Convergence of Learning Dynamics in Information Retrieval Games

Omer Ben-Porat, Itay Rosenberg, Tennenholtz
Technion - Israel Institute of Technology
Haifa 32000 Israel
{omerbp@campus, itayrose@campus, moshet@ie}.technion.ac.il

Abstract

We consider a game-theoretic model of information retrieval with strategic authors. We examine two different utility schemes: authors who aim at maximizing exposure and authors who want to maximize active selection of their content (i.e., the number of clicks). We introduce the study of author learning dynamics in such contexts. We prove that under the probability ranking principle (PRP), which forms the basis of the current state-of-the-art ranking methods, any better-response learning dynamics converges to a pure Nash equilibrium. We also show that other ranking methods induce a strategic environment under which such a convergence may not occur.

1 Introduction

Information retrieval is probably the most central task carried out by consumers and users of on-line media. The basic information retrieval task involves ranking documents in a corpus by their relevance to the information needs expressed in a query. In adversarial retrieval settings such as the Web, information resources (contents) are owned by strategic bodies - website owners (henceforth *authors*). Authors can strategically change their content in order to improve their rankings in response to a query in a practice referred to as search engine optimization (SEO) (Gyöngyi and Garcia-Molina 2005). Therefore, the authors are players in a game, altering their content to increase their *utility*: increase exposure of their content (in a plain content setting) or to increase selection of their content ("clicks" in a sponsored content setting). In this strategic game, the search engine serves as a *mediator* between users and authors, and attempts to match queries and websites.

Despite the tremendous amount of work on information retrieval and SEO published during past decades, mathematical modeling of the aforementioned strategic behavior has only been formally suggested and studied recently (Ben-Basat, Tennenholtz, and Kurland 2015; Ben-Basat, Tennenholtz, and Kurland 2017; Raifer et al. 2017). One central question in this regard is whether learning dynamics, whereby at every step one author alters her content to increase her utility, is likely to converge. Convergence would suggest that authors should only invest a considerably limited amount of time altering their websites until their utility cannot be further improved. An accompanying question is whether such convergence occurs when state-of-the-art approaches to information retrieval, aiming at ranking documents in the corpus according to estimated relevance probabilities with respect to a given query, are used. The basis for all such retrieval methods is the probability ranking principle (PRP) (Robertson 1977).

In this paper we introduce what is, to the best of our knowledge, the first attempt to explore the learning dynamics of strategic behavior in information retrieval systems such as the Web, through a formal theoretical model. Our main result proves that under the PRP, any better-response learning dynamics converges to a pure Nash equilibrium. This result is obtained for the two prevalent utility schemes: authors seeking content exposure (i.e., exposure-targeted), and authors seeking to increase "clicks" in content selection (i.e., action-targeted). Interestingly, this learning dynamics convergence property, which rarely exists in games, is obtained even though our class of games are not potential games (Monderer and Shapley 1996). We also show that other plausible ranking methods may not induce such convergence, which further highlights the significance of our results.

1.1 Related Work

The concept of mediators in strategic environments is widely known to the game-theory community (Ashlagi, Monderer, and Tennenholtz 2009; Aumann 1974; Monderer and Tennenholtz 2009), and the design of a mediator (or in a different terminology, a mechanism) is often called mechanism design (Nisan and Ronen 1999). In the context of information retrieval, a search engine can be viewed as a mediator between two parties: users and authors.

Considering strategic behavior in an information retrieval context is the aim of Ben-Basat, Tennenholtz, and Kurland (2017). The work of Ben-Basat, Tennenholtz, and Kurland presents a game-theoretic approach to information retrieval, and illustrates that the myopic static view falls short in dynamic and adversarial settings. Ben-Basat, Tennenholtz, and Kurland explicitly assume that users will select the highest ranked result, a somewhat strong assumption but nevertheless justified by a large body of empirical work (Butman et al. 2013; Joachims et al. 2005; Liu and Wei 2016; Ghose, Goldfarb, and Han 2012). Note that in this case, PRP coincides with ranking the most relevant document highest. Ben-Basat, Tennenholtz, and Kurland analyze the user social welfare, defined as the quality of documents available in the presence of strategic behavior of the authors. Interestingly, they demonstrate that introducing randomization into a ranking function can sometimes lead to social welfare that transcends that of applying the PRP. In this paper we also adopt the game-theoretic approach to information retrieval, but explore a different criterion, which is the learning dynamics in games induced by the selection of the PRP as the mediator. Furthermore, beyond the action-targeted utility suggested in Ben-Basat, Tennenholtz, and Kurland, we also analyze exposure-targeted utility.

Ben-Porat and Tennenholtz (2018b) consider mediator design in recommendation systems with strategic content providers. They highlight several fairness-related properties that a mediator should arguably satisfy, along with the requirement of pure Nash equilibrium existence. They claim against PRP, as they show that in their mathematical model the PRP mediator (termed TOP in their work) satisfies the fairness-related properties, but may lead to a game without pure Nash equilibria and hence without better-response convergence. However, their mathematical model differs from the one in this paper, since e.g. they allow the mediator to present an empty list of documents, which is highly unlikely in information retrieval settings.

Designing a mediator for improved social welfare was recently proposed by Ben-Porat et al. (2019), who also make the connection between recommendation systems and facility location games (Hotelling 1929). In their model as well, matching users with their nearest facility may yield a low social welfare in case the content providers are strategic. Their goal is to design a mediator that optimizes welfare in equilibrium and does not intervene too much.

In this work, however, we do not study the social welfare, but rather focus on the *learning dynamics*. Learning dynamics is an important concept in machine learning and game theory (Cesa-Bianchi and Lugosi 2006; Claus and Boutilier 1998; Freund and Schapire 1999; Palaiopanos, Panageas, and Piliouras 2017; Syrgkanis et al. 2015; Meir et al. 2010; Lev and Rosenschein 2012), and work on learning dynamics in games is considered instrumental, e.g., to understanding ad auctions (Cary et al. 2014). Better-response learning dynamics are appealing to the (algorithmic) game theory community, as they only assume a minimal form of rationality: under any given profile, a player will act to increase her individual utility. However, general techniques for showing better-response learning convergence in games are rare, and are based typically on coming up with a potential function (Monderer and Shapley 1996), see e.g. (Garg and Jaakkola 2016; Palaiopanos, Panageas, and Piliouras 2017; Ben-Porat and Tennenholtz 2018a). However, as exact potential functions imply the games are congestion games (Rosenthal 1973), it is easy to observe that our games do not fit that category.

Another interesting class of games which are not potential games for which better-response dynamics always converge is (Milchtaich 1996). However, that setting is quite remote from ours, as in Milchtaich's work the players share a common set of strategies.

1.2 Our Contribution

Our main conceptual contribution is the explicit analysis of learning dynamics in information retrieval systems that is motivated by strategic behavior. Our demonstration of convergence serves as an important justification for the use of the PRP, and should be taken into account when designing stable and robust information retrieval systems.

The key technical contribution of this paper is the proof that under PRP any better-response dynamics converges to a pure Nash equilibrium. We prove this claim for both exposure-targeted and action-targeted utility schemes. As stated above, the convergence of better-response learning dynamics in our setting is obtained although the class of games we consider do not have an exact potential function. Moreover, we show that other ranking methods induce a strategic environment under which such convergence may not occur. Together, our results provide strong novel game-theoretic justification to the PRP and illustrate its applicability in an adversarial context such as the Web.

1.3 Paper Organization

The rest of the paper is organized as follows. Section 2 formalizes the model we adopt, as well as an informal introduction to the relevant core game-theoretic concepts and an illustrative example. In Section 3 we analyze better-response learning with the PRP mediator for both utility schemes. In Section 4 we show non-learnability of mediators other than the PRP, and Section 5 is devoted to discussion and future work. Due to space limitations, some of the proofs of this paper are deferred to the supplementary material.

2 Problem Statement

An authors game is composed of a set of *authors* $N = [n] \overset{\text{def}}{=} \{1, 2, \dots, n\}$, each owning one document/website/blog. $M = [m]$ is the set of *topics*, and we assume both n and m are finite. An author's pure strategy space is the set of all topics, i.e., she can choose to write her document on any topic. We further assume that each document is concerned with a single topic. The set of all pure strategy profiles is denoted by $A = M^n$, and each strategy profile $\boldsymbol{a} = (a_1, \dots a_n)$ corresponds to a set of documents. A query distribution D over M is publicly known, where each query symbolizes the user mass associated with that topic. Given a topic k, we denote by $D(k)$ the demand for topic k. We further assume w.l.o.g. that $D(1) \geq D(2) \geq \dots \geq D(m)$. That is, the topics are sorted according to the query distribution mass in a non-increasing order.

The matrix $Q \in [0, 1]^{n \times m}$ is the *quality matrix*, where $Q_{j,k}$ represents the quality for author j's document if she decides to write on topic k. This modeling allows an author to have remarkable aptitude for one topic and poor aptitude for another. For example, an economic guru is able to write about sports, but his writing quality w.r.t. sports is substantially lower than economics.

The function R is the *mediator*, which plays the role of a ranking function or a search engine. The mediator ranks the documents selected by the authors w.r.t. a given query (or equivalently, a topic). We assume for simplicity that users always read the document ranked first. This assumption is consistent with many applications, e.g. the use of personal assistants in mobile devices, where only the first ranked item is shown to the user. Thus, we let $R(Q, k, \boldsymbol{a})$ denote a distribution over the set of documents selected under $\boldsymbol{a}$ w.r.t. a topic $k \in M$, which represents the probability of being displayed in the first position. For ease of notation, we shall also denote $R_j(Q, k, \boldsymbol{a})$ as the probability that author j is ranked first under the distribution $R(Q, k, \boldsymbol{a})$.

The last component u is the *utility function*, which maps every strategy profile to a real-valued vector of length n. In this paper, we consider two different utility functions which are motivated by current applications.

Under the exposure-targeted utility, denoted by u^{Ex} , an author's utility is the number of impressions her document receives. Formally,

Definition 1 (Exposure-targeted utility)**.** *The exposure-targeted utility of author j under a strategy profile $\boldsymbol{a}$ is given by*

$$u_j^{Ex}(\boldsymbol{a}) \overset{\text{def}}{=} \sum_{k=1}^{m} \mathbb{1}_{a_j=k} \cdot D(k) \cdot R_j(Q, k, \boldsymbol{a}).$$

Note that u^{Ex} depends solely on the user mass of the topic she writes on and the probability of the mediator displaying her document. The other utility function is the action-targeted utility, denoted by u^{Ac}.

Definition 2 (Action-targeted utility)**.** *The action-targeted utility of author j under a strategy profile $\boldsymbol{a}$ is given by*

$$u_j^{Ac}(\boldsymbol{a}) \overset{\text{def}}{=} \sum_{k=1}^{m} \mathbb{1}_{a_j=k} \cdot D(k) \cdot R_j(Q, k, \boldsymbol{a}) \cdot Q_{j,k}.$$

Namely, an author's utility is the user mass of her selected topic times the probability she is ranked first times the quality of her document.

Overall, an authors game can be represented as a tuple $\mathcal{G} = \langle N, M, D, Q, R, u \rangle$.

It is convenient to quantify the following; given a strategy profile $\boldsymbol{a}$, let $B_k(\boldsymbol{a})$ denote the highest quality of a document on topic k , i.e.,

$$B_k(\boldsymbol{a}) \overset{\text{def}}{=} \max_{1 \le j \le n} \{Q_{j,k} \cdot \mathbb{1}_{a_j=k}\}.$$

Moreover, we denote by $H_k(\boldsymbol{a})$ the number of authors whose documents have the highest quality among those who write on topic k under $\boldsymbol{a}$,

$$H_k(\boldsymbol{a}) \overset{\text{def}}{=} |\{j \mid j \in [n], Q_{j,k} \cdot \mathbb{1}_{a_j=k} = B_k(\boldsymbol{a})\}|.$$

Unless stated otherwise, we analyze games with a particular mediator, which is based on the PRP. Since we restrict the ranking list to include one rank only, the PRP coincides with ranking first the highest quality document on that topic. We denote by R^{PRP} the mediator that displays the document with the highest quality. In case there are several documents with the highest quality, R^{PRP} ranks first each one of them with equal probability. Formally,

Definition 3 (The PRP Mediator)**.** *Given a quality matrix Q, a topic k and a strategy profile $\boldsymbol{a}$, the R^{PRP} ranks first the document of each author j with a probability of*

$$R_j^{PRP}(Q, k, \boldsymbol{a}) \overset{\text{def}}{=} \begin{cases} \frac{1}{H_k(\boldsymbol{a})} & Q_{j,k} \cdot \mathbb{1}_{a_j=k} = B_k(\boldsymbol{a}) \\ 0 & \textit{otherwise} \end{cases}.$$

2.1 Further Game Theory Notation

We now informally introduce some basic game theory concepts used throughout this paper. For an action profile $\boldsymbol{a} = (a_1, \dots, a_j, , \dots, a_n) \in A$, we denote by $\boldsymbol{a}_{-j} = (a_1, \dots, a_{j-1}, a_{j+1}, \dots, a_n) \in A_{-j}$ the action profile of all authors except author j. A strategy $a'_j \in A_j$ is called a *better response* of author j w.r.t. a strategy profile $\boldsymbol{a}$ if $u_j(a'_j, \boldsymbol{a}_{-j}) > u_j(\boldsymbol{a})$. Similarly, $a'_j \in A_j$ is said to be a *best response* if $u_j(a'_j, \boldsymbol{a}_{-j}) \geq \max_{a_j \in A_j} u_j(a_j, \boldsymbol{a}_{-j})$. We say that a strategy profile $\boldsymbol{a}$ is a *pure Nash equilibrium* (herein denoted PNE) if every author plays a best response under $\boldsymbol{a}$.

Given a strategy profile $\boldsymbol{a} \in A$, an *improvement step* is a profile $(a'_j, \boldsymbol{a}_{-j})$ such that a'_j is a better response of author j w.r.t. $\boldsymbol{a}$. An *improvement path* $\gamma = (\boldsymbol{a}^1, \boldsymbol{a}^2, \dots)$ is a sequence of improvement steps, where the improvements can be performed by different authors. Namely, in any improvement step along the improvement path exactly one author deviates from the strategy she selected in the previous step, but different authors can deviate in different steps. When the path γ is clear from the context, we denote by p_r the author that improves in step r. Since the number of strategy profiles is finite, every infinite improvement path must contain an improvement cycle. A non-cooperative game $\mathcal{G}$ has the *finite improvement property* (FIP for brevity) if all the improvement paths are finite; in such a game every better-response dynamics converges to a PNE (Monderer and Shapley 1996).

2.2 An Illustrative Example

To further clarify our notation and setting, we provide the following example. Consider a game with $n = 2$ authors, $m = 3$ topics, a query distribution mass D such that $D(1) = 0.5, D(2) = 0.3, D(3) = 0.2$, a quality matrix

$$Q = \begin{pmatrix} 0.1 & 0.4 & 0.8 \\ 0.9 & 0.4 & 0.2 \end{pmatrix},$$

and R^{PRP} as the mediator. Given the utility function, the induced game can be viewed as a normal form bi-matrix game, as presented in Figure 1.

First, consider the exposure-targeted utility function. Consider the strategy profile $(a_1, a_2) = (2, 2)$. Under this strategy profile the two authors write on topic 2, and their quality on that topic is the same, i.e., $Q_{1,2} = Q_{2,2} = 0.4$; thus, $R_1(Q, 2, (2, 2)) = R_2(Q, 2, (2, 2)) = 0.5$ and

$$u_1^{Ex}(2, 2) = u_2^{Ex}(2, 2) = \frac{D(2)}{2} = 0.15.$$

Notice that author 2 can improve her utility by deviating to topic 1, i.e., to the strategy profile $(2, 1)$. Indeed, this is an improvement step w.r.t. $(2, 2)$. In this case, her utility is $u_2^{Ex}(2, 1) = 0.5$. Clearly $(2, 1)$ is a PNE of this game.

$$\begin{array}{cc} & \begin{array}{ccc}\text{topic 1} & \text{topic 2} & \text{topic 3}\end{array} \\ \begin{array}{c}\text{topic 1}\\ \text{topic 2}\\ \text{topic 3}\end{array} & \left[\begin{array}{ccc} 0, 0.5 & 0.5, 0.3 & 0.5, 0.2 \\ 0.3, 0.5 & 0.15, 0.15 & 0.3, 0.2 \\ 0.2, 0.5 & 0.2, 0.3 & 0.2, 0 \end{array}\right] \end{array}, \quad \begin{array}{cc} & \begin{array}{ccc}\text{topic 1} & \text{topic 2} & \text{topic 3}\end{array} \\ \begin{array}{c}\text{topic 1}\\ \text{topic 2}\\ \text{topic 3}\end{array} & \left[\begin{array}{ccc} 0, 0.45 & 0.05, 0.12 & 0.05, 0.04 \\ 0.12, 0.45 & 0.06, 0.06 & 0.12, 0.04 \\ 0.16, 0.45 & 0.16, 0.12 & 0.16, 0 \end{array}\right] \end{array}$$

(a) exposure-targeted (b) action-targeted

Figure 1: The normal form games induced by the example in Subsection 2.2. Subfigure (a) represents the utilities of author 1 (row) and author 2 (column) under the exposure-targeted utility function, while Subfigure (b) represents the utilities under action-targeted utility function.

The action-targeted utility function induces a different bimatrix game. The reader can verify that under this utility scheme, the unique PNE is $(3, 1)$.

3 Better-Response Learning with the PRP Mediator

In this section we show that under the PRP mediator, every better-response dynamics converges to a PNE, for both utility schemes. To make this claim more concrete, we use the following definition.

Definition 4. *We say that a mediator R is u-learnable if every game induced by R and the utility function u has the FIP property.*

Clearly, if any game that consists of (R, u) has the FIP property, then the authors can learn a PNE using any better-response dynamics. We use the above definition to crystallize our goals for this section: we wish to show that R^{PRP} is both u^{Ex}-learnable and u^{Ac}-learnable. Namely, in Subsection 3.1 we show that under the PRP mediator and the exposure-targeted utility function, every improvement path is finite. In Subsection 3.2 we prove the equivalent statement for the action-targeted utility function.

Before we go on, we claim that the class of games induced by the PRP mediator does not have an exact potential.

Proposition 1. *The class of games induced by R^{PRP} and either one of u^{Ex} or u^{Ac} does not have an exact potential.*

Proof sketch of Proposition 1. We show that the necessary condition for the existence of an exact potential (Monderer and Shapley 1996) does not hold for a general authors game with $n \geq 3$ authors. This result is obtained for both utility schemes. □

As mentioned in Section 1 above, showing the convergence of any better-response dynamics in the lack of exact potential is challenging, and is nevertheless our goal for the rest of this section. In light of that, we shall introduce a further notation.

Definition 5. *Given a finite improvement path $\gamma = (\boldsymbol{a}^1, \dots \boldsymbol{a}^l)$, we define*

$$W_k(\gamma) \overset{\text{def}}{=} \min_{1 \leq r \leq l} \{H_k(\boldsymbol{a}^r)\},$$

i.e., $W_k(\gamma)$ is the minimal number of authors writing documents with the highest quality on topic k.

Note that the minimum is taken over all steps in γ.

3.1 Exposure-Targeted Utility

We now focus on games with R^{PRP} and u^{Ex}, namely the PRP mediator and the exposure-targeted utility function. We show that every improvement path is finite, suggesting that any better-response dynamics converges. The proof of this convergence relies on several supporting claims.

The following Proposition 2 claims that in every improvement step, the improving author writes with a quality of at least the highest quality obtained in the preceding improvement step, on that particular topic.

Proposition 2. *Let γ be a finite improvement path, and let $a_{p_r}^{r+1} = k$ for an arbitrary improvement step r. It holds that $Q_{p_r,k} \geq B_k(\boldsymbol{a}^r)$.*

We now bound the utility the improving author obtains in the corresponding improvement step, when her document's quality does not exceed the highest quality (on that particular topic) in the preceding improvement step.

Proposition 3. *Let γ be a finite improvement path, and let $a_{p_r}^{r+1} = k$ for an arbitrary improvement step r. If $Q_{p_r,k} \leq B_k(\boldsymbol{a}^r)$, then*

$$u_{p_r}^{Ex}(\boldsymbol{a}^{r+1}) \leq \frac{D(k)}{W_k(\gamma) + 1}.$$

Next, we characterize a property that must hold in improvement cycles, under the false assumption that such exist. We prove that if an improvement cycle exists, the quality of the first-ranked document is constant throughout the improvement cycle; this must hold for every topic.

Lemma 1. *If $c = (\boldsymbol{a}^1, \dots, \boldsymbol{a}^l = \boldsymbol{a}^1)$ is an improvement cycle, then for every improvement step r and every topic k it holds that $B_k(\boldsymbol{a}^r) = B_k(\boldsymbol{a}^{r+1})$.*

Proof sketch. We give here a high-level overview of the proof and refer the reader to the appendix for the formal proof.

Under the false assumption that an improvement cycle exists, assume that the claim does not hold. Namely, assume that $c = (\boldsymbol{a}^1, \dots, \boldsymbol{a}^l = \boldsymbol{a}^1)$ is an improvement cycle (w.l.o.g. c is a simple improvement cycle), and that there exist an improvement step r and a topic k such that $B_k(\boldsymbol{a}^r) \neq B_k(\boldsymbol{a}^{r+1})$.

Recall that $D(1) \geq \cdots \geq D(m)$, i.e., the topics are sorted according to the query distribution mass in a non-increasing order. We prove by induction on the topic index k

that $B_k(\boldsymbol{a}^r) \le B_k(\boldsymbol{a}^{r+1})$ holds for every r, $1 \le r \le l-1$. Clearly, if this holds for every improvement step r then

$$B_k(\boldsymbol{a}^1) \le \cdots \le B_k(\boldsymbol{a}^l) = B_k(\boldsymbol{a}^1);$$

thus, all inequalities hold in equality and $B_k(\boldsymbol{a}^r) \ne B_k(\boldsymbol{a}^{r+1})$ cannot occur.

Base, $k = 1$: Assume the assertion does not hold for $k = 1$; hence, there exists r, $1 \le r \le l-1$, such that $B_1(\boldsymbol{a}^r) > B_1(\boldsymbol{a}^{r+1})$. This means that there exists an author who writes with the highest quality on topic 1 in the step r, and then she deviates to another topic in step $r+1$. Moreover, due to the strict inequality, that author is the unique author to write with the highest quality on topic 1 in step r; hence, her utility in step r is exactly $D(1)$. When she deviates, she can obtain at most $D(2)$, but recall that $D(1) \ge D(2)$; hence, this deviation is not beneficial.

Step: Assume the assertion holds for $k \in \{1, 2, \dots K-1\}$, i.e., $B_k(\boldsymbol{a}^r) = B_k(\boldsymbol{a}^{r+1})$ for every step r. We show that $B_K(\boldsymbol{a}^r) > B_K(\boldsymbol{a}^{r+1})$ for a step r implies that the improving author in improvement step r deviates to a topic with a lower index. Using the bound obtained in Proposition 3 and the induction hypothesis, we show that there must be an improving author which does not increase her utility after preforming the deviation, which is clearly a contradiction. □

Lemma 1 implies that the only element that varies throughout an improvement cycle, if such exists, is the number of authors who write on each topic. In particular, the highest quality on each topic remains constant. It also suggests that any improving author is not the only author writing the highest quality document on the topic to which she deviated.

Consider an arbitrary improvement step, and denote by k the topic that the improving author writes on in the improvement step. The improving author joins a (non-empty) set of authors which are already writing documents with the highest quality on topic k. Since we deal with a cycle, at some point an author abandons topic k, and deviates to another topic, say k'. In Lemma 2 we bound the utility of the improving author (deviating to topic k) with that of the author who deviated to k'.

Lemma 2. *If $c = (\boldsymbol{a}^1, \dots, \boldsymbol{a}^l = \boldsymbol{a}^1)$ is an improvement cycle, then for every improvement step r and topic k such that $a_{p_r}^{r+1} = k$ there exist (r', k') such that $a_{p_{r'}}^{r'+1} = k'$ and*

$$\frac{D(k)}{W_k(c)+1} < \frac{D(k')}{W_{k'}(c)+1}.$$

Proof sketch. Let r, k be such that $a_{p_r}^{r+1} = k$. By definition of improvement step $a_{p_r}^r \ne k$. From Lemma 1 we know that $B_k(\boldsymbol{a}^r) = B_k(\boldsymbol{a}^{r+1})$; thus, $Q_{p_r,k} = B_k(\boldsymbol{a}^r)$ and $H_k(\boldsymbol{a}^r) \ne H_k(\boldsymbol{a}^{r+1})$. Afterwards, we prove another claim which guarantees that there exists r' such that

$$\frac{D(k)}{W_k(c)+1} = u_{p_{r'}}^{Ex}(\boldsymbol{a}^{r'})$$

holds. In addition, $p_{r'}$ is the improving author, and so

$$\frac{D(k)}{W_k(c)+1} = u_{p_{r'}}^{Ex}(\boldsymbol{a}^{r'}) < u_{p_{r'}}^{Ex}(\boldsymbol{a}^{r'+1}). \tag{1}$$

Clearly, $a_{p_{r'}}^{r'+1} = k' \ne k$. Lemma 1 indicates that $B_{k'}(\boldsymbol{a}^{r'}) = B_{k'}(\boldsymbol{a}^{r'+1})$; hence, $Q_{p_{r'},k'} \le B_{k'}(\boldsymbol{a}^{r'})$. Having showed that the condition of Proposition 3 holds, we invoke it for r', k' and conclude that

$$u_{p_{r'}}^{Ex}(\boldsymbol{a}^{r'+1}) \le \frac{D(k')}{W_{k'}(c)+1}.$$

Combining this fact with Equation (1), we get

$$\frac{D(k)}{W_k(c)+1} < \frac{D(k')}{W_{k'}(c)+1}.$$

□

In Theorem 1 below we leverage Lemma 2 to show that improvement cycles cannot exist.

Theorem 1. *R^{PRP} is u^{Ex}-learnable.*

Proof of Theorem 1. To show that R^{PRP} is u^{Ex}-learnable it suffices to show that every improvement path is finite. Moreover, every improvement path cannot contain more than a finite number of different strategy profiles, as m and n are finite; therefore, if γ is infinite it must contain an improvement cycle. We are left to prove that γ cannot contain an improvement cycle.

Assume by contradiction that γ contains an improvement cycle $c = (\boldsymbol{a}^1, \boldsymbol{a}^2, \dots, \boldsymbol{a}^l = \boldsymbol{a}^1)$. Let r_1 be an arbitrary improvement step and denote by k_1 the topic such that $a_{p_{r_1}}^{r_1+1} = k_1$. From Lemma 2 we know that there exist (r_2, k_2) such that $a_{pr_2}^{r_2+1} = k_2$ and

$$\frac{D(k_1)}{W_{k_1}(c)+1} < \frac{D(k_2)}{W_{k_2}(c)+1}.$$

Since $a_{p_{r_2}}^{r_2+1} = k_2$, we can now use Lemma 2 again in order to find (r_3, k_3) such that $a_{p_{r_3}}^{r_3+1} = k_3$ and

$$\frac{D(k_2)}{W_{k_2}(c)+1} < \frac{D(k_3)}{W_{k_3}(c)+1}.$$

This process can be extended to achieve additional $k_4, k_5, \dots, k_{m+1}$ such that

$$\frac{D(k_1)}{W_{k_1}(c)+1} < \frac{D(k_2)}{W_{k_2}(c)+1} < \dots < \frac{D(k_{m+1})}{W_{k_{m+1}}(c)+1}.$$

Since there are only m topics and that the inequality above contains $m+1$ elements, there are at least two elements which are identical; thus we obtain a contradiction. We deduce that an improvement cycle cannot exist. □

Theorem 1 concludes the analysis of the exposure-targeted utility function.

3.2 Action-Targeted Utility

After analyzing games with exposure-targeted utility, we proceed to action-targeted utility. The main result of this subsection is that R^{PRP} is u^{Ac}-learnable, which is analogous to the main result of the previous one. Interestingly, achieving this result requires a more subtle treatment. To motivate it, consider the following: under u^{Ex}, in a case where the quality of an author's document on topic k exceeds the quality of all other authors writing on topic k, she will not deviate to a topic with a higher index (a topic with a lower or equal user mass). This, however, is not true for u^{Ac}. For instance, consider the strategy profile $(2,1)$ in the example given in Subsection 2.2. Under u^{Ex}, author 1 cannot increase her utility by deviating to topic 3 (a topic with a lower user mass). In contrast, under u^{Ac}, author 1 *can* improve her utility by deviating to topic 3. To assist in that, let $S_k(\gamma)$ denote the highest quality of a document written on topic k throughout a finite improvement path γ. Formally,

Definition 6. *Given a topic k and an improvement path $\gamma = (\boldsymbol{a}^1, \dots, \boldsymbol{a}^l)$,*

$$S_k(\gamma) \overset{\text{def}}{=} \max_{1 \le r \le l} \{B_k(\boldsymbol{a}^r)\}.$$

In Proposition 4 we bound the utility of an improving author in an improvement step.

Proposition 4. *Let γ be a finite improvement path, and let $a_{p_r}^{r+1} = k$ for an arbitrary improvement step r. If $Q_{p_r,k} \le B_k(\boldsymbol{a}^r)$, then*

$$u_{p_r}^{Ac}(\boldsymbol{a}^{r+1}) \le \frac{D(k) \cdot S_k(\gamma)}{W_k(\gamma) + 1}.$$

Notice that $S_k(\gamma) \le 1$ for every k and every γ; thus, the bound given in Proposition 3 trivially holds for u^{Ac}. However, proving this tighter bound becomes essential for refuting the existence of improvement cycles under u^{Ac}. By proving additional supporting lemmas (which are further elaborated in the appendix), we show that

Theorem 2. *R^{PRP} is u^{Ac}-learnable.*

4 Non-Learnability under Other Mediators

In the previous section we showed a powerful result: R^{PRP} is both u^{Ex}-learnable and u^{Ac}-learnable. In other words, when using R^{PRP}, any better-response dynamics converges; this is true for both utility schemes. In fact, R^{PRP} is not the only mediator under which such convergence occurs. For instance, Let R^{RAND} be the *random* mediator, such that for any author j and any topic k,

$$R_j^{RAND}(Q, k, \boldsymbol{a}) \overset{\text{def}}{=} \begin{cases} \frac{1}{\sum_{i=1}^{n} \mathbb{1}_{a_i = k}} & a_j = k \\ 0 & \text{otherwise} \end{cases}.$$

By showing that under u^{Ex} any game with R^{RAND} can be reduced to a game with R^{PRP}, we conclude that

Proposition 5. *R^{RAND} is u^{Ex}-learnable.*

Proof sketch. We prove the claim by showing that under u^{Ex} any game with R^{RAND} can be reduced to a game with R^{PRP}, such that the two games are strategically equivalent. This is done by taking any game G with R^{RAND} as the mediator and a quality matrix Q, and reduce it to a game G' with R^{PRP} as the mediator and Q' as the quality matrix, such that $Q'_{j,k} = 1$ for every $j \in N$ and $k \in M$.

Since both G, G' consists of the exposure-targeted utility function, we omit the super-script Ex and use the super-script G to specify the utility of author j under the strategy profile $\boldsymbol{a}$ in G, i.e., $u_j^G(\boldsymbol{a})$, and equivalently $u_j^{G'}(\boldsymbol{a})$ for G'. By definition of exposure-targeted utility and R^{PRP}, for every valid j and $\boldsymbol{a}$ it holds that

$$\begin{aligned} u_j^{G'}(\boldsymbol{a}) &= \sum_{k=1}^{m} \mathbb{1}_{a_j = k} \cdot D(k) \cdot R_j^{PRP}(Q', k, \boldsymbol{a}) \\ &= D(a_j) \cdot R_j^{PRP}(Q', a_j, \boldsymbol{a}) \\ &= D(a_j) \cdot \frac{1}{H_{a_j}(\boldsymbol{a})} \\ &= D(a_j) \cdot R_j^{RAND}(Q, a_j, \boldsymbol{a}) \\ &= \sum_{k=1}^{m} \mathbb{1}_{a_j = k} \cdot D(k) \cdot R_j^{RAND}(Q, k, \boldsymbol{a}) = u_j^G(\boldsymbol{a}). \end{aligned}$$

Since G' possesses R^{PRP} as the mediator, Theorem 1 guarantees that G' has the FIP property. Since we showed G and G' are strategically equivalent, G also has the FIP property, and in particular does not contain improvement cycles. □

Notice that R^{RAND} treats every document the same, regardless of its quality. However, in many (and perhaps even most) scenarios mediators seek to promote high-quality content. Therefore, the reader may wonder whether other plausible mediators are u^{Ex}-learnable or u^{Ac}-learnable. We now focus on a wide and intuitive family of mediators, which we term scoring mediators.

Definition 7. *Let R be a mediator. We say that R is a scoring mediator if there exists a non-decreasing function $f : \mathbb{R} \to \mathbb{R}_+$ such that for every $Q, k, \boldsymbol{a}$ and author index j it holds that*

$$R_j(Q, k, \boldsymbol{a}) \overset{\text{def}}{=} \begin{cases} \frac{f(Q_{j,k})}{\sum_{i=1}^{n} \mathbb{1}_{a_i = k} \cdot f(Q_{i,k})} & a_j = k \\ 0 & \textit{otherwise} \end{cases}.$$

It this case, we denote $R = R^f$ for the corresponding f.

Under a scoring mediator every author receives a probability according to the proportion of her score over the sum of the scores of all author writing on that topic. Notice that if R^f is a scoring mediator such that the corresponding f is constant, we get $R^f = R^{RAND}$. In addition, this family also includes celebrated mediators, e.g. the softmax function (for $f(Q_{j,k}) = e^{Q_{j,k}}$), which is very popular in machine learning applications, or the linear function (for $f(Q_{j,k}) = Q_{j,k}$) that is common in probabilistic models for decision making (for instance, in the Bradley—Terry model (Bradley and Terry 1952)). Noticeably, the R^{PRP} is not a scoring mediator. In the rest of this section, we show non-convergence of better-response dynamics for general families of scoring mediators.

4.1 Exposure-Targeted Utility

In this subsection we prove that, under mild assumptions, scoring mediators are not u^{Ex}-learnable (as opposed to R^{RAND}). We restrict ourselves to mediators for which the corresponding function f is continuous, and exhibits the following property: the ratio between the score of the highest quality and the lowest quality is greater than two (note that this property holds trivially if the score of the lowest quality is zero, i.e., $f(0) = 0$). Among others, this class of mediators contains mediators based on softmax and linear functions, as described above.

Theorem 3. *Let R^f be a scoring mediator. If f is a continuous function such that $f(1) > 2f(0)$, then R^f is not u^{Ex}-learnable.*

Proof sketch. It is sufficient to show that for every f that satisfies the theorem's conditions, we can construct a game instance with an improvement cycle. We exploit the properties of f to construct a game with four authors and three topics, and show that an improvement cycle exists. Let R^f be a scoring mediator with the corresponding function f, which we assume exhibits $f(1) > 2f(0)$. Due to the Intermediate Value Theorem, there exist x_1, x_2, x_3 such that $0 < x_3 < x_2 < x_1 \leq 1$ and

$$\frac{f(x_2)}{f(x_3)} > \frac{2f(x_1)}{f(x_2)} > 2.$$

For brevity, denote $c_1 = \frac{f(x_1)}{f(x_2)}$ and $c_2 = \frac{f(x_2)}{f(x_3)}$, and observe that $c_2 > 2c_1$. Consider a game with $|N| = 4$ authors, $|M| = 3$ topics and a quality matrix Q such that

$$\begin{pmatrix} x_1 & 0 & 0 \\ x_1 & x_2 & 0 \\ x_2 & 0 & x_3 \\ 0 & x_3 & x_2 \end{pmatrix}.$$

The only missing ingredient is the distribution D over the topics. The selection of such D is crucial: we shall select D to allow improvement cycles. Denote

$$D(1) = \frac{1}{2-3\epsilon}, D(2) = \frac{1-2\epsilon}{2(2-3\epsilon)}, D(3) = \frac{1-4\epsilon}{2(2-3\epsilon)},$$

for some $0 < \epsilon \leq \frac{1}{4}$. It can be verified that D is a valid distribution over the set of topics. Consider the strategy profiles

$$\boldsymbol{a}^1 = (1,1,1,2), \quad \boldsymbol{a}^2 = (1,2,1,2), \quad \boldsymbol{a}^3 = (1,2,3,2),$$
$$\boldsymbol{a}^4 = (1,2,3,3), \quad \boldsymbol{a}^5 = (1,1,3,3), \quad \boldsymbol{a}^6 = (1,1,1,3).$$

In the rest of the proof we show that ϵ can be selected such that the cycle $c = (\boldsymbol{a}^1, \boldsymbol{a}^2, \boldsymbol{a}^3, \boldsymbol{a}^4, \boldsymbol{a}^5, \boldsymbol{a}^6, \boldsymbol{a}^7 = \boldsymbol{a}^1)$ is an improvement cycle of the game we constructed. More precisely, we prove that for every r, $1 \leq r \leq 6$, $u^{Ex}_{p_r}(\boldsymbol{a}^r) < u^{Ex}_{p_r}(\boldsymbol{a}^{r+1})$. This suggests that R^f is not u^{Ex}-learnable. □

While all it takes to prove Theorem 3 is to show a *single* game instance with an improvement cycle, we can actually construct infinitely many games which do not possess FIP. Moreover, our construction can be viewed as a sub-game in a much broader game, i.e., with more authors and topics.

4.2 Action-Targeted Utility

When analyzing scoring mediators, an additional difference between the two utility schemes emerges. In the improvement cycle constructed in the proof of Theorem 3, there exists an improvement step in which the improving author decreases the quality of her document but still increases her utility. Under the action-targeted utility function, such a decrease may not be translated to improved utility. Namely, the technique employed in Theorem 3 for constructing a game that possesses an improvement cycle might not work here. Nevertheless, the following theorem shows non-learnability under u^{Ac} of scoring mediators that boost high-quality content. For example, a mediator R^f where the corresponding f satisfies $f(1) > 6f(\frac{1}{2})$ assigns a substantially higher score to the highest quality than a mediocre one.

Theorem 4. *Let R^f be a scoring mediator. If f is a continuous function such that $f(1) > 2(2\alpha - 1)f\left(\frac{1}{\alpha}\right)$ for some $\alpha > 1$, then R^f is not u^{Ac}-learnable.*

Notice the resemblance between the condition of Theorem 3 to that of Theorem 4. Due to space limitations, additional results on the non-learnability of other scoring mediators under u^{Ac} are omitted and further elaborated in the appendix.

5 Discussion

We introduced the study of learning dynamics in the context of information retrieval games. Our results address learning in the framework introduced by Ben-Basat, Tennenholtz, and Kurland (2017), where authors are action-targeted as well as for a complementary type of information retrieval game in which the authors' aim is to maximize their exposure. In particular, our results show that a mediator that operates according to the PRP (Robertson 1977) induces a game in which learning-dynamics converges; the latter is true for both exposure-targeted and action-targeted utility schemes. Moreover, we have also demonstrated that this convergence is a virtue of the PRP, and does not apply for other relevant mediators.

One prominent question is the time required for the authors to converge, namely, finding the worst-case length of an improvement path. It turns out that there is a class of games where the length of the best-response paths is easy to analyze.

Consider the exposure target utility, and assume that D is strictly decreasing, the number of authors equals the number of topics, and that the matrix Q is generic, i.e., has $n \times m$ distinct values. The induced game exhibits a unique equilibrium: topic 1 is assigned to the author with the highest quality w.r.t. topic 1. Topic 2 is assigned to the author with the highest quality on that topic, from the set of authors who were not assigned before. Clearly, the PNE is computed by following this process until every author/topic is assigned. Consequently, any best-response dynamics where the authors play in a round-robin fashion will converge after at most a quadratic number of improvement steps in the number of authors. A similar observation applies to action targeted utility under a slightly different notion of generality of Q. The general question of convergence rate is nevertheless left open.

Our model, as any other novel model that pretends to explain theoretical aspects of real-world systems, has its limitations. To name a few, we assume the set of authors and topics are fixed, while in reality they are often dynamic; we assume that the quality of documents is perfectly observed by the mediator, which only approximates modern search engines. Although not ultimate, we do believe that our modeling, which extends a model that is already acknowledged as valuable (Ben-Basat, Tennenholtz, and Kurland 2017), serves as an important justification for the use of the PRP, and may be an important step for future work to circumvent the limitations presented above. We note that our learning dynamics is based on applying an author's response to the current behavior of other authors. In fact, this assumes that the only information available to the author is the quality of the documents currently published, and assumes nothing about information available to an author on other authors' (unobserved) qualities. Relaxing the assumption that published documents' qualities can be observed goes beyond the scope of our work, and may be a subject for future research.

An interesting future direction is to expand the information retrieval setting to a setup where each author's document may include several topics. This issue is treated in a preliminary manner in (Ben-Basat, Tennenholtz, and Kurland 2017) and it may be of interest to see whether our results can be extended to that context as well. It may be also interesting to study the quality of the equilibrium (as far as users' social welfare is concerned) reached under PRP. Would the best equilibrium be obtained under better-response learning dynamics?

Acknowledgments

This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement n° 740435).

References

Ashlagi, I.; Monderer, D.; and Tennenholtz, M. 2009. Mediators in position auctions. *Games and Economic Behavior* 67(1):2–21.

Aumann, R. 1974. Subjectivity and correlation in randomized strategies. *Journal of Mathematical Economics* 1:67–96.

Ben-Basat, R.; Tennenholtz, M.; and Kurland, O. 2015. The probability ranking principle is not optimal in adversarial retrieval settings. In *Proceedings of ICTIR*, 51–60.

Ben-Basat, R. B.; Tennenholtz, M.; and Kurland, O. 2017. A game theoretic analysis of the adversarial retrieval setting. *Journal of Artificial Intelligence Research* 60:1127–1164.

Ben-Porat, O., and Tennenholtz, M. 2018a. Competing prediction algorithms. *arXiv preprint arXiv:1806.01703*.

Ben-Porat, O., and Tennenholtz, M. 2018b. A game-theoretic approach to recommendation systems with strategic content providers. In *Advances in Neural Information Processing Systems (NIPS) 2018*.

Ben-Porat, O.; Goren, G.; Rosenberg, I.; and Tennenholtz, M. 2019. From recommendation systems to facility location games. In *Proceedings of the Thirty-Third National Conference on Artificial Intelligence (AAAI 2019)*.

Bradley, R. A., and Terry, M. E. 1952. Rank analysis of incomplete block designs: I. the method of paired comparisons. *Biometrika* 39(3/4):324–345.

Butman, O.; Shtok, A.; Kurland, O.; and Carmel, D. 2013. Query-performance prediction using minimal relevance feedback. In *Proceedings of ICTIR*, 7.

Cary, M.; Das, A.; Edelman, B.; Giotis, I.; Heimerl, K.; Karlin, A. R.; Kominers, S. D.; Mathieu, C.; and Schwarz, M. 2014. Convergence of position auctions under myopic best-response dynamics. *ACM Transactions on Economics and Computation* 2(3):9.

Cesa-Bianchi, N., and Lugosi, G. 2006. *Prediction, learning, and games*. Cambridge Univ Press.

Claus, C., and Boutilier, C. 1998. The dynamics of reinforcement learning in cooperative multiagent systems. *AAAI/IAAI* 1998:746–752.

Freund, Y., and Schapire, R. E. 1999. Adaptive game playing using multiplicative weights. *Games and Economic Behavior* 29:79–103.

Garg, V., and Jaakkola, T. 2016. Learning tree structured potential games. In *Advances In Neural Information Processing Systems*, 1552–1560.

Ghose, A.; Goldfarb, A.; and Han, S. P. 2012. How is the mobile internet different? search costs and local activities. *Information Systems Research* 24(3):613–631.

Gyöngyi, Z., and Garcia-Molina, H. 2005. Web spam taxonomy. In *Proceedings of AIRWeb*, 39–47.

Hotelling, H. 1929. Stability in competition. In the Economic Journal 39 (153): 4157, 1929.

Joachims, T.; Granka, L. A.; Pan, B.; Hembrooke, H.; and Gay, G. 2005. Accurately interpreting clickthrough data as implicit feedback. In *Proceedings of SIGIR*, 154–161.

Lev, O., and Rosenschein, J. S. 2012. Convergence of iterative voting. In *Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems-Volume 2*, 611–618. International Foundation for Autonomous Agents and Multiagent Systems.

Liu, C., and Wei, Y. 2016. The impacts of time constraint on users' search strategy during search process. *Proceedings of the Association for Information Science and Technology* 53(1).

Meir, R.; Polukarov, M.; Rosenschein, J. S.; and Jennings, N. R. 2010. Convergence to equilibria in plurality voting. In *AAAI*, volume 10, 823–828.

Milchtaich, I. 1996. Congestion games with player-specific payoff functions. *Games and Economic Behavior* 13.

Monderer, D., and Shapley, L. 1996. Potential games. *Games and Economic Behavior* 14:124–143.

Monderer, D., and Tennenholtz, M. 2009. Strong mediated equilibrium. *Artif. Intell.* 173(1):180–195.

Nisan, N., and Ronen, A. 1999. Algorithmic mechanism design. Proceedings of STOC-99.

Palaiopanos, G.; Panageas, I.; and Piliouras, G. 2017. Multiplicative weights update with constant step-size in congestion games: Convergence, limit cycles and chaos. In *Advances in Neural Information Processing Systems*, 5874–5884.

Raifer, N.; Raiber, F.; Tennenholtz, M.; and Kurland, O. 2017. Information retrieval meets game theory: The ranking competition between documents' authors. In *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 465–474. ACM.

Robertson, S. E. 1977. The probability ranking principle in IR. *Journal of Documentation* 33:294–304.

Rosenthal, R. W. 1973. A class of games possessing pure-strategy nash equilibria. *International Journal of Game Theory* 2(1):65–67.

Syrgkanis, V.; Agarwal, A.; Luo, H.; and Schapire, R. E. 2015. Fast convergence of regularized learning in games. In *Advances in Neural Information Processing Systems*, 2989–2997.