Everthing you ever wanted to know about Punks and Apes but never dated to ask

3rd Annual Boca Corporate Finance and Governance Conference, December 2022

Carol Alexander

Professor of Finance, University of Sussex Business School

Alexander and Chen (2022)

Many thanks to Dr. Peter M. Williams for very useful feedback and discussions
and to OutRank for implementing these and other models for ranking and pricing NFTs

Outline

1. Background
2. Measuring Rarity
3. Generalized Power Means
4. Score and Rank Invariance
5. Collection Characters

Background

Main Types of Non-Fungible Token (NFT)

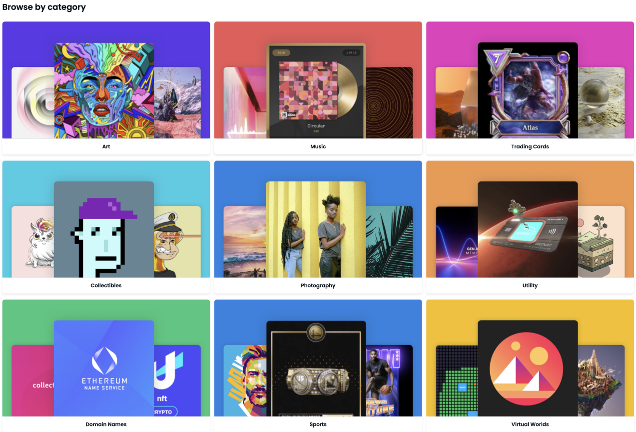
- Music and Videos (open and limited editions)
- Event Tickets (sports games or concerts, etc.)
- Real Estate (physical and digital)
- Game Accessories (weapons, land, fashion items, clothes)
- Photography and Art (physical and digital)

NFT Ecosystem

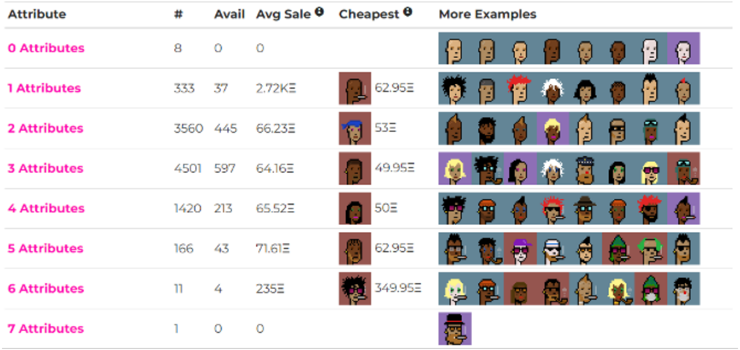
- Value of NFT sales in 2021 $\sim$ \$22 billion
- Market cap of top 100 NFTs $\sim$ \$16.7 billion (based on floor price)
- Ecosystem: marketplaces, exchanges, price and rarity analytics providers, PFP generators, data aggregators, lending platforms, betting shops
- Examples of valuations/amount raised:
  - Nansen: raised \$75m, Dec 2021, Series B
  - OpenSea: valuation \$13.3bn, Jan 2022, Series C
  - NFTgo: raised \$6.7m, Feb 2022, Series A
  - Upshot: raised \$29m, Mar 2022, Series A
  - NFTbank: raised \$12m, April 2022, Series A
  - Curio: raised \$7m, May 2022, seed
  - Magic Eden: valuation \$1.6bn, Jun 2022, Series B

NFT Marketplaces

OpenSea



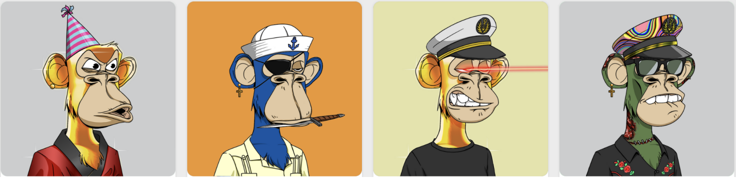
CryptoPunks (Larva Labs 2017)



Attribute	#	Avail	Avg Sale	Cheapest	More Examples
0 Attributes	8	0	0		
1 Attributes	333	37	2.72KΞ	62.95Ξ	
2 Attributes	3560	445	66.23Ξ	53Ξ	
3 Attributes	4501	597	64.16Ξ	49.95Ξ	
4 Attributes	1420	213	65.52Ξ	50Ξ	
5 Attributes	166	43	71.61Ξ	62.95Ξ	
6 Attributes	11	4	235Ξ	349.95Ξ	
7 Attributes	1	0	0		

- Average sale price $\sim$ 200 ETH
- Total value Punks sold exceeds \$2 billion
- Highest price ever paid \$23 million

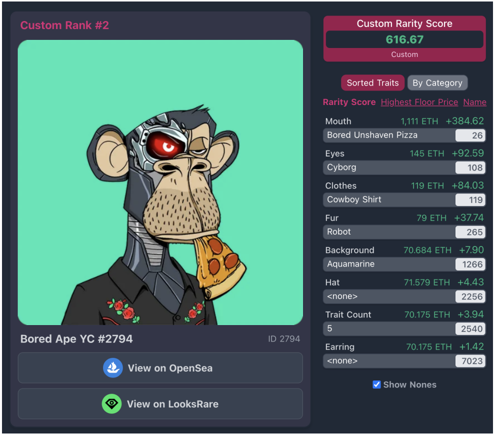
Bored Ape Yacht Club (Yuga Labs 2021)



- Sold out @ 0.08 ETH within a week
- Seven traits: background, clothes, earring, eyes, fur, hat and mouth
- Very rare attributes: solid gold fur, dagger mouth, x-ray eyes, etc.
- Highest price ever paid \$3.4 million
- Trait distributions $\sim$ independent

Rarity Tools

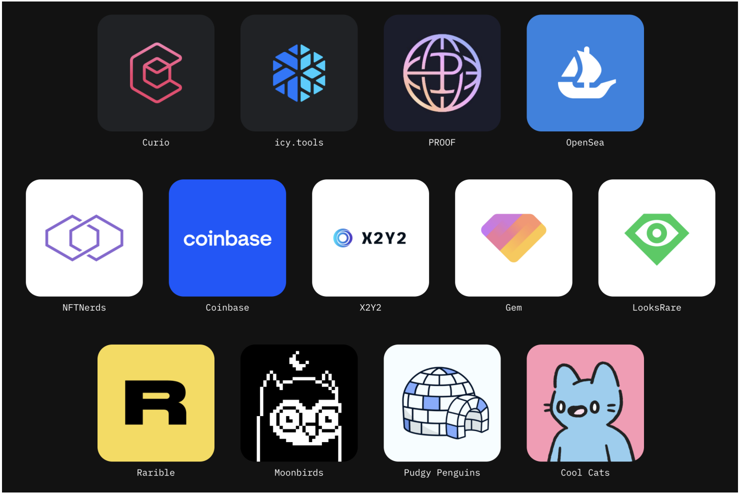
- Harmonic rank 2
- Arithmetic rank 5040



Rarity Metrics in Use

Platform	Minimum	Harmonic	Geometric	Arithmetic	Unknown
Collective.xyz					✓
CryptoSlam					✓
HowRare.is		✓*	✓		
Icy.tools			✓		
LuckyTrader		✓*			
Nansen				✓	
NFTEXP					✓
NFTgo				✓	
NFTinit					✓*
NFTonchained		✓			
NFTSniff		✓			
NFTStats†		✓			
OpenRarity			✓		
RankNFT		✓			
Rarity.tools	✓*	✓*	✓*	✓*	
RarityMon†		✓			
RaritySniffer†		✓*			
Rarity Sniper		✓			
Traitsniper		✓			

OpenRarity


Curio
icy.tools
PROOF
OpenSea
NFTNerds
coinbase
Coinbase
X2Y2
X2Y2
Gem
LooksRare
Rarible
Moonbirds
Pudgy Penguins
Cool Cats

OpenRarity

Surprisal Ranking Algorithm

Information content is an alternative way of expressing probabilities that is more well suited for assessing rarity. Think of it as a measure of how surprised someone would be upon discovering something.

1. Probabilities of 1 (i.e. every single token has the Trait) convey no rarity and add zero information to the score.
2. As the probability approaches zero (i.e. the Trait becomes rarer), the information content continues to rise with no bound. See equation below for explanation.
3. It is valid to perform linear operations (e.g. addition or arithmetic mean) on information, but not on raw probabilities.

Information content is used to solve lots of problems that involve something being unlikely (i.e. rare or scarce). This video shows how it was used to solve Wordle and also has an explanation of the equations, along with graphics to make it easier to understand. You can skip straight to the part on information theory if you'd like.

The score is defined as:

$$\frac{I(x)}{\mathbb{E}[I(x)]} \text{ where } I(x) = \sum_{i=1}^{n} -\log_2(P(trait_i))$$

NFTgo

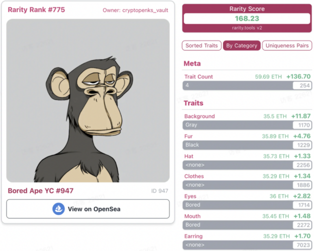
NFTGo Rarity Model

Rarity model created by NFTGo.io

Essentially, the rarity of an object depends largely on its differences from other objects: the greater the difference is, the more special and rarer this object will be. Quantifying the combined difference between an object and other things within the group reflects the object's rarity. Based on this logic, NFTGo.io uses a statistical approach called Jaccard distance to evaluate the real rarity of NFTs. An explanation of the specific algorithm can be found below.

What is Jaccard distance?

NFTGo.io's innovative rarity evaluation model goes beyond the mainstream approaches employed by products like rarity.tools. The common way of approaching NFT rarity is flawed and it has a low accuracy in some cases. An example of an inaccurate prediction from these models is the rarity score for Ape #947 from the Bored apes club. Rarity.tools gave a rarity score of 168.23. Even though the NFT does not seem to have any unique attributes.



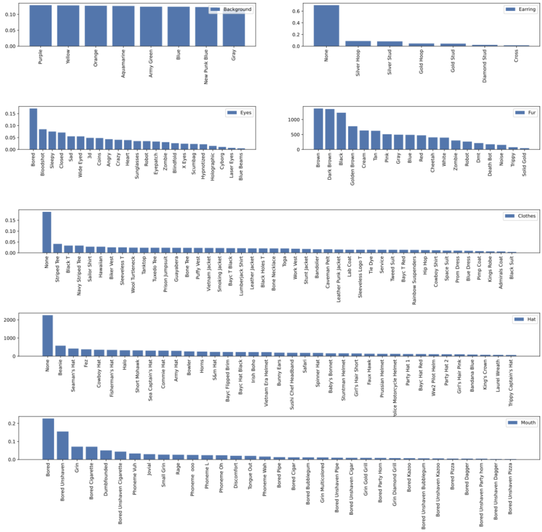
Generation of PFP Collections

- Very large number ($\sim$ 10,000) of tokens, each certifying ownership of a different computer-generated avatar
- PFP collection first defined by setting traits – example: background, gender, hair, eyes and clothes
- Each trait has a certain number of attributes – example: 10 possible background colours, 3 possible gender attributes, 4 hair attributes (including ‘none’, meaning no hair), 5 eye attributes and 8 clothes attributes
- Each avatar has a unique attribute map which assigns exactly one attribute to each trait
- So total number of tokens $\leq 10 \times 3 \times 4 \times 5 \times 8 = 4,800$ in this example, but there may be less than 4,800 because some attribute combinations may not be represented by an avatar

Rarity Analytics Platforms

- Rarity based entirely on trait distributions BAYC
- A token's raw rarity measurement is calculated using a rarity metric applied to the token's attribute counts or frequencies
- Raw measurement converted to a rarity score and a rarity rank where most rare token is rank 1 and least rare is rank 10,000 (or whatever the number of tokens in the collection is)
- Currently there are ~ 20 different rarity analytics platforms. OpenRarity and NFTgo each claim to provide a new rarity metric which they also propose as an industry standard
- However, every platform provides identical rankings to those derived from one (or more) of four basic metrics and these methods are almost all mathematically incorrect

BAYC Trait Distributions



Traits, Attributes and Counts

Traits X_i, No. Attributes θ_i, $i = 1, \ldots, n$

Unique Token ID:

$$A^k \longleftrightarrow \{X_1 = x_1^k, \ldots, X_n = x_n^k\}$$

Count of a particular attribute x_{ij} for trait X_i is:

$$m_{ij} = |k \ni x_i^k = x_{ij}| = \sum_{k=1}^{m} \mathbb{1}_{x_i^k = x_{ij}}$$

Every token has exactly one of each attribute, hence:

$$\sum_{j=1}^{\theta_i} m_{ij} = m, \qquad \text{for } i = 1, \ldots, n$$

Setting $m_i^k = m_{ij} | x_{ij} = x_i^k$ for $i = 1, \ldots, n$ yields injection:

$$A^k \longrightarrow \{m_1^k, \ldots, m_n^k\}$$

Attribute Frequency

$$p_{ij} = P\left(X_i = x_{ij}\right) = \frac{m_{ij}}{m}$$

Every token has exactly one of each attribute:

$$\sum_{j=1}^{\theta_i} p_{ij} = 1, \qquad \text{for } i = 1, \ldots, n$$

Injective map:

$$A^k \longrightarrow \left\{p_1^k, \ldots, p^k\right\}$$

where for $i = 1, \ldots, n$

$$p_i^k = \frac{m_i^k}{m}$$

Weighted Power Means

- Let p be any non-zero real number
- Let $\{a_1, \ldots, a_n\}$ be a set of positive real numbers
- Let $\{\omega_1, \ldots, \omega_n\}$ be a set of positive weights

Weighted power mean with exponent p:

$$M_p\left(a_1, \ldots, a_n | \omega_1, \ldots, \omega_n\right) = \left(\frac{\sum_{i=1}^{n} w_i a_i^p}{\sum_{i=1}^{n} w_i}\right)^{1/p}$$

Special Cases

Limits:

$$M_{-\infty}(a_1,\ldots,a_n|\omega_1,\ldots,\omega_n) = \min\{\omega_1 a_1,\ldots,\omega_n a_n\}$$

$$M_{\infty}(a_1,\ldots,a_n|\omega_1,\ldots,\omega_n) = \max\{\omega_1 a_1,\ldots,\omega_n a_n\}$$

$$M_0(a_1,\ldots,a_n|\omega_1,\ldots,\omega_n) = \left(\prod_{i=1}^{n} a_i^{\omega_i}\right)^{1/\sum_{i=1}^{n}\omega_i}$$

Trait Normalization (Simple):

$$\{a_1,\ldots,a_n\} = \{m_1^k,\ldots,m_n^k\} \quad \text{and} \quad \{\omega_1,\ldots,\omega_n\} = \{\theta_1,\ldots,\theta_n\}$$

Pythagorean Means

$$\{\omega_1, \dots, \omega_n\} = \{1, \dots, 1\} \to \text{Generalized Means}$$

1. Minimum :

$$r^k_{\text{min}} = \min_i p^k_i$$

2. Harmonic:

$$r^k_{\text{harm}} = \frac{n}{\sum_{i=1}^n \left(p^k_i\right)^{-1}}$$

3. Geometric:

$$r^k_{\text{geo}} = \prod_{i=1}^n \left(p^k_i\right)^{1/n}$$

4. Arithmetic:

$$r^k_{\text{arith}} = n^{-1} \sum_{i=1}^n p^k_i$$

Rarity Scores and Ranks

- Rarity score $r^k \to \bar{r}^k \in [0,1]$

$$\bar{r}^k = \frac{r^k - r_{\text{min}}}{r_{\text{max}} - r_{\text{min}}}$$

- Rarity rank $r^k \to R^k \in \{\ 1, \ldots, m\}$

Rank $1 =$ most rare, ... , Rank $m =$ least rare

- Pythagorean mean ordering:

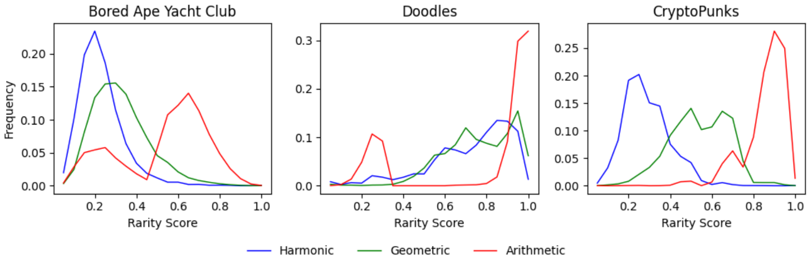
$$r^k_{\text{min}} \leq r^k_{\text{harm}} \leq r^k_{\text{geo}} \leq r^k_{\text{arith}} \leq r^k_{\text{max}}$$

Ordering lost on conversion to scores or ranks

- Correlations between scores (and ranks) derived from different metrics can be very low, depending on the collection and the two metrics selected

Score Distributions

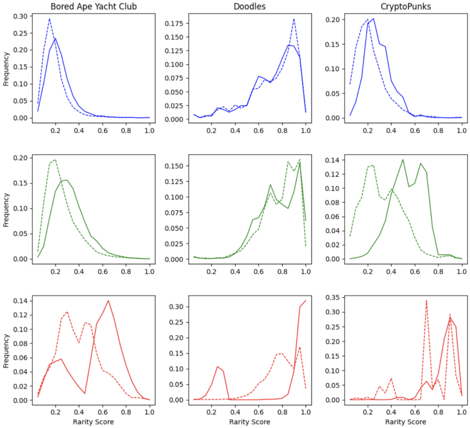
Figure: Rarity Scores from Harmonic, Geometric and Arithmetic Means. The relative frequency plots of rarity scores derived from harmonic ($p = -1$), geometric ($p = 0$) and arithmetic ($p = 1$) mean rarity metrics. These empirical densities are computed using each collection's trait data which are downloaded from OpenSea. Pearson correlation (scores) between 0.14 and 0.87; Kendall's τ (ranks) between 0.45 and 0.77.



Correlations between Scores and Ranks

	BoredApeYachtClub			Doodles			CryptoPunks		
Panel A: Pearson's ρ on Rarity Scores									
	Har.	Geo.	Arith.	Har.	Geo.	Arith.	Har.	Geo.	Arith.
Har.	1	0.872	0.463	1	0.792	0.14	1	0.805	0.461
Geo.	0.872	1	0.780	0.792	1	0.67	0.805	1	0.850
Arith.	0.463	0.780	1	0.14	0.67	1	0.461	0.850	1
Panel B: Spearman's ρ on Rarity Ranks									
	Har.	Geo.	Arith.	Har.	Geo.	Arith.	Har.	Geo.	Arith.
Har.	1	0.827	0.514	1	0.805	0.605	1	0.809	0.593
Geo.	0.827	1	0.866	0.805	1	0.935	0.809	1	0.917
Arith.	0.514	0.866	1	0.605	0.935	1	0.593	0.917	1
Panel C: Kendall τ on Rarity Ranks									
	Har.	Geo.	Arith.	Har.	Geo.	Arith.	Har.	Geo.	Arith.
Har.	1	0.641	0.374	1	0.654	0.453	1	0.635	0.442
Geo.	0.641	1	0.716	0.654	1	0.798	0.635	1	0.773
Arith.	0.374	0.716	1	0.453	0.798	1	0.442	0.773	1

Effect of Trait Normalization



OutRank

Invariance Results

1. Rarity scores are invariant (reversed) under increasing (decreasing) linear transformations of the raw rarity metric

   For example, we can set $\{a_1, \ldots, a_n\}$ to be either counts or frequencies, the score is the same

2. Rarity ranks are invariant under strict monotonic transformations of a rarity metric

   (Assumes decreasing transformations reverse rank order map)

3. The kurtosis of a rarity score distribution increases (decreases) leptokurtic under convex (concave) strict monotonic transformations of a rarity metric

Equivalence Results

- OpenRarity information content is a decreasing, non-linear strictly monotonic transformation of the geometric mean frequency, so

  OpenRarity $\equiv$ ranking by geometric mean

- NFTgo Jaccard distance is a decreasing, linear transformation of the arithmetic mean frequency, so

  NFTgo $\equiv$ ranking by arithmetic mean

- Every rarity metric currently in public domain is a strict monotonic transformation of, and therefore yields a ranking equivalent to, a weighted power mean with $p = -1, 0, 1$ or $-\infty$

The Ultimate Guide

Consider two tokens:

$$A^k \longleftrightarrow \{x_1^k, \dots, x_n^k\}$$

and

$$A^l \longleftrightarrow \{x_1^l, \dots, x_n^l\}$$

Jaccard similarity is proportion of attributes in common:

$$n^{-1}|i \ni x_i^k = x_i^l, i = 1, \dots, n| = n^{-1} \sum_{i=1}^{n} \mathbb{1}_{x_i^k = x_i^l}$$

Jaccard distance is one minus the Jaccard similarity:

$$JD(A^k, A^l) = 1 - n^{-1} \sum_{i=1}^{n} \mathbb{1}_{x_i^k = x_i^l}. \tag{1}$$

Equivalence Proof

NFTgo define their rarity metric as the arithmetic average of a token's Jaccard distances from all the other tokens

$$r^k = m^{-1} \sum_{l=1}^{m} JD(A^k, A^l). \tag{2}$$

Combining (1) and (2) yields:

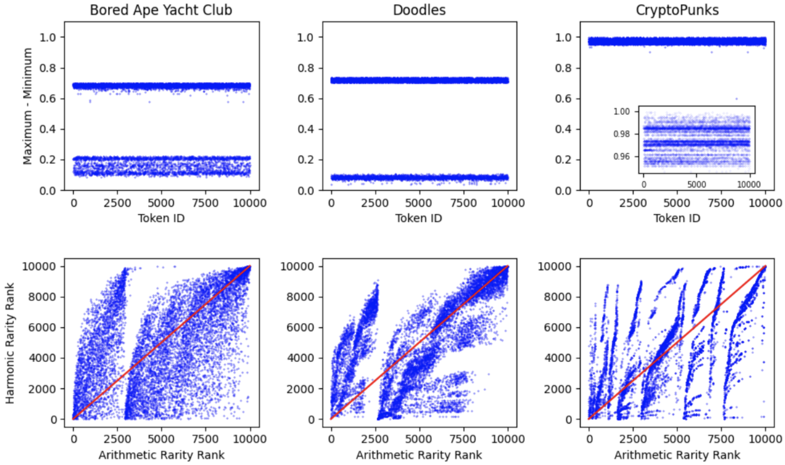
$$\begin{aligned} r^k &= m^{-1} \sum_{l=1}^{m} \left(1 - n^{-1} \sum_{i=1}^{n} \mathbb{1}_{x_i^k = x_i^l} \right) \\ &= 1 - n^{-1} \sum_{i=1}^{n} \sum_{l=1}^{m} m^{-1} \mathbb{1}_{x_i^k = x_i^l} \\ &= 1 - n^{-1} \sum_{i=1}^{n} p_i^k = 1 - r^k_{\text{arith}} \end{aligned}$$

OutRank's Collection Characters

Trait Independence

Panel A: Bored Ape Yacht Club							
	Background	Clothes	Earring	Eyes	Fur	Hat	Mouth
A.1: Independence test							
Background		**359**	40	149	109	282	219
Clothes	342		271	995	760	**1961**	1372
Earring	58	296		126	125	269	174
Eyes	184	1019	160		377	**1163**	675
Fur	153	840	133	443		628	529
Hat	290	1641	251	859	708		1332
Mouth	260	1463	225	767	633	1232	
A.2: Cramer's V							
Background		0.0716	0.0257	0.0461	0.0395	0.0635	0.0560
Clothes			0.0672	0.0673	0.0650	0.0738	0.0655
Earring				0.0459	0.0456	0.0670	0.0538
Eyes					0.0458	0.0727	0.0554
Fur						0.0591	0.0542
Hat							0.0645

Bar Codes and QR Codes



Main Results

1. Rarity ranks are invariant under strict monotonic transformations

2. Rarity scores are invariant (reversed) under increasing (decreasing) linear transformations

3. If and only if traits are independent then the geometric mean count or frequency, or any strict monotonic transformation thereof, are statistically correct rarity metrics

4. All other rarity metrics in the public domain are not statistically correct

5. And when traits are not independent an entirely different approach needs to be developed

OutRank