Game Theory Analysis of SEO Strategies: From Methods to Models

1st Yuzhe Yang
School of Data Science
121090684@link.cuhk.edu.cn

2nd Xiayu Ni
School of Science and Engineering
121010047@link.cuhk.edu.cn

3rd RongXin Cao
School of Data Science
120090102@link.cuhk.edu.cn

***Abstract*—In the realm of Search Engine Optimization (SEO), enhancing a website's visibility is paramount. This paper focuses on developing methods to increase the importance score of a homepage, leveraging insights from Google's page ranking algorithm. The study compares several methods to determine their effectiveness and demonstrates the results through actual system implementation.**

The analyzed methods include optimizing content quality, improving page speed, and enhancing user engagement. By implementing these strategies, the goal is to boost the website's visibility in search engine results and increase organic traffic.

Through systematic comparison and evaluation, the paper aims to identify the most effective methods for increasing the homepage's importance score. The results of the study are supported by the implementation of these methods in a real-world system.

***Index Terms*—Search Engine Optimization, Game Theory**

I. Introduction

In the ever-evolving digital landscape, the visibility and importance of a website are paramount for its success. With the majority of online experiences beginning with a search engine query, understanding and optimizing for search engine algorithms is crucial. Among the various search engines, Google's page ranking algorithm stands out as one of the most influential and widely used methods for determining the relevance and importance of web pages.

This project aims to delve into the intricacies of Google's page ranking algorithm, exploring ways to enhance a homepage's importance score through the development and implementation of a customized optimization model. By conducting a comparative analysis of various methods, we seek to identify the most effective strategies for improving website ranking, providing valuable insights for web developers and businesses alike.

II. Objective

1) Development of an Optimization Model: Creation of a custom optimization model tailored to boost a homepage's importance score based on the identified factors.

2) Comparative Analysis: Evaluation of multiple optimization methods to determine their effectiveness in improving the importance score.

3) Implementation and Validation: Design and launch a homepage applying the developed model to validate the effectiveness of the chosen optimization strategies.

III. Literature Review

A. Review of Google Ranking Policy

According to our textbook Networked Life [2], we google matrices that denote the importance score in Google ranking policy: H, $\hat{H}$, and G.

$$G = \theta\hat{H} + (1-\theta)\frac{1}{N}11^T$$

Which indicates that given the initialization vector $\pi[0]$, the corresponding iterative procedure:

$$\pi^T[k] = \pi^T[k-1]G$$

Would converge as k goes to infinity. And we could thus calculate the Page Rank algorithm:

$$\pi^{*T} = \pi^{*T}G$$

B. SEO Strategy

SEO strategy involves the systematic planning and implementation of techniques to optimize a website's content, structure, and relevance, with the ultimate goal of improving its visibility in search engine results. This strategic approach focuses on organizing and presenting information in a way that aligns with search engine algorithms, thus increasing the likelihood of attracting organic traffic and enhancing overall online presence.

Consideration of Mobile SEO is crucial when formulating a comprehensive strategy [4]. Mobile optimization entails guaranteeing that website and its content are accessible to visitors on mobile devices. This ensures that they can enjoy an equivalent experience and derive the same value as users on desktop browsers.

An SEO strategist can specialize in two distinct types of SEO:

1) On-page SEO [5]: This approach centers on optimizing the content present on website pages to improve the site's ranking for specific keywords. It involves refining the content to align with search engine algorithms and enhance relevance.

2) Off-page SEO [5]: This strategy concentrates on building links from external sources to the website. The number and quality of backlinks from reputable sources play a crucial role in establishing trust with search algorithms, ultimately influencing rankings.

IV. Factors affect Page Rank

Before building our website and collecting data, we need to understand factors affecting SEO criteria. Tsuei et al. (2018) [1]summarised and split two main dimensions of different SEO criteria: internal and external website optimization. They also emphasized that internal website optimization requires attention to website design, meta tags, and keywords, which involves optimizing various elements such as page names, photos, links, and content texts on every page. It also extends to handling related texts and pages in different languages, while external website optimization involves connecting the website to the site guide, leveraging social media factors, and incorporating links from other optimized websites to the relevant webpage. This comprehensive approach helps enhance the overall performance and visibility of the website in the digital landscape (Yalçın and Köse 2010) [3]. Moreover, influence of the traffic plays a role in search result rankings. This is because search results that rank higher in position on Google are generally from websites with larger traffic.

TABLE I
Factors Affect Page Rank

Dimensions	Criteria
Internal	Keywords
	Meta Tags
	Website Design
External	Site Guide
	Social Media
	Linkage
	Traffic

To make our project workable, we considered the following factors:

- Keywords frequency;
- Traffic;
- Linkage (outlink and inlink).

V. New Google's Rank Algorithm

A. Mathematical Model

Keep the above factors in mind, we would create our own enhanced ranking model that includes the aforementioned four factors: keywords frequency, traffic, outgoing links, and ingoing links. As we know, Google's strategy from the textbook only considers about the number of outgoing links that bind to the website, and they are equally distributed. Thus, there is an improved way to evaluate the popularity of each link.

Firstly, let L_w^{in} as the weight of the link from page j to page i is determined by considering both the quantity of incoming links to page i from j and the total number of outgoing links to all reference pages of page j.

$$L_w^{in} = \frac{I_i}{\sum O_j}$$

Thus, the new model that relates linkage is given by:

$$\pi_{newi}^* = \pi_i^{*T} L_w^{in}$$

Where π^{*T} is the Pagerank calculated by Google matrix.

Next, we would like to consider the effect of the number of Keywords, we denote $\frac{N_k}{N_T}$ as the fraction, where N_k means the number of keywords that appear in the web page and N_T means the total word number appears at the web page. The ratio between then shows how frequently the key word appears and illustrate how relevant the website is, and then the model becomes as:

$$\pi_{newi}^* = \pi_i^{*T} L_w^{in} (1 + \frac{N_k}{N_T})$$

We collected the traffic per day of the above websites. At last, since the scale of the traffic in a website differs drastically, we would get the final model by adding the normalized traffic $\hat{T}$, where

$$\hat{T} = 0.1 + \frac{T - T_{min}}{T_{max} - T_{min}}$$

Here we add 0.1 means the minimum normalized traffic is 0.1.Then the final equation becomes:

$$\pi_{newi}^* = \pi_i^{*T} L_w^{in} (1 + \frac{N_k}{N_T}) + \hat{T}$$

B. Evaluation of the New Page Rank Algorithm

To evaluate the new Pagerank algorithm, We implemented both the improved algorithm and the standard Page Rank algorithms to assess and compare their outcomes.

Initially, we select the initial website www.nikon.com.cn for our testing website, we used the keyword term ”mirrorless camera” and generated a sitemap by web spider (Fig. 1). Using this sitemap, we computed both the standard PageRank (SPR) and our new version (NPR). The results, including the original PageRank, and the improved PageRank, are presented in Table 2.

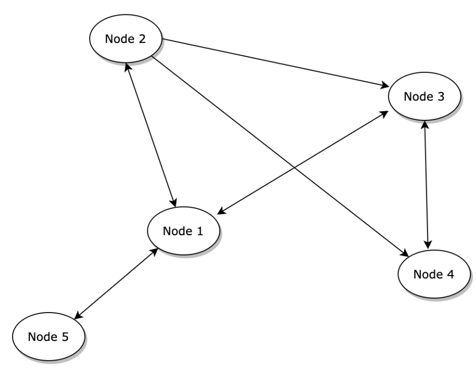

Fig. 1. Site Map

Through Table 2, it is easy to observe that the overall trend of PageRank generated by our algorithm is similar to the original one. For webpages 2 and 5, as well as webpages 3 and 4, the rankings remain the same in the original algorithm due to the equal number of outgoing links. However, with our algorithm's cumulative approach, we can achieve a more refined ranking between the same number of outlinks.

Furthermore, concerning web pages 2 and 4, we made an interesting discovery by comparing our model with the original

TABLE II
PAGE RANK PRODUCED STANDARD ALGORITHM (SPR), AND NEW ALGORITHM (NPR)

Index	SPR	Page rank number	L	N_k/N_T	$\hat{T}$	NPR	Pagerank number
1	1	0.340	0.5	1.042	1.1	1	1.277
2	4	0.102	0.25	1.032	0.322	3	0.348
3	2	0.228	0.33	1.030	0.306	2	0.383
4	2	0.228	0.33	1.023	0.179	4	0.256
5	4	0.102	0.25	1.023	0.1	5	0.126

one. In the original algorithm, web page 4 achieved the same ranking with web page 2. However, in our algorithm, webpage 2 has a higher ranking than web page 4. This is because, although web page 2 has only slightly fewer linkages than web page 4, the keyword density and traffic of web page 2 have seen a greater improvement. Therefore, the comprehensive score of web page 2 is higher than that of web page 4, which aligns with our intuition.

VI. Website Analysis

In our previous research, we conducted a comprehensive analysis to assess the diverse factors that influence PageRank, utilizing authentic data obtained from the online environment. Subsequently, we rigorously validated these findings within a real-world network context. By delving deeply into the meticulous examination of these factors, we are poised to implement strategic optimizations on our website, with the overarching goal of significantly enhancing its overall performance.

In Fig. 2, the depicted homepage serves as a meticulously crafted personal photography website. Its primary purpose is to share the creator's photography portfolio with the online community, thereby offering a repository of high-quality image resources for Internet users. Through this platform, users can access and enjoy a diverse collection of visually captivating photographs while also benefiting from convenient links to various social media platforms for enhanced engagement and interaction.

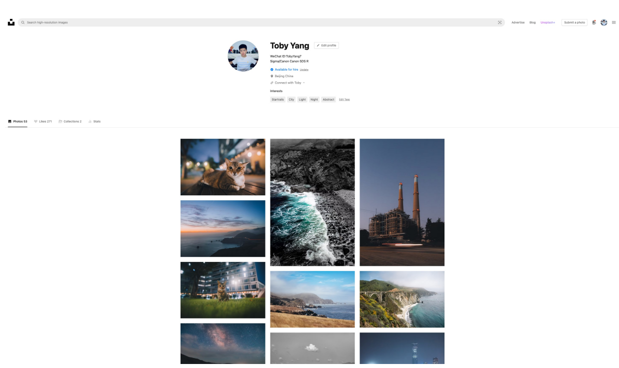

Fig. 2. Unsplash Homepage: https://unsplash.com/@tobyyang

In Fig. 3, we have provided a simplified representation of the overall structure of our website. Here, we can observe that our website is established beneath the primary domain name. Within this domain, various homepages are organized. Under each homepage, different pieces of content are neatly arranged. These content sections encompass essential elements such as keywords, meta tags, and the images we intend to showcase. It is worth noting that these elements are the portions that external websites may choose to link to.

Furthermore, we have also depicted several outbound links, which illustrate the interconnected nature of the internet, showing how various pages are linked to one another, creating a web of connections and relationships across the online landscape.

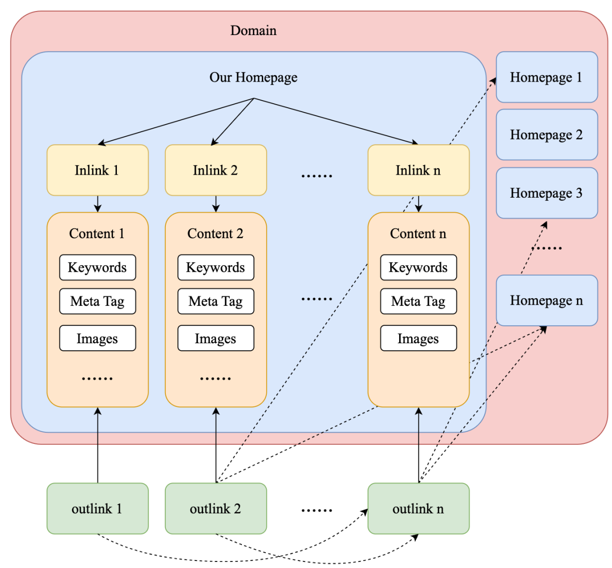

Fig. 3. The framework of our homepage

A. SEO Analysis

To attain a higher PageRank, we employed several strategies, including:

1) Selecting a Powerful Domain: In our pursuit of optimizing website performance and enhancing PageRank, we undertook a comprehensive analysis leveraging six months' worth of statistical data. Our focus was on quantifying the volume of devices accessing our website during this period, allowing us to gauge the extent of traffic our site attracts. Greater traffic signifies heightened exposure, thus affording our homepage greater potential for PageRank improvement.

Our analytical comparisons encompassed four analogous photography-themed websites. As elucidated in Fig. 4, the data unequivocally underscored Unsplash's preeminent standing in terms of user traffic within this cohort. It is this unequivocal dominance in user engagement that underpinned our deliberate choice to establish our homepage within the Unsplash domain. We discerned that aligning our website with a domain boasting such a formidable user base held the promise of substantially amplifying our website's visibility and PageRank prospects.

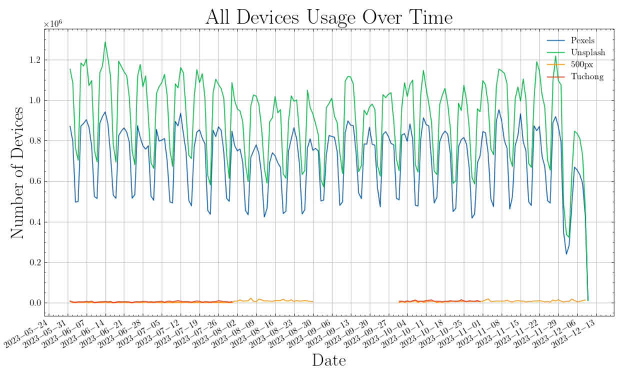

Fig. 4. Websites traffic data

2) Labeling Keywords: In the realm of Google search results, two pivotal metrics come to the forefront: significance and relevance. It is noteworthy that an increase in the quantity of keywords can contribute to heightened visibility and prominence of specific key terms. This augmentation not only enhances the importance of these keywords but also bolsters their contextual relevance within the search ecosystem. As a result, optimizing the presence of relevant keywords becomes an integral facet of an effective search engine strategy, aiming to align content with user queries and drive improved search rankings.

In our keyword optimization efforts, we leveraged our substantial content portfolio dedicated to Chongqing and capitalized on our existing traffic base. This strategic approach involved a deliberate expansion of keywords related to Chongqing, aimed at enhancing the relevance of our content and bolstering our ranking within Chongqing-related search queries. By aligning our content with the specific interests and queries of our target audience, we aimed to strengthen our online presence and cater to the needs of users seeking information about Chongqing.

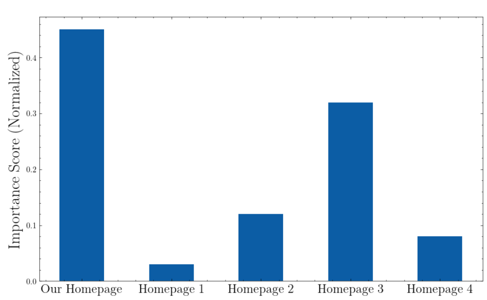

Fig. 5. Keywords Rank

In Fig. 5, we conducted concurrent assessments of the relevance of Chongqing keywords across four different homepages, including ours. Remarkably, our strategy yielded the highest level of relevance among all the evaluated homepages. This finding underscores the effectiveness of our approach in optimizing content for Chongqing-related keywords, further solidifying our position as a prominent source of information for users seeking content related to Chongqing. In Fig. 6, we showcase the results of a Google search, which reflects the impact of our strategy on our search engine visibility.

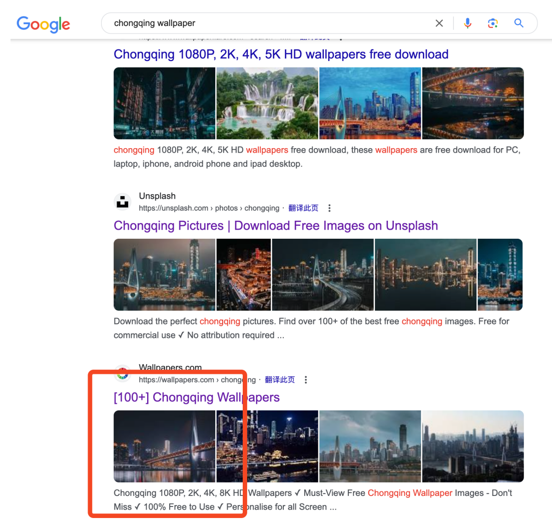

Fig. 6. Google Search Result

3) Building Quality Backlinks: In our prior analysis, we believed in the pivotal role of links in influencing PageRank. We fully recognized that backlinks play a critical role in determining a website's authority and credibility. Therefore, our objective was to establish a high-quality network of backlinks. Since we cannot directly control which high-quality external websites link to our homepage, our strategy involved disseminating our homepage across various high-traffic platforms, such as Instagram, Twitter, and other photography website platforms like Tuchong, 500px, and Pinterest. We utilized web crawling to gather backlink data, resulting in a total of 391 backlinks. The Fig. 7 illustrates the interconnected network of these links.

B. Historical Data Analysis

In a parallel effort, we utilized web crawling tools to amass comprehensive data regarding all the backlinks to our homepage throughout the year 2023. As depicted in Fig. 8, a discernible trend emerges, showing a consistent and upward trajectory in the accumulation of backlinks. This pattern is indicative of the efficacy of our SEO strategy, which has systematically bolstered our website's online presence and authority.

Of particular interest is the observation that around May 2023, there was a remarkable surge in the growth of backlinks. We postulate that this sudden upswing may be attributed to the emergence of a newly established photography website that engaged in the wholesale scraping of our website's content.

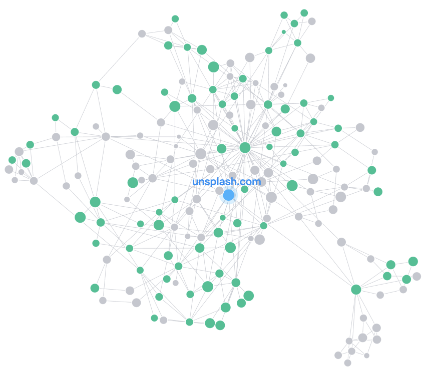

Fig. 7. Interconnected Network

This activity led to a rapid influx of backlinks from their site to ours.

From another perspective, this influx served to amplify the influence of our website within the digital landscape. It facilitated an expanded reach, encouraging more web pages to reference and cite our content, thereby organically generating new backlinks. This multifaceted development underscores the dynamic nature of our online presence and the far-reaching impact of our content within the digital ecosystem.

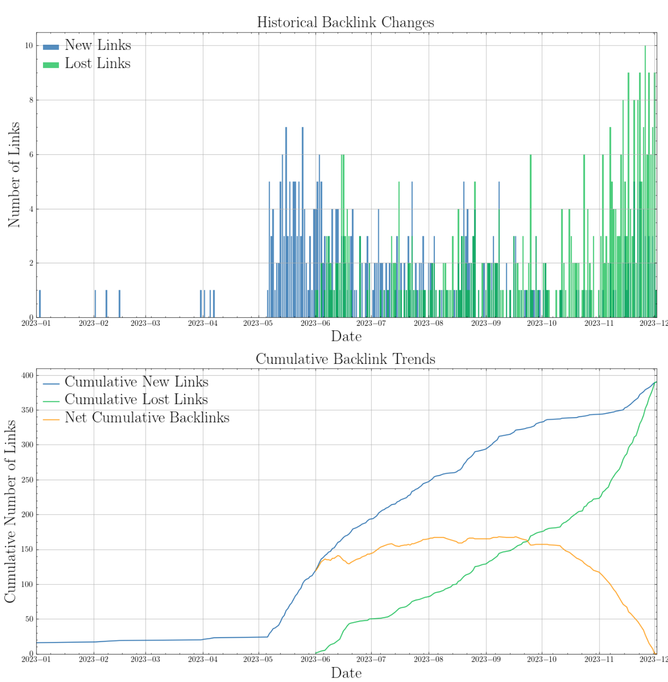

Fig. 8. Backlinks Historical Data

VII. Game Theory Model

In our ongoing research, we conducted a comprehensive analysis of our SEO strategy, examining it through the lens of a game theory model.

In the real-world context, SEO operates within a complex web of interactions. PageRank, being a dynamic metric, is significantly influenced by a myriad of external factors. Our website serves as an illustrative case in point: when other websites crawl and reference our content, they inherently contribute to the improvement of our PageRank. Conversely, if other websites remove links pointing to our site, it can result in a decrease in our PageRank. These interactions exemplify the dynamic nature of the online ecosystem.

Applying game theory principles to SEO facilitates our understanding and prediction of competitor behaviors, enabling us to formulate more effective strategies. This strategic insight equips us with the ability to make informed decisions that account for the moves and counter-moves of competitors within the ever-evolving online landscape.

A. α Random Walk PageRank Calculation

Consider a directed graph $G = (V, E)$ with a probability distribution q on the vertex set V and a damping factor α where $0 < \alpha < 1$. The PageRank of each vertex in G is determined by simulating an α-random walk, representing a balance between following direct links and random jumps to any vertex. This approach addresses the accessibility and rank distribution of web pages in the network, ensuring a comprehensive representation of their relative importance.

1) Start at a vertex v chosen according to the distribution q.
2) With probability $1-\alpha$, move to a random adjacent vertex.
3) With probability α, jump to a vertex chosen according to q.
4) The PageRank π is the stationary distribution of this random walk.

The potential ϕ_{uv} is the probability that an α random walk starting from vertex u visits vertex v before jumping. The PageRank π_v for a vertex v can be computed as follows:

$$\pi_v = \alpha \sum_{u \in V} \frac{q_u \phi_{uv}}{1 - (1-\alpha) \sum_{i \in \Gamma(v)} \phi_{iv}}$$

where $\Gamma(v)$ denotes the set of out-neighbors of v in the graph. Thus, we can calculate π^*_{newv} by other factors.

B. Nash Equilibrium

First we consider a strategy for link deletion. In the context of a vector x, where $x_i = 0$ corresponds to the deletion of link i, we can explore the implications of such a strategy.

Hence, the Pagerank of vertex v under strategy x is

$$\pi^{*T}_{newv}(\mathbf{x}) = \alpha \frac{\sum_{u \in V} q_u \phi_{uv}(\mathbf{x})}{1 - (1-\alpha) \frac{\sum_{i \in \Gamma(v)} \phi_{iv}(\mathbf{x}) x_i}{\mathbf{1}^T \mathbf{x}}} (L^{in}_w (1 + \frac{N_k}{N_T}) + \hat{T})$$

For simplicity (if we know vector $a \geq 0, b \geq 0$, and consider c is a constant), we rewrite the formula as:

$$\pi^{*T}_{newv} = \mathbf{1}^T \mathbf{x} \frac{a^T \mathbf{x}}{b^T \mathbf{x}} c$$

However, for a single page v, we need the maximize its Pagerank score to ensure that this particular page gains the highest possible visibility and authority within the context of the broader network. We solved for$\mathbf{1}^T\mathbf{x} = 1, 2, ...\Gamma(v)$ Hence we can write this problem as linear-fractional problem:

$$\max_{\mathbf{x}} \quad \frac{a^T\mathbf{x}}{b^T\mathbf{x}}c$$

and we can solve it by using Dinkelbach's method. We transformed the original equation as:

$$\max_{\mathbf{x}} \quad a^T\mathbf{x} - y \cdot b^T\mathbf{x}$$

Let $y[t]$ be the parameter at iteration t.It can be proved that convergence is guaranteed by alternatively updating y and solving for x because y is non-decreasing after each iteration. Hence we can solve this linear-fraction programing by updating:

$$y[t+1] = \frac{a^T\mathbf{x}[t]}{b^T\mathbf{x}[t]}$$

For Nash Equilibrium, we can write:

$$\pi^{*T}_{newv}(\mathbf{1}) \geq \pi^{*T}_{newv}(\mathbf{x})$$

In a Nash Equilibrium state for web pages, each page has honed its linking strategy to the extent that no single page can enhance its PageRank by altering its links in isolation. This signifies a highly competitive and stable environment where each page's position in search engine rankings is strategically balanced against others. The example illustrated in Fig. 1 from earlier serves as a concrete demonstration of such a Nash Equilibrium.

In practical SEO terms, this implies that the focus should shift towards other optimization aspects, such as enhancing content quality and improving the keywords frequency. Relying solely on link-based strategies would no longer yield significant improvements in search engine rankings. Any substantial shifts in the competitive landscape would typically occur only if multiple pages simultaneously change their strategies or if external factors introduce new dynamics into the equation.

C. Multi-phase Strategy

Due to factors like the dynamic nature of internet links, changing keyword trends, and other variables, the intensity of our strategy often needs to adapt in real-time. However, since SEO operates on a particularly lengthy cycle, we can only showcase this evolving process through simulated data. The results of these changes are depicted in Fig. 9.

D. Limitation

SEO is constrained by its extended timeframes, the intricate nature of the real online environment, limited access to competitor data, and the theoretical nature of decision models. Achieving tangible SEO results takes time and adaptation. Real-world SEO involves numerous complex variables, and obtaining competitor data can be challenging. Game theory-based models offer theoretical insights but may not perfectly mirror real-world complexities. Consequently, these models serve as strategic guides rather than definitive predictors, emphasizing the importance of a balanced approach in the dynamic SEO landscape.

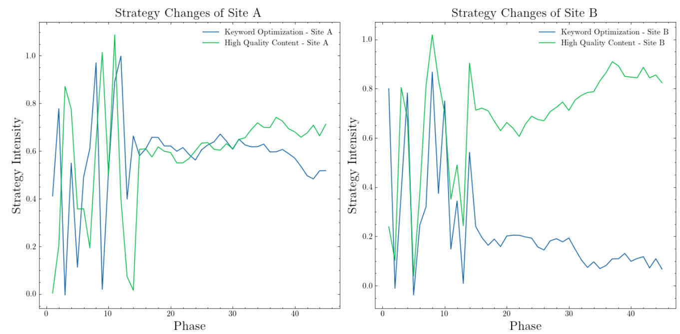

Fig. 9. Strategy Simulation

VIII. Contribution

A. Yuzhe Yang (Team Leader): 50%

Homepage construction, model construction, coding

B. Xiayu Ni: 50%

Introduction, literature review and new page rank algorithm building.

C. RongXin Cao: 0%

IX. Webpages Used

- www.nikon.com.cn
- www.nikon.com
- www.nikonimglib.com
- www.nikonimgsupport.com
- www.petapixel.com

Github Repo

https://github.com/TobyYang7/EIE3280_Project

References

[1] H.-J. Tsuei, W.-H. Tsai, F.-T. Pan, and G.-H. Tzeng, "Improving search engine optimization (SEO) by using hybrid modified MCDM models - Artificial Intelligence Review," SpringerLink, https://link.springer.com/article/10.1007/s10462-018-9644-0 (accessed Nov. 18, 2023).

[2] M. Chiang, "How does Google rank webpages?," in Networked Life: 20 Questions and Answers, Cambridge: Cambridge University Press, 2012, pp. 44–60.

[3] N. Yalçın and U. Köse, "What is search engine optimization: Seo?," Procedia - Social and Behavioral Sciences, vol. 9, pp. 487–493, 2010. doi:10.1016/j.sbspro.2010.12.185

[4] P. B. Godhwani, "A research paper on SEO aid in mobile website optimization," *International Journal of Management, IT and Engineering*, vol. 3, no. 6, pp. 282-288, 2013.

[5] V. M. Patil and A. V. Patil, "SEO: On-Page + Off-Page Analysis," in *2018 International Conference on Information, Communication, Engineering and Technology (ICICET)*, Pune, India, 2018, pp. 1-3, doi: 10.1109/ICICET.2018.8533836.