Applying the Yule-Simon Process to Ranking Network Accesses

Naoshi Sakamoto

Department of Information and Communication Engineering, School of Engineering,
Tokyo Denki University, Tokyo, Japan
Email: sakamoto@c.dendai.ac.jp

***Abstract*—Power-law distributions have been observed in various phenomena throughout history. In this study, we examined the access rankings of video streaming sites and the follower rankings of music streaming sites. While videos and music span a wide range of genres, we conducted observations for each genre separately. As a result, we found that these rankings followed a power-law distribution with exponents that fell within a remarkably narrow range. To explain this behavior, we discussed the applicability of the Yule-Simon process as a natural model. This process models positive feedback in which fans belonging to a particular genre generate new fans with equal probability. Consequently, a one-parameter model was successfully applied to video streaming sites, while a two-parameter model was applicable to music streaming sites.**

***Keywords*—power law, Yule-Simon process, internet access analysis**

I. Introduction

The probability of measuring a specific value of a certain quantity has been observed to inversely vary with the power of that value in various contexts. This phenomenon is referred to as a power-law relationship. It is also known as Zipf’s law or the Pareto distribution, and it has been observed across a wide range of fields. According to Newman’s review article [1], power-law relationships appear broadly in physics, biology, earth and planetary sciences, economics, finance, computer science, demography, and social sciences.

In various user behaviors on the internet, individuals primarily act based on their own will. However, when observed at the group level, statistical patterns often emerge. Notably, behaviors following power-law relationships have been observed. In the context of network service sites, understanding and predicting user behavior is valuable for estimating the necessary resources, such as server capacity, required response speed, and network bandwidth. Therefore, statistical analysis of user behavior is crucial.

In recent years, Application Programming Interfaces (APIs) that provide statistical information have been made publicly available by video platforms and music services. These APIs enable statistical analysis and allow for the study of user behavior. In this study, we collected and analyzed data from the video streaming site Niconico [2] and the music information service Spotify [3] to model user behavior. Niconico organizes uploaded videos by genre, and we observed the number of views for videos within each genre. Niconico is primarily a Japan-based platform with a unique feature where user comments are displayed over the video screen.

For Spotify, where music is categorized by genre, we observed the number of followers for each artist. Initially, we hypothesized that similar power-law relationships would be observed across all genres. However, while power-law distributions were indeed observed in all genres, the exponents of these power-law relationships varied across genres. Nevertheless, the exponents were found to follow a relatively narrow and specific distribution.

To explain this phenomenon, we adopted the Yule-Simon process as a model. This approach naturally accounts for the narrow distribution of the exponents, providing a coherent explanation for the observed behavior.

The number of accesses to specific content or followers of a particular artist is largely determined by human actions. As a result, there are often more outliers compared to the values predicted by the model. However, in the case of Spotify, significant deviations were observed, too large to be treated simply as noise. To address this, we introduced a model incorporating time-dependent parameters that account for changes in user behavior over time, which successfully explained the observed discrepancies.

This paper is structured as follows: Section II introduces related works. Section III presents the observational results from two network content services. Section IV applies the probability distribution derived from the Yule-Simon process to rankings and demonstrates how the model parameters can be interpreted through the application of the Yule-Simon process. Finally, Section V provides the conclusion.

II. Related Works

The property that rankings and sizes follow an inverse power-law relationship, known as Zipf’s law, has been

Manuscript received September 27, 2024; revised December 22, 2024, accepted March 4, 2025; published June 5, 2025.

extensively studied for many years [4]. Newman wrote a comprehensive paper explaining the connections between power-law relationships, the Pareto distribution, and Zipf's law. He summarized research on power-law relationships across various fields and provided explanations of important theories [1]. These laws share characteristics that commonly appear in systems with extreme frequency or distribution imbalances. Specifically, the Pareto distribution is applied to wealth distribution, and Zipf's law to distributions like language [4, 5] and city population sizes [6] both grounded in power-law principles. The paper examines their unified theoretical foundations and practical examples across diverse fields.

Ha *et al*. [7] extended Zipf's law from words to phrases in their research. While the frequency distribution of words was already known to follow an inverse power law, this study demonstrated similar patterns for phrases. By analyzing the relationship between phrase length and frequency, they confirmed the universality of Zipf's law in natural language, suggesting broader applications in linguistic data and natural language processing fields.

Zanette explored Zipf's law and multiplicative processes in urban growth [8]. Zipf's law indicates that the size distribution of cities follows a specific pattern, with few large cities and many small ones. This paper attributed the emergence of Zipf's law to multiplicative processes—where cities grow proportionally—and examined the dynamics of urban growth, aligning the theory with empirical data.

Chen *et al.* [9] analyzed Zipf's law using an exponential approach. Their study measured the accuracy of the law using its exponent (power-law exponent) and provided detailed analyses of size distributions in cities and other complex systems. They also examined deviations and exceptions caused by variations in the exponent, exploring more generalized regularities in these systems.

The Yule-Simon process discussed in this paper is based on a model proposed by Yule [10] and later analyzed by Simon [11]. Their work demonstrated that this class of distributions can be applied to various statistical phenomena, providing concrete examples, and analyzing its properties. As a result, it introduced a theoretical framework useful for modeling imbalanced data in the real world.

Building on this, Baur and Bertoin extended the traditional Yule-Simon distribution by proposing a model with two parameters [12]. This extension enhanced the flexibility of the distribution and improved its fit to data.

In this study, we apply the two-parameter Yule-Simon distribution to model the observed phenomena.

III. Observing Rankings in Actual Network Services

In this section, we observe the rankings of actual network services. First, we collect data on all major genres from each site. Next, for each genre, we gather ranking data up to approximately the top 1000 entries. The data is then visualized by plotting it on a log-log graph.

Additionally, to visualize the distribution of slopes and mitigate the impact of outliers, we overlay a graph that divides the values by the 1/4 power of 1000, which corresponds to approximately the 5th position in logarithmic scale (the 25th percentile). This approach helps reduce the influence of extreme values and provides a clearer view of the slope distribution.

Next, we calculated the exponents for each genre. This was done by performing linear regression on the logarithms of the values. However, only the first 5% of the data was used for this calculation. This is because power-law relationships are typically used as an approximation for larger values (i.e., when the ranks are low), and the behavior of the lower-ranked items differs from the higher-ranked ones.

A. Observation of View Count in Niconico

Niconico [2] is a video streaming site primarily serving the Japanese audience. A distinctive feature of Niconico is that viewer comments are displayed directly on the video screen as the video plays. Videos are categorized by genres and tags, and the number of views can be obtained through the site's API.

The English translations of the genres published by Niconico are as follows: entertainment, MMD, virtual, try2draw, game, Idolmaster, TOHO, live talk, TRPG, sports, others, diary, you-know-what, try2dance, radio, music, ASMR, Vocaloid, try2sing, try2play, indies, Niconico tutorials, history, science, Niconico tech club, craftwork, try2make, nature, politics, onboard camera, rail, animals, travel, cooking, and R-18.

The ranking of view counts up to the 1000th position was plotted for the 35 genres in Fig. 1.

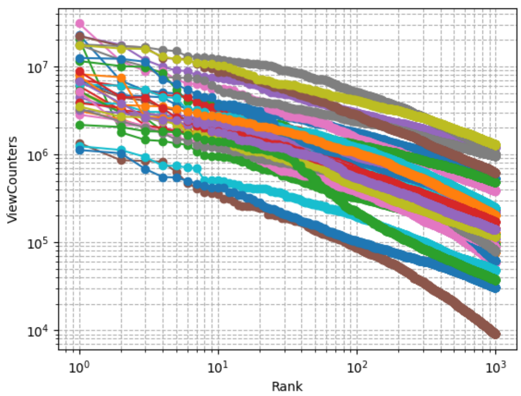

Fig. 1. Rank number of ViewCounter for 35 genres in Niconico.

In this graph, while the heights of the lines differ, the slopes are similar. Additionally, there are outliers in the higher ranks. To address this, when the data is normalized by dividing by the number of followers of the 5th rank, it results in Fig. 2. Furthermore, we calculate the slope for each genre. In Fig. 2, the rankings for each genre are not a perfect straight line, and the slope is different between the first and second halves of the data. For this reason, we calculated an exponential regression from $1000^{1/4} \sim 5$ to $1000^{2/4} \sim 33$, which are the quartiles in a double logarithmic graph.

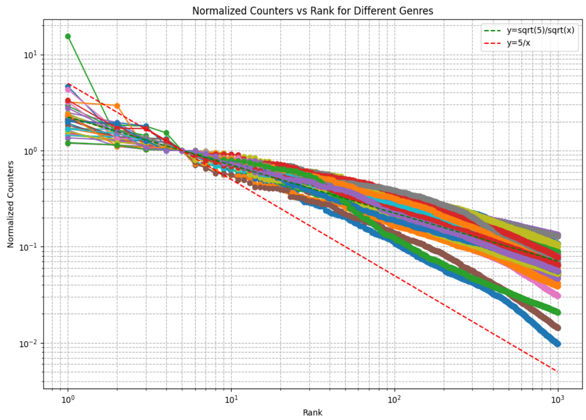

Fig. 2. Relationship between number of followers and ranking, normalized by dividing by number of followers of 5th place for each genre on Niconico.

Fig. 3 shows the empirical distribution function of the exponential coefficient obtained from the exponential regression. As you can see, the slope is in the vicinity of −1/2.

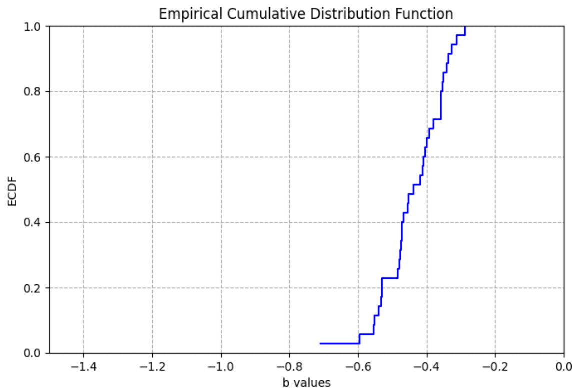

Fig. 3. Empirical distribution function of exponential coefficients obtained by power regression of the top parts of each genre on Niconico.

B. Observation of Number of Followers in Spotify

In this section, we observed the follower counts of artists on the music streaming service Spotify [3] as a characteristic of content service rankings. Spotify provides an API, and by adhering to the access limitations, it is possible to download music-related information.

Artists on Spotify are categorized into genres. Then, we collected information on the top 1000 artists based on their follower counts for each of the 44 genres as follows: acoustic, afrobeat, alt-rock, alternative, ambient, anime, black-metal, bluegrass, blues, brazil, breakbeat, british, cantopop, chicago-house, children, chill, classical, club, comedy, country, dance, dancehall, deathmetal, deep-house, detroit-techno, disco, drum-andbass, dub, dubstep, edm, electro, electronic, emo, folk, forro, french, funk, garage, german, gospel, goth, grindcore, groove, grunge, guitar, happy, hard-rock, hardcore, hardstyle, heavy-metal, hip-hop, honky-tonk, house, idm, indian, indie, indie-pop, industrial, iranian, j-dance, j-idol, j-pop, j-rock, jazz, k-pop, kids, latin, latino, malay, mandopop, metal, metalcore, minimaltechno, mpb, new-age, opera, pagode, party, piano, pop, pop-film, power-pop, progressive-house, psychrock, punk, punk-rock, r-n-b, reggae, reggaeton, rock, rock-n-roll, rockabilly, romance, sad, salsa, samba, sertanejo, show-tunes, singer-songwriter, ska, sleep, songwriter, soul, spanish, study, summer, swedish, synthpop, tango, techno, trance, trip-hop, turkish, and worldmusic. The follower counts for these artists, ranked by their follower numbers, are displayed in a log-log graph in Fig. 4.

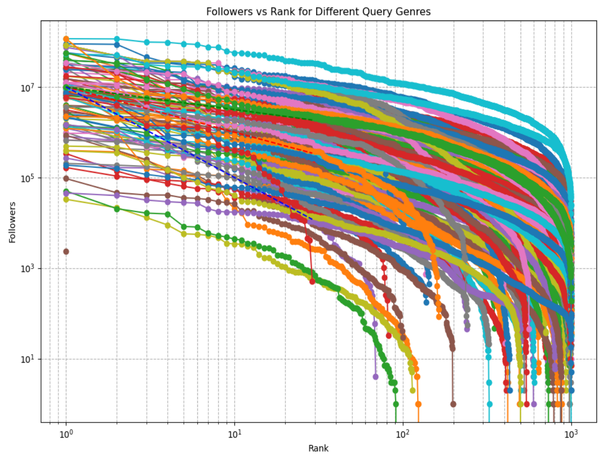

Fig. 4. The relationship between the number of followers and ranking of 44 genres on Spotify.

In also this graph, while the heights of the lines differ, the slopes and shapes of the graphs are similar. By normalizing the data by dividing by the follower count of the 5th-ranked artist, the resulting graph is shown in Fig. 5. As shown, the slopes for many genres fall between –1 and –1/2. When focusing on this range, the data can be explained by a one-parameter Yule-Simon process. The remaining deviations can be interpreted as errors or unforeseen fluctuations.

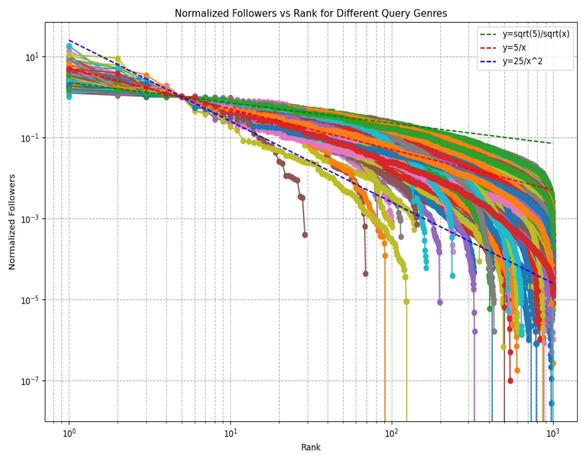

Fig. 5. The relationship between the number of followers and ranking, normalized by dividing by the number of followers of the 5th highest ranking in each genre on Spotify.

The exponents of the genres, sorted in descending order, are plotted in Fig. 6. Additionally, the major values are provided in Table I.

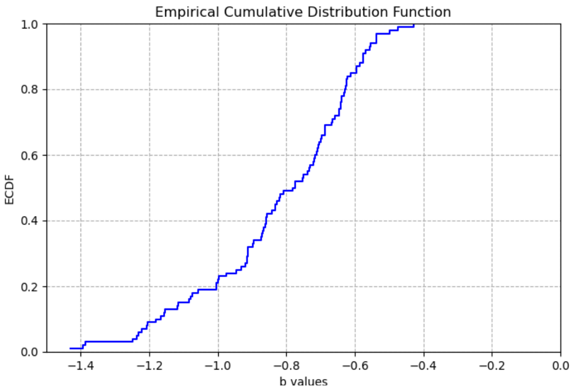

Fig. 6. Empirical distribution function of exponential coefficients obtained by power regression of the top rankings of each genre on Spotify.

TABLE I. EXPONENT FOR MAJOR GENRE

genre	exponent
alternative	−0.428
hip-hop	−0.555
folk	−0.568
country	−0.594
rock	−0.621
soul	−0.624
gospel	−0.626
pop	−0.640
jazz	−0.686
j-pop	−0.703
blues	−0.711
disco	−0.718
k-pop	−0.733
samba	−0.737
classical -	−0.809
hardcore	−0.828
club	−0.832
punk	−0.869
anime	−0.894
reggae	−0.897
ambient	−1.075
house	−1.115
heavy-metal	−1.249

IV. YULE-SIMON PROCESS FOR RANKING

In this section, we consider applying the Yule-Simon process to the ranking of content services.

A. Model Examination

It is evident from various observations that the rankings of content services are governed by a power-law distribution.

There are various causes for the emergence of power laws, but Newman summarizes them as follows [1].

1) Combinations of exponentials
2) Inverses of quantities
3) Random walks
4) The Yule process
5) Phase transitions and critical phenomena
6) Self-organized criticality
7) Other mechanisms for generating power laws.

Among these, we will examine whether the access distribution in public services can be explained using a simple model.

(1) We will examine the Combinations of Exponentials. This principle states that if a quantity y follows an exponential distribution, then the related variable x may exhibit a power-law behavior of the form $x \sim e^{by}$. For example, the frequency of generated words from random letters can follow a power-law distribution due to this principle [13, 14]. However, when considering quantities such as access numbers in content services, it is challenging to assume a logarithmic behavior relative to some distribution of the content service.

(2) We will examine Inverses of Quantities. This principle suggests that if two quantities are related as inverses, when one follows a distribution, the other follows a power-law distribution. This behavior is observed in physical systems, such as the Ising model, where there is a reciprocal relationship between quantities [15, 16].

However, it is difficult to conceive a natural model where access quantities are inversely related to another quantity, making this approach challenging for modeling access distributions in content services.

(3) We will examine Random Walks. A simple example of a power-law observed in random walks is the return time to the origin in a one-dimensional random walk. This principle has been applied to the distribution of species lifetimes in fossil records [17]. However, since this is a model for a single time series, it is not suitable for modeling sequences where access numbers continuously increase, as observed in content services. Therefore, it does not provide a fitting framework for explaining the growth patterns of access numbers over time.

(4) The Yule process. Yule hypothesized that the distribution of the number of species within a genus in the context of evolutionary theory follows a random pattern [10]. Species are classified into genera based on the following assumptions:

1. A speciation event occurs in which a new species arises from an existing species, and the new species belongs to the same genus.

2. Separately, a new species may suddenly arise a new genus that does not belong to any existing genus.

In this model, it is assumed that speciation event occur with equal probability across all species. Additionally, it is assumed that the creation of a new genus occurs once for every m speciation events. Analysis of such a stochastic process reveals that the probability distribution of genera containing k species approximates a power law. In the context of network content services, considering specific content as a "genus" and users of the service as "species," it is not unreasonable to assume the following:

1. A randomly chosen user contributes to increasing the user count of the content they are already using.

2. For every certain number of new users, a new content item with a single user emerges.

This analogy aligns well with the assumptions of the Yule process.

(5) The consideration of "phase transitions and critical phenomena" in the context of percolation theory explores how critical points in complex networks lead to scale-free behaviors, particularly in systems with random connectivity [18]. Percolation theory examines situations

where nodes in a network, which are geographically or topologically close, become connected under certain conditions, forming clusters with power-law distributions.

However, when applying this model to network services, it becomes clear that "user distances" or connections between users are not typically considered in simple content ranking models. This suggests that the percolation-based model may not directly apply to the analysis of user behavior or content rankings.

That being said, "community analysis" among users, where users within certain clusters are more likely to interact with one another (forming cohesive groups), could benefit from this model. In such a case, communities within the network might exhibit scale-free behavior similar to what is observed in percolation theory, where the size distribution of these user communities follows a power law.

(6) "Self-organized criticality (SOC)" is a concept used to describe systems that naturally evolve towards a critical state without requiring fine-tuning of parameters. This critical state is marked by a scale-invariant behavior, where small events can trigger large-scale phenomena, often following a power-law distribution. Examples include systems like forest fires, where a small spark can potentially lead to a massive blaze [19], or the spreading of links in the web, where the linking behavior [20] may trigger large-scale effects once a critical threshold is surpassed.

In the context of "network content services," the concept of self-organized criticality may not directly apply to explaining "content rankings" or "user behavior." The idea of small changes triggering large-scale events could potentially relate to viral content or a sudden surge in users accessing a particular piece of content.

(7) Other mechanisms for generating power laws. This section discusses mechanisms such as: Tolerance mechanisms [21, 22], Coherent noise mechanisms [23], and the approximation of power laws through the multiplication of random variables, which leads to a lognormal distribution [24–26].

The reason for choosing the Yule-Simon process is its simplicity and the small number of parameters.

1) Simplicity: The Yule-Simon process primarily relies on two main parameters: the probability of generation and the frequency at which new categories (species, content, etc.) appear. This simple model provides an intuitive framework for understanding real-world phenomena.
2) Applicability to real-world phenomena: The Yule-Simon process is particularly suited for systems with cumulative evolution, where new items or categories emerge from existing ones.
3) Few parameters: Despite being a random growth process, the Yule-Simon process determines behavior with relatively few parameters. This makes the model easy to apply and analyze.

For a Yule-Simon process, we consider a model where all accesses are independent and occur with the same probability. In this case we consider quantiles of multiple Poisson distributions. When all accesses do not occur uniformly, a ranking is derived based on the relative size of each access, which is a Gumbel distribution [27]. This is an exponential distribution rather than a power law, so it differs from the distribution obtained as it gets small quickly on a log-log graph and does not show linearity.

B. *Quantile for Yule-Simon Process*

The model applying the Yule-Simon process suggests that for content within a specific genre, when m users increase, a new piece of content is added with one user. The parameter m which determines the number of users needed to add new content, can vary depending on the genre.

In this case, Simon provided the following analysis. Let p_k represent the proportion of species with k individuals in a genus. If we define $\rho = 1 + \frac{1}{m}$, the following result is obtained.

This result refers to the distribution of the number of species (or content) within a given genus (or genre), where p_k is the probability that a genus contains k species. By using ρ, which depends on the parameter m, the process models how content grows in terms of user adoption.

The distribution follows a power law, which is a typical result in systems where growth is driven by the preferential attachment mechanism, as in the Yule-Simon process. The specific form of the distribution depends on m, which characterizes the rate of content growth in the model.

$$p_k = \rho B(k, \rho + 1), \tag{1}$$

In the Yule-Simon process, species are considered immortal, meaning once a species is created, it continues to behave in the same way indefinitely. In this case, the behavior is solely determined by the parameter m, which governs the rate at which new species are generated.

To determine the quantiles in this model, we analyze the distribution of the number of species with respect to m.

Theorem 1: In When the total number of species is denoted by n, and the number of species at rank r is denoted by s,

$$r \sim n\rho\Gamma(\rho + 1)\zeta(\rho + 1, s).$$

Proof.

$$\begin{aligned}\frac{r}{n} &\sim \sum_{k=s}^{\infty} p_k = \sum_{k=s}^{\infty} \rho B(k, \rho + 1) \\ &= \sum_{k=s}^{\infty} \rho \frac{\Gamma(k)\Gamma(\rho + 1)}{\Gamma(k + \rho + 1)} \\ &= \rho\Gamma(\rho + 1) \sum_{k=s}^{\infty} \frac{\Gamma(k)}{\Gamma(k + \rho + 1)} \\ &= \rho\Gamma(\rho + 1) \sum_{k=s}^{\infty} \frac{1}{k(k + 1)\cdots(k + \rho + 1)}.\end{aligned}$$

Given the assumption $s \gg \rho + 1$, the summation of the rank distribution can be approximated using the Hurwitz zeta function $\zeta(s, q)$ is defined as:

$$\begin{aligned}&\sim \rho\Gamma(\rho + 1) \sum_{k=s}^{\infty} \frac{1}{k^{\rho+1}} \\ &\sim \rho\Gamma(\rho + 1)\zeta(\rho + 1, s).\end{aligned}$$

Q.E.D.

Theorem 2: When s is enough large,

$$s \sim Cr^{-1+\frac{1}{m+1}}.$$

where C is a positive constant.

Proof. When s is enough large, we can approximate $\zeta(\rho+1,s) \sim \frac{1}{\rho s^{\rho}}$. Then,

$$\begin{aligned} r &\sim n\rho\Gamma(\rho+1)\zeta(\rho+1,s) \\ &\sim n\rho\sqrt{2\pi\rho}\left(\frac{\rho}{s}\right)^{\rho}\frac{1}{\rho s^{\rho}}. \\ s &\sim Cr^{-\frac{1}{\rho}} = Cr^{-1+\frac{1}{m+1}}. \end{aligned}$$

Q.E.D.

Baur and Bertoin investigated a two-parameter Yule-Simon process where the rate of divergence of a species with age $a \geq 0$ follows the function $e^{-\theta a}$, where θ is a parameter. This means that older species and newer species have different probabilities of diversifying.

In this model, the parameter θ determines how quickly the species' rate of diversification decreases as their age increases. The larger the value of θ, the more rapidly the rate of diversification diminishes for older species. When $\theta = 0$, this model reduces to the standard Yule-Simon process, where the rate of diversification is independent of the age of the species.

The introduction of the parameter θ allows for a more flexible and realistic representation of species' evolution, as it accounts for the fact that older species may become less likely to give rise to new species over time, while newer species may be more likely to diversify.

Theorem 3: (Theorem 4.1 see [12])

Let $\rho > 0$.

1. If $\theta < 1$, then there exists a constant $C = C(\theta,\rho) \in (0,\infty)$ such that, as $n \to \infty$:
$$\Pr\left[X_{\theta,\rho} > n\right] \sim Cn^{-\frac{\rho}{1-\theta}}.$$
2. If $\theta > 1$, then as $n \to \infty$:
$$\operatorname{Ln}\Pr\left[X_{\theta,\rho} > n\right] \sim -\left(\ln\theta - 1 + \frac{1}{\theta}\right)n.$$

As a result, we derive the ranking formula:

Theorem 4: In a two-parameter Yule-Simon process in which a genus is generated once every m differentiations, if $\theta < 1$ and the differentiation rate of species with age $a \geq 0$ is $e^{-\theta a}$, then the number of species s in the rth most common genus can be approximated by

$$s \sim Cr^{-(1-\theta)\left(1-\frac{1}{m+1}\right)},$$

If the number of families is sufficiently large and s is also sufficiently large.

Proof. For $\rho = 1 + \frac{1}{m}$ and $\theta < 1$, let $X_{\theta,\rho}$ be a random variable of a two-parameter Yule-Simon process.

In this case, if the total number of genera is n and the number of species of rank r is s,

$$\Pr\left[X_{\theta,\rho} > n\right] = \frac{r}{n}. \tag{2}$$

Now, by applying the results of Baur *et al.* [12],

$$\begin{aligned} \frac{r}{n} &= \Pr\left[X_{\theta,\rho} \geq s\right] \sim Cs^{-\rho/(1-\theta)} \\ s &\sim \left(\frac{r}{Cn}\right)^{-\frac{1-\theta}{\rho}} \\ &= C'r^{-(1-\theta)(1-\frac{1}{m+1})} \end{aligned} \tag{3}$$

Q.E.D.

According to Theorem 4, assuming that $\theta < 1$ is a universal value and considering the condition $m > 0$, the range of the exponent for ranking r is determined as follows:

$$-(1-\theta) \leq -(1-\theta)\left(1-\frac{1}{m+1}\right) < 0. \tag{4}$$

In particular, when $\theta = 0$,

$$-1 \leq -\left(1-\frac{1}{m+1}\right) < 0. \tag{5}$$

C. Discussion

In the case of Niconico's ViewCount, the distribution closely followed the assumptions of the Yule-Simon process for each genre. Specifically, the tight distribution of the slopes indicates that the genres are aligned with the concentration of videos. This suggests that in each genre, the pace at which new content is supplied is similar, and the volume of content tends to be comparable across genres.

On the other hand, for Spotify's artist follower count, the results for higher ranks closely matched the model. However, for lower ranks, there may be a lack of content supply, leading to a sharp decline on the right-hand side, which decreases faster than a power law. For the higher ranks, the model for follower growth can be considered as following the Yule-Simon process, where the probability is proportional to the number of followers. However, looking at Fig. 6, there are certain genres where the exponent is below –1, and it would not be appropriate to ignore this as an error. In other words, applying the 1-parameter Yule-Simon process with $\theta = 0$ is not feasible, so a 2-parameter Yule-Simon process is applied. Assuming the minimum exponent is –1.4, θ becomes –0.4, which is a negative value. This suggests that over time, it becomes easier to attract other followers. Since it seems natural for followers to increase after content becomes widespread, a negative value for θ is not unreasonable.

Furthermore, regarding the increase in follower counts on Spotify, if we assume that θ does not change across genres, each exponent will represent the relationship between the growth of followers and the growth of artists. A larger exponent means that m is smaller, indicating that it becomes easier for artists to increase in number. Since the exponent represents the relative speed of growth, it does not necessarily mean that a smaller value indicates that the genre is more popular. However, from a subjective perspective, I tend to associate a smaller exponent with genres that have spread more widely.

D. Application

In In this study, we found that network access follows a power law, and that the exponential coefficient of the power law also falls within a certain range.

The Yule-Simon process is a simple model, and by applying this, it is expected that even if the network access genre is fixed, it will follow a power law characterized by one or two parameters. On the other hand, if it can be explained by the Yule-Simon process, the probability that a user will attract other users can be considered to be equal. As a result, while the total number of users increases exponentially, it is expected that the number of users at the top of the rankings can be roughly estimated.

On the other hand, in a competitive relationship, when aiming to change the ranking, it is possible to estimate how many users are required to raise the ranking. In other words, by estimating $-\alpha$ of the power regression for a specific ranking, it becomes possible to estimate that the number of users needs to be 10^{α} times higher if you want to reduce the ranking to 1/10. This may be useful for estimating the scale of marketing and advertising costs, etc.

V. Conclusions

In this study, we observed the ranking characteristics of network content services by analyzing the genre-specific ViewCount of Niconico and the genre-specific follower counts of Spotify. As a result, we found that the relationship between ranking r and follower count s generally follows a power-law of the form $s = Cr^{\alpha}$ for all genres. However, the value of α varied by genre. In Niconico, α ranged from –0.7 to –0.3, while in Spotify, it ranged from –1.4 to –0.4.

To explain the exponents below –1.0, we applied the two-parameter Yule-Simon process. This revealed that followers tend to increase as time passes, showing a growing trend. On the other hand, it seemed that the more popular genres had a larger value of α.
In the future, we aim to clarify the mechanism behind the apparent uniformity of the variability in the exponents observed across genres, and to better understand user behavior through more precise models.

Conflict of Interest

The authors declare no conflict of interest.

References

[1] M. E. J. Newman, "Power laws, Pareto distributions and Zipf's law," *Contemp. Phys.*, vol. 46, no. 5, pp. 323–351, 2005.
[2] Niconico. [Online]. Available: https://www.nicovideo.jp/
[3] Spotify. [Online]. Available: https://spotify.com/
[4] G. K. Zipf, "The unity of nature, least-action, and natural social science," *Sociometry*, vol. 5, no. 1, pp. 48–62, 1942.
[5] J. B. Estoup, *Les Gammes Stenographiques*, Institut Stenographique de France, 1916.
[6] U.S. Census Bureau, "Summary file 3, table pct019: Place of birth for the foreign-born population," Technical report, Washington, DC, 2000.
[7] L. Q. Ha, E. I. Sicilia-Garcia, J. Ming, and F. J. Smith, "Extension of Zipf's law to words and phrases," in *Proc. COLING*, 2002.
[8] D. H. Zanette, "Zipf's law and city sizes: A short tutorial review on multiplicative processes in urban growth," arXiv preprint, arXiv:0704.3170, 2007.
[9] Y.-S. Chen and F. F. Leimkuhler, "Analysis of Zipf's law: An index approach," *Inf. Process. Manag.*, vol. 23, no. 3, pp. 171–182, 1987.
[10] G. U. Yule, "A mathematical theory of evolution, based on the conclusions of Dr. J. C. Willis, F.R.S.," *Philos. Trans. R. Soc. Lond. B*, vol. 213, pp. 21–87, 1925.
[11] H. A. Simon, "On a class of skew distribution functions," *Biometrika*, vol. 42, no. 3–4, pp. 425–440, 1955.
[12] E. Baur and J. Bertoin, "On a two parameter Yule-Simon distribution," in *Proc. Springer Int. Publ.*, Cham, 2021, pp. 59–82.
[13] G. A. Miller, "Some effects of intermittent silence," *Am. J. Psychol.*, vol. 70, no. 2, pp. 311–314, 1957.
[14] W. Li, "Random texts exhibit Zipf's-law-like word frequency distribution," *IEEE Trans. Inf. Theory*, vol. 38, no. 6, pp. 1842–1845, 2006.
[15] N. Jan, L. Moseley, T. Ray, and D. Stauffer, "Is the fossil record indicative of a critical system?" *Adv. Complex Syst.*, vol. 2, no. 2, pp. 137–141, 1999.
[16] R. H. Swendsen and J.-S. Wang, "Nonuniversal critical dynamics in Monte Carlo simulations," *Phys. Rev. Lett.*, vol. 58, pp. 86–88, 1987.
[17] K. Sneppen, P. Bak, H. Flyvbjerg, and M. H. Jensen, "Evolution as a self-organized critical phenomenon," in *Proc. the National Academy of Sciences*, 1995, vol. 92, no. 11, pp. 5209–5213.
[18] P. J. Reynolds, H. E. Stanley, and W. Klein, "A real-space renormalization group for site and bond percolation," *J. Phys. C: Solid State Phys.*, vol. 10, no. 8, L167, 1977.
[19] P. Grassberger, "Critical behaviour of the Drossel-Schwabl forest fire model," *New J. Phys.*, vol. 4, no. 1, 17, 2002.
[20] A.-L. Barabási and R. Albert, "Emergence of scaling in random networks," *Science*, vol. 286, no. 5439, pp. 509–512, 1999.
[21] J. M. Carlson and J. Doyle, "Highly optimized tolerance: A mechanism for power laws in designed systems," *Phys. Rev. E*, vol. 60, no. 2, pp. 1412–1427, 1999.
[22] J. M. Carlson and J. Doyle, "Highly optimized tolerance: Robustness and design in complex systems," *Phys. Rev. Lett.*, vol. 84, pp. 2529–2532, 2000.
[23] K. Sneppen and M. E. J. Newman, "Coherent noise, scale invariance and intermittency in large systems," *Physica D*, vol. 110, no. 3, pp. 209–222, 1997.
[24] R. Cont and D. Sornette, "Convergent multiplicative processes repelled from zero: Power laws and truncated power laws," *J. Phys. I*, vol. 7, pp. 431–444, 1996.
[25] D. Sornette, "Multiplicative processes and power laws," *Phys. Rev. E*, vol. 57, pp. 4811–4813, 1998.
[26] X. Gabaix, "Zipf's law for cities: An explanation," *Q. J. Econ.*, vol. 114, no. 3, pp. 739–767, 1999.
[27] E. J. Gumbel, *Statistics of Extremes*. Columbia University Press, 1958.