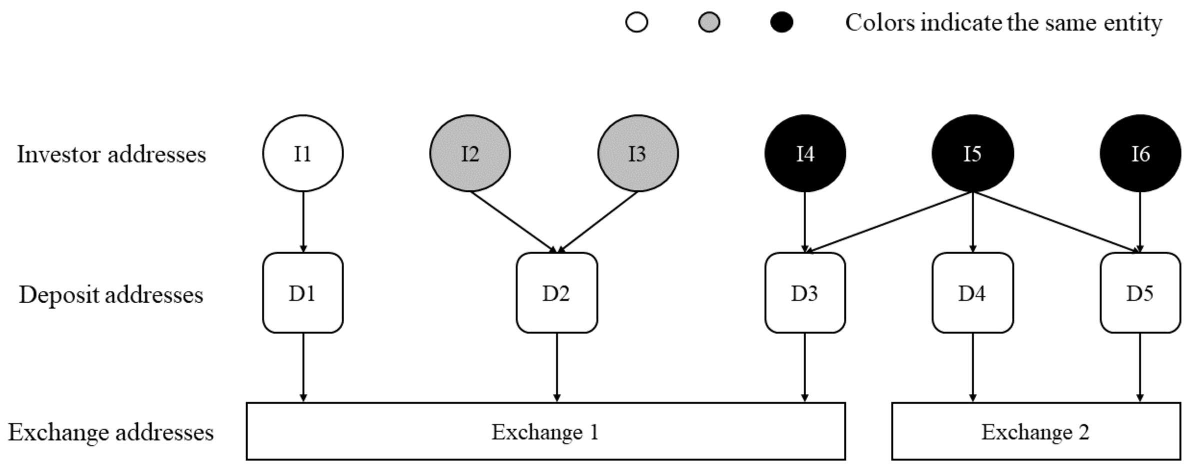
We leverage the newly developed method of mapping multiple addresses to the same user (Victor 2020) to solve the multiple address issues. Then, we conduct analyses based on two newly constructed samples: (i) merging multiple addresses into a single one; (ii) dropping addresses with multiple address concerns. We find consistent results, indicating the robustness of our findings.

I1. Methods of Mapping Multiple Addresses to the Same User

Existing studies aim to solve the multiple addresses issues for two types of blockchains: the blockchain with an unspent transaction output model (e.g., Bitcoin) (Foley et al. 2019, Yin et al. 2019) and the blockchain with an account model (e.g., Ethereum) (Victor 2020). Specifically, Victor (2020) proposed the first clustering heuristics for Ethereum's account model.

The logic for mapping multiple addresses to the same user in the work of (Victor 2020) is illustrated below. This clustering approach is based on a well-known Ethereum phenomenon: the reuse of exchange deposit addresses. Any Ethereum user who wants to sell their Ether tokens must transfer them to crypto exchanges. To credit the assets to the correct account, crypto exchanges typically create "deposit addresses" to receive asset transfers from blockchain individuals and pass them to the main account of the exchange. These "deposit addresses" are created per exchange customer. Thus, several addresses sending Ether to the same deposit address most likely belong to the same user. Such an assumption can be used to cluster addresses.

Figure I1 illustrates the logic of this method briefly. Figure H1 presents three types of addresses: investor addresses (I1-I6), deposit addresses (D1-D5), and exchange addresses (Exchange1-Exchange2). Investor addresses represent the Ethereum addresses of potential investors in our analyses. Potential investors need to transfer their crypto assets to "exchange addresses" via the "deposit addresses" to sell out. For example, both investor addresses "I2" and "I3" use deposit address "D2" for assets exchange and thus may belong to the same entity.



Notes. This figure is modified based on Figure 1 of Victor (2020). I2-I6 denotes investor addresses. D1-D5 denotes the deposit addresses. Exchange 1 and 2 denote the exchange addresses.

Figure I1. Logics to Cluster Multiple Addresses by Deposit Address

I2. Analyses based on Two Newly Constructed Samples

Following the work of (Victor 2020), we implement a clustering algorithm to map multiple addresses belonging to the same user together. The results reveal that out of the total 9,295,371 regular investor addresses (already eliminating deposit addresses) in our full sample, 488,647 addresses involve in multiple addresses issue. Within the optimal bandwidth, 2,928 (2.5 percent) out of 117,186 addresses may have multiple address issues. These 2,928 addresses belong to 2,735 unique blockchain entities/users, and those unique entities/users also have 9,376 addresses out of the optimal bandwidth. Accordingly, we conduct analyses based on the following two newly constructed samples:

- In the first sample, we aggregate the multiple addresses from the same user together and then reconstruct the variables for them. Such analysis could help us map the transaction history of several addresses to form a precise understanding of the investors' transaction behavior.
- In the second sample, we exclude the addresses with multiple address issues.

We report the results for the above two samples in Table I1 and Table I2, respectively. We find qualitatively similar results, which indicate the robustness of our findings. Thus, the multiple addresses issue may be less of a concern in our context.

Table I1. Estimation after Aggregating Addresses with Multiple Addresses Issues

	Prob.		log (Amt. + 1)	
	(1)	(2)	(3)	(4)
I(Balance>0.1)	**0.0031***** **(0.0005)**	0.0055*** (0.0008)	**0.0204***** **(0.0033)**	0.0357*** (0.0053)
I(Balance>0.1) × Project similarity		**-0.0055***** **(0.0006)**		**-0.0359***** **(0.0042)**
Project similarity		-0.0021*** (0.0002)		-0.0148*** (0.0016)
Balance-0.1	-0.6199*** (0.1043)	-0.5660** (0.1854)	-4.3102*** (0.7289)	-3.9158** (1.3153)
I(Balance>0.1) × (Balance-0.1)	0.4605* (0.1991)	-0.0288 (0.3137)	3.3628* (1.3918)	0.1894 (2.2250)
Bandwidth	0.008	0.008	0.008	0.008
Kernel function	Triangular	Triangular	Triangular	Triangular
Eff. observations	115,824	57,594	115,824	57,594
Adj. R^2	0.0007	0.0058	0.0007	0.0052

*Notes. *p<0.05; **p<0.01; ***p<0.001. Standard errors in parentheses.*

Table I2. Estimation after Excluding Addresses with Multiple Addresses Issues

	Prob.		log (Amt. + 1)	
	(1)	(2)	(3)	(4)
I(Balance>0.1)	**0.0029***** **(0.0005)**	0.0056*** (0.0007)	**0.0193***** **(0.0032)**	0.0359*** (0.0052)
I(Balance>0.1) × Project similarity		**-0.0056***** **(0.0006)**		**-0.0360***** **(0.0041)**
Project similarity		-0.0020*** (0.0002)		-0.0137*** (0.0015)
Balance-0.1	-0.5599*** (0.1012)	-0.5094** (0.1800)	-3.9157*** (0.7058)	-3.5673** (1.2750)
I(Balance>0.1) × (Balance-0.1)	0.3930* (0.1937)	-0.0945 (0.3045)	2.9296* (1.3502)	-0.1881 (2.1569)
Bandwidth	0.008	0.008	0.008	0.008
Kernel function	Triangular	Triangular	Triangular	Triangular
Eff. observations	114,258	56,210	114,258	56,210
Adj. R^2	0.0006	0.0058	0.0006	0.0052

*Notes. *p<0.05; **p<0.01; ***p<0.001. Standard errors in parentheses.*

Appendix J. Alternative Weighting Methods

We alternatively construct project similarity by weighting the projects with the amount ($) invested, except for the number of transactions in our main analysis. Table J1 presents the results. We find qualitatively similar results, indicating the robustness of our findings.

Table J1. Moderating Effect of Project Similarity Weighted by Investment Amount

	Prob.	log (Amt. + 1)
	(1)	(2)
I(Balance>0.1)	0.0055***	0.0356***
	(0.0007)	(0.0052)
I(Balance>0.1) × Project similarity	**-0.0039*****	**-0.0252*****
	(0.0005)	**(0.0037)**
Project similarity	-0.0020***	-0.0142***
	(0.0002)	(0.0017)
Balance-0.1	-0.5885***	-3.9078**
	(0.1788)	(1.2706)
I(Balance>0.1) × (Balance-0.1)	-0.0234	-0.0138
	(0.3027)	(2.1515)
Bandwidth	0.008	0.008
Kernel function	Triangular	Triangular
Eff. observations	58,837	58,837
Adj. R^2	0.0050	0.0045

*Notes. *p<0.05; **p<0.01; ***p<0.001. Standard errors in parentheses.*

Appendix K. Alternative Similarity Metrics

We use two alternative measurements of the project similarity: Cosine similarity and Jaccard similarity, besides the inverse of Euclidean distance in the main analysis. We report the estimations in Table K1. We find qualitatively similar results, indicating the robustness of our findings.

Table K1. Estimations with Alternative Similarity Metrics

Panel A: Investment Probability as the DV

	Similarity measure 1: Cosine similarity	Similarity measure 2: Jaccard similarity
I(Balance>0.1)	0.0060***	0.0060***
	(0.0007)	(0.0007)
I(Balance>0.1) × Project similarity	**-0.0063*** **	**-0.0061*** **
	(0.0006)	**(0.0006)**
Project similarity	-0.0020***	-0.0019***
	(0.0002)	(0.0002)
Balance-0.1	-0.6686***	-0.6474***
	(0.1760)	(0.1767)
I(Balance>0.1) × (Balance-0.1)	0.0583	0.0354
	(0.3006)	(0.3011)
Bandwidth	0.008	0.008
Kernel function	Triangular	Triangular
Eff. observations	58,837	58,837
Adj R^2	0.0063	0.0060

*Notes. *p<0.05; **p<0.01; ***p<0.001. Standard errors in parentheses.*

Panel B: log (Investment Amount ($) + 1) as the DV

	Similarity measure 1: Cosine similarity	Similarity measure 2: Jaccard similarity
I(Balance>0.1)	0.0391***	0.0388***
	(0.0052)	(0.0052)
I(Balance>0.1) × Project similarity	**-0.0408*** **	**-0.0397*** **
	(0.0042)	**(0.0042)**
Project similarity	-0.0140***	-0.0135***
	(0.0015)	(0.0015)
Balance-0.1	-4.5055***	-4.3506***
	(1.2507)	(1.2559)
I(Balance>0.1) × (Balance-0.1)	0.6195	0.4367
	(2.1365)	(2.1400)
Bandwidth	0.008	0.008
Kernel function	Triangular	Triangular
Eff. observations	58,837	58,837
Adj R^2	0.0055	0.0053

*Notes. *p<0.05; **p<0.01; ***p<0.001. Standard errors in parentheses.*

Appendix L. Constructing Project Similarity using TFIDF and LSA

We construct the project similarity through the topics embedded in the whitepapers using the topic modeling techniques in our main analysis. We additionally construct the project similarity through the textual similarity of their whitepapers using the Term Frequency-Inverse Document Frequency (TF-IDF) model and Latent Semantic Analysis (LSA) model, which are widely used algorithms to generate textual similarity (Guzman and Li 2023, Henry and Leone 2016, Larsen and Bong 2016). TF-IDF model reflects how important a word is to a document in a collection or corpus by weighting a word based on its frequency in a text but compensated by its frequency in the corpus. The LSA model further transforms the TF-IDF vector space into a semantic space with a lower dimensionality to overcome the data sparsity issue and identify patterns and relationships between words. We report the estimations based on TFIDF and LSA in Table L1. We find qualitatively similar results, indicating the robustness of our findings.

Table L1. The Moderating Effect of Project Similarity

	Term Frequency – Inverse Document Frequency (TF-IDF)		Latent Semantic Analysis (LSA)	
	Prob.	log (Amt. + 1)	Prob.	log (Amt. + 1)
	(1)	(2)	(3)	(4)
I(Balance>0.1)	0.0053***	0.0345***	0.0053***	0.0341***
	(0.0007)	(0.0052)	(0.0007)	(0.0052)
I(Balance>0.1) × Project Similarity	**-0.0039*****	**-0.0248*****	**-0.0038*****	**-0.0244*****
	(0.0006)	**(0.0041)**	**(0.0006)**	**(0.0041)**
Project Similarity	-0.0017***	-0.0125***	-0.0017***	-0.0120***
	(0.0002)	(0.0016)	(0.0002)	(0.0016)
Balance-0.1	-0.5876**	-3.8363**	-0.6052***	-3.9802**
	(0.1808)	(1.2848)	(0.1805)	(1.2828)
I(Balance>0.1) × (Balance-0.1)	0.0743	0.5670	0.1002	0.7738
	(0.3037)	(2.1585)	(0.3036)	(2.1573)
Bandwidth	0.008	0.008	0.008	0.008
Kernel function	Triangular	Triangular	Triangular	Triangular
Eff. observations	58,837	58,837	58,837	58,837
Adj R2	0.0038	0.0034	0.0036	0.0033

*Notes. *p<0.05; **p<0.01; ***p<0.001. Standard errors in parentheses.*

Appendix M. Project Characteristics

We present the key characteristics of our focal project, and the 13 projects distribute their promotional tokens randomly to Ethereum users in Table M1.

Table M1. Key Characteristics of Blockchain Projects

Panel A

	Focal project	Project A	Project B	Project C	Project D
Announcements before the airdrop	NO	NO	NO	NO	NO
Airdrop distributing rule	Cutoff (0.1 ETH)	Randomly	Randomly	Randomly	Randomly
Number of token receivers	926,957	71,107	297,710	178,231	98,501
Rewarding for token receivers	5% bonus	5% bonus	5% bonus	fixed quantity	5% bonus
Industry	Platform services	Platform services	Data storage	Finance	Data storage
Token Type	Utility token	Utility token	Utility token	Utility token	Utility token
Whether connected with DApp	Yes	Yes	Yes	Yes	Yes
Targeted funding size	$10,000,000	$3,600,000	$5,000,000	$148,009,000	$5,000,000
Raised capital	$45,252,684	$18,638,214	$4,699,950	$153,121,448	$5,000,000
Whether listed successfully	Yes	Yes	No	Yes	No

Panel B

	Project E	Project F	Project G	Project H	Project I
Announcements before the airdrop	NO	NO	NO	NO	NO
Airdrop distributing rule	Randomly	Randomly	Randomly	Randomly	Randomly
Number of token receivers	71,109	16,123	8,522	7,840	7,403
Rewarding for token receivers	30% bonus	5% bonus	5% bonus	fixed quantity	5% bonus
Industry field	Health Care	Finance	Platform services	Trading	Finance
Token Type	Utility token	Utility token	Utility token	Utility token	Security token
Whether connected with DApp	Yes	Yes	Yes	Yes	No
Targeted funding size	$2,000,000	$3,000,000	$200,000	$3,000,000	$150,000
Raised capital	$19,703,656	$37,189,195	$200,000	$3,000,000	$276,420
Whether listed successfully	No	Yes	No	No	No

Panel C

	Project J	Project K	Project L	Project M
Announcements before the airdrop	NO	NO	NO	NO
Airdrop distributing rule	Randomly	Randomly	Randomly	Randomly
Number of token receivers	4,565	2,999	1,634	1,109
Rewarding for token receivers	5% bonus	5% bonus	5% bonus	5% bonus
Industry field	Entertainment	Chemical industry	Lending	Finance
Token Type	Utility token	Security token	Utility token	Utility token
Whether connected with DApp	Yes	No	Yes	Yes
Targeted funding size	$5,000,000	$18,500,000	$750,000	$2,000,000
Raised capital	$5,000,000	$2,135,000	$750,000	$11,230,000
Whether listed successfully	No	No	No	Yes