Journal of Artificial Intelligence Research 60 (2017) 1127-1164 Submitted 4/17; published 12/17

A Game Theoretic Analysis of the Adversarial Retrieval Setting

Ran Ben Basat SRAN@CS.TECHNION.AC.IL
Faculty of Computer Science, Technion,
Haifa 32000003, Israel

Moshe Tennenholtz MOSHET@IE.TECHNION.AC.IL
Oren Kurland KURLAND@IE.TECHNION.AC.IL
Faculty of Industrial Engineering and Management, Technion,
Haifa 32000003, Israel

Abstract

The main goal of search engines is ad hoc retrieval: ranking documents in a corpus by their relevance to the information need expressed by a query. The Probability Ranking Principle (PRP) — ranking the documents by their relevance probabilities — is the theoretical foundation of most existing ad hoc document retrieval methods. A key observation that motivates our work is that the PRP does not account for potential post-ranking effects; specifically, changes to documents that result from a given ranking. Yet, in adversarial retrieval settings such as the Web, authors may consistently try to promote their documents in rankings by changing them. We prove that, indeed, the PRP can be sub-optimal in adversarial retrieval settings. We do so by presenting a novel game theoretic analysis of the adversarial setting. The analysis is performed for different types of documents (single-topic and multi-topic) and is based on different assumptions about the writing qualities of documents' authors. We show that in some cases, introducing randomization into the document ranking function yields an overall user utility that transcends that of applying the PRP.

1. Introduction

The core task that search engines address is ad hoc document retrieval: ranking documents in a corpus in response to a query by their relevance to the information need the query expresses. Numerous retrieval methods and frameworks have been devised throughout the years; e.g., the vector space model (Salton, Wong, & Yang, 1975), the probabilistic approach, (Sparck Jones, Walker, & Robertson, 2000), the language modeling framework (Ponte & Croft, 1998), the divergence from randomness framework (Amati & van Rijsbergen, 2002), and learning-to-rank approaches (Liu, 2009).

Probabilistic retrieval methods are based on estimating the probability that a document is relevant to a query (Sparck Jones et al., 2000), or more precisely, satisfies the information need underlying the query. The theoretical basis for all probabilistic retrieval methods, and under one interpretation or another, for all retrieval frameworks mentioned above, is the *Probability Ranking Principle* (PRP) (Robertson, 1977). According to the PRP, documents in the corpus should be ranked by the probability that they are relevant to the query so as to maximize the utility of the user who posted the query; relevance probabilities are assumed to be estimated using all information available to the search system (Robertson, 1977)[1]. For example, a relevance estimate can be devised based on the query-document textual match, query-independent properties of the document (e.g., its content breadth, see Bendersky, Croft, & Diao, 2011) and other indicators.

1. There is also a definition of the PRP for interactive retrieval (Fuhr, 2008). Interactive retrieval is outside the scope of this paper.

There are two main scenarios where the PRP may result in sub-optimal utility. The first is when users have different utility functions (i.e., personalization effects) (Robertson, 1977). The second scenario is when the relevance of one document depends on that of another; i.e., diversification effects (Robertson, 1977; Carbonell & Goldstein, 1998).[2]

However, there is an additional fundamental aspect of the PRP that has not been accounted for in past work. The PRP, and retrieval methods devised based on the PRP, are based on a static view of the ad hoc retrieval setting. That is, the goal is to optimize user utility for the current query given the current *fixed* snapshot of the corpus[3]. The optimization is performed *regardless* of the incentives of the authors who created the documents. Thus, it does not account for potential post-retrieval effects — e.g., changes of the corpus, and therefore rankings for future queries, as a result of the ranking induced for the current query. More generally, an underlying implicit premise of the standard approach of designing ranking functions is that optimizing user utility along time can be based on providing the best possible relevance ranking for each query without accounting for future effects. In this paper, we show that this static view falls short in dynamic and adversarial settings.

In adversarial retrieval settings such as the Web, documents might be changed by their authors, henceforth *publishers*, so as to have them promoted in rankings induced in response to queries. This practice is referred to as search engine optimization (Gyöngyi & Garcia-Molina, 2005). Thus, the assumption that the ranking induced for one query will not affect the potential utility attained for future queries is simply wrong; that is, each induced ranking can lead to changes in the corpus upon which future search will be performed. Naturally, these changes can affect the utility attained for future queries. To illustrate this point, consider the following example. Suppose that a Web page is composed of a few sections, some of which contain valuable information which is unique to the page. In terms of page visibility, the publisher of the Web page is mainly concerned about how the page is ranked for queries that match the non-unique sections of the page. Having the page ranked low in response to these queries might lead the publisher to remove some of the unique sections to better emphasize those which should "attract" the queries of concern. Such an action can lead to diminishing the utility of users looking for the unique information.

We present a novel game theoretic analysis of the adversarial ad hoc retrieval setting. The analysis accounts for the following facts (i) document publishers are "players" with *incentives*; namely, having their documents ranked high in response to some queries; (ii) the search engine is a *mediator* which affects the actions of publishers by the rankings it induces in response to queries; and therefore, (iii) the utility attained for the current query can affect that attained for future queries by the virtue of post-retrieval effects on the corpus (i.e., changes to documents). Our analysis assumes that all users have the same utility function (i.e., we do not account for personalization effects) and that the relevance of one document is independent of that of others (i.e., we do not account for the importance of diversifying search results but rather focus on their relevance). In that respect, our analysis is committed to the same conditions under which the PRP was shown to be optimal for a specific query and a static corpus.

A particular advantage of taking a game theoretic approach to modeling the adversarial retrieval setting is the ability to make insightful statements about the steady state of the dynamics of the setting, namely, an equilibrium. The players — publishers in our case — who act selfishly, do not have an incentive to deviate from the equilibrium. Thus, analyzing properties of the equilibrium is important for understanding the implications of a specific (dynamic) game. We note that any change to the actions that the players can take — in our case, the types of documents they can produce — and to the ranking function which serves as the mediator induces a new game with potentially new equilibria.

Inspired by standard practice in work on (algorithmic) game theory, our treatment of the adversarial retrieval setting as a game focuses on the notion of *social welfare*: the sum of utilities

2. A probability ranking principle using quantum probability principles was proposed to address inter-document relations (Zuccon & Azzopardi, 2010).
3. In some cases, information about previous snapshots of the corpus might also be utilized (Elsas & Dumais, 2010).

provided to users by rankings produced in response to their queries. Indeed, search engines should opt to maximize the social welfare of their users. More specifically, we analyze the worst possible social welfare attained in *any* equilibrium with respect to the best possible social welfare that can be attained. This concept, a.k.a. *price of anarchy* (PoA)(Koutsoupias & Papadimitriou, 1999; Roughgarden & Tardos, 2002), is an important tool for analyzing games. In our case, the PoA is a means to contrasting different types of ranking functions in terms of the worst possible social welfare they can lead to in a steady state of the game.

We build up our analysis by starting from the case of having documents that discuss a single topic. We show that applying the PRP in this case, on a per-query basis, results in optimal social welfare if publishers are assumed to have the same writing quality for all topics. Later, we explore the more realistic scenario of publishers having differential writing qualities. We show that applying the PRP in this setting is sub-optimal. As it turns out, the PRP is also sub-optimal in the case of having multi-topic documents and publishers with equal writing qualities for all topics. Interestingly, in several cases of our analysis, we show that introducing *randomization* into a ranking function can sometimes lead to social welfare that transcends that of applying the PRP.

A major portion of our analysis of the games focuses on a setting where some m topics can be queried — and in the multi-topic document setting, written about — and 2 publishers compete.[4] We note that, indeed, in many realistic settings, two major publishers (or publishers' networks) compete over a set of topics. In a few cases in the analysis, we consider the case of 2 publishers and 2 topics, mainly as it is enough in these cases for proving the sub-optimality of the PRP.

We believe that the sub-optimality of the PRP, and more generally, the sub-optimality entailed by the static view taken by current retrieval methods that are based on the PRP, have important implications. For example, our theoretical findings imply that learning ranking functions by optimizing rankings for a train set of queries and a static corpus, as is the current state-of-affairs (Liu, 2009), is a sub-optimal practice. That is, the learning should also account for post-retrieval effects in terms of potential changes applied to documents in the corpus by incentivized document authors.

The task just mentioned, of devising ranking functions that optimize social welfare, is essentially a *mediator design* challenge as we explain next. Mechanism design (Mas-Colell, Whinston, & Green, 1995) deals with the design of game rules such that when agents play a solution of the game (e.g., some equilibrium) the desired behavior will be implemented. Work in computer science under the title of algorithmic mechanism design (Nisan & Ronen, 2001) had a special focus on dealing in incentive-compatible manner with the tractable implementation of such mechanisms. An interesting variation of mechanism design, titled *mediator design*, relates to situations in which rather than designing games/interactions, the designer has only the power to reliably modify existing interactions in order to lead participants into desired behaviors. Mediators can be viewed as entities that aim at directing decision-makers' behavior in a given game. In our setting, the mediator (i.e., the ranking function) directs the behavior (changes to documents) of decision makers (publishers) by inducing rankings. As noted above, our work establishes the motivation for, and importance of, a new research task: designing mediators (ranking functions) that optimize social welfare for the ad hoc retrieval setting.

The two main contributions of this paper are as follows[5]:

4. The notable exceptions are in (i) parts of the analysis for the most realistic setting of publishers with differential writing qualities; and (ii) all the connections we make to *congestion games* (Rosenthal, 1973; Monderer & Shapley, 1996). In these cases, n publishers and m topics are assumed.
5. This paper extends our initial conference paper (Ben-Basat, Tennenholtz, & Kurland, 2015). The main additions with respect to the conference paper are: (i) making connections to congestion games in Section 4.1.2, theorems 2, 6, and 10, and Lemma 1; (ii) providing a proof for the case of $c = 1/2$ in Theorem 7 that was only proven for $c < 1/2$ in the conference version; (iii) providing the proof for Theorem 8 which was Theorem 6 in the conference version; (iv) Theorem 13 which improves the bound stated in Theorem 10 in the conference version from 1.7 to 1.683; in addition, in contrast to the conference version, a full proof is provided; (v) A proof of Theorem 14 for a general m (which was Theorem 10 in the conference paper with a proof sketch for $m = n = 2$) with an improved bound of 1.683 rather than 1.7; (vi) Theorem 15 which extends the Theorem 11 from the conference paper by using general n rather than $n = 2$.

- Presenting a novel game theoretic analysis of the adversarial ad hoc retrieval setting. The analysis accounts for the incentives of document authors to have their documents ranked high in response to queries.

- Showing that the standard practice of inducing a ranking based on the Probability Ranking Principle — i.e., optimizing relevance ranking for a given query and static corpus — can be sub-optimal in terms of the overall utility attained by users of the search engine in the face of changes that the corpus goes through.

2. Related Work

There are two lines of work that pertain to ours which we discuss next: adversarial search and mediator design.

2.1 Adversarial Search

The vast majority of published work on adversarial information retrieval has focused on spam classification (mainly content-based and link-based spam), and more generally, on estimating the quality and/or authoritativeness of Web pages (e.g., see Fetterly, Manasse, & Najork, 2004; Gyöngyi & Garcia-Molina, 2005; Benczúr, Csalogány, Sarlós, & Uher, 2005; Wu & Davison, 2005; Krishnan & Raj, 2006; Cormack, 2007; Jones, Hawking, & Sankaranarayana, 2007, AIRWEB, 2009; Cormack, Smucker, & Clarke, 2010; Bendersky et al., 2011, WIC, 2012). There has also been some recent work on estimating for which queries content manipulation of documents is more likely (Raiber, Collins-Thompson, & Kurland, 2013) and on estimating global term statistics in adversarial peer-to-peer search systems (Richardson & Cox, 2014). Work on adversarial classification treats the classifier and the data generator as two adversaries (Dalvi, Domingos, Mausam, Sanghai, & Verma, 2004; El-Yaniv & Nisenson, 2010). A game theoretic approach was also used to optimize learning-to-rank functions in a standard non-adversarial retrieval setting (Wang, Xing, Asif, & Ziebart, 2015).

In contrast to all this prior work, we present a game theoretic modeling of the adversarial search (i.e., ad hoc retrieval) setting wherein publishers are incentivized to have their documents promoted in rankings. Our analysis does not depend on the nature of actions employed by publishers, namely, whether these are black hat (e.g., spamming), grey hat or white hat search engine optimization efforts (Gyöngyi & Garcia-Molina, 2005). Furthermore, showing that the Probability Ranking Principle is sub-optimal in adversarial settings is a novel contribution of the work presented here.

There has been some (none game theoretic) work on modeling the past versions of a Web page — independently of those of other pages — so as to improve retrieval effectiveness (Elsas & Dumais, 2010). In contrast, we provide a game theoretic analysis of the effect (specifically, in a steady state) of having multiple publishers change the content of their documents throughout time in response to multiple queries.

There has recently been much interest in studying the dynamic aspects of interactive retrieval where the user has on-going interaction with the search system through which the queries and ranking function might change (e.g., see Fuhr, 2008; Sloan & Wang, 2012; Zhang & Zhai, 2015; Yang et al., 2016). Changes of clickthrough patterns have also been explored (Sloan & Wang, 2012). In contrast to this line of work, we focus on the dynamics of documents in the corpus as a result of induced rankings.

Studying and predicting the dynamics of the Web corpus, specifically, for improving crawling policies and personalizing content delivery, was the focus of some studies (Radinsky & Bennett, 2013; Santos, Pasini, & Freire, 2016). The focus of our work is very different: studying the dynamics with respect to the "ranking competition" between publishers of documents.

There is recent work on devising relevance-based dueling strategies for search engines (Izsak, Raiber, Kurland, & Tennenholtz, 2014). Specifically, two search engines compete with each other assuming static publishers and documents. In contrast to our work, analysis of game theoretic aspects (e.g., equilibrium and social welfare) was not presented. Furthermore, our focus is on the competition between dynamic publishers (and documents) given that *some* search engine is used.

In work (Ben-Basat & Kravi, 2016) that follows the initial version of this paper (Ben-Basat et al., 2015), the merits of non-deterministic ranking functions similar in spirit to those we theoretically advocate here, were demonstrated using a *simulation* of a ranking competition between publishers who stuff query terms in their documents (Ben-Basat & Kravi, 2016). No game theoretical analysis as that we present here was performed.

Our focus in this paper is on analyzing the content dynamics of a corpus driven by incentivized publishers. Addressing other types of dynamics — e.g., hyperlink dynamics that results from the incentive to improve PageRank scores (cf. Hopcroft & Sheldon, 2008) — is left for future work.

2.2 Mediators

As noted above, in the game setting we address in this paper, the ranking function plays the role of a mediator in the retrieval setting.

In his seminal work, Aumann (Aumann, 1974) presented a form of mediation called *Correlated Equilibrium* (see also Ashlagi, Monderer, & Tennenholtz, 2008). In this setting, a mediator is allowed to send signals to each player. The signal is designed to help the players achieve higher payoffs.

Early work on social laws for artificial agent societies (Shoham & Tennenholtz, 1995) deals with mediators that can constrain participants' behavior.

Monderer and Tennenholtz (2004) address mediators who can promise positive payments. A different type of mediator is that which is allowed to change the players' payoffs (Monderer & Tennenholtz, 2004). The goal of the mediator is then to "force" players into playing some desired subset of their potential strategies with minimal interference with the payoff function.

Additional work on mediation includes mediation with incomplete information (Ashlagi, Monderer, & Tennenholtz, 2009), where players may only have partial knowledge about the world; strong mediated-equilibria (Monderer & Tennenholtz, 2006) where no *set* of players has incentive to deviate from the equilibrium; and, routing games (Rozenfeld & Tennenholtz, 2007).

The type of mediator we consider here, the search engine's ranking function, is different than those described above. For example, the mediator only affects payoffs by the virtue of the ranking it induces. Furthermore, the mediator in our case does not send explicit signals to the players. The only information the players are provided with is the query distribution (over queries) used in the search engine. Indeed, for some commercial search engines, some information about the most frequent queries is available[6].

There has been game-theoretic work on addressing ambiguous queries, which target a specific result, using a diversified *set* of retrieved results (Eliaz & Spiegler, 2016). The goal was to devise a ranking mechanism (mediator) that maximizes social welfare given that the user uses random sequential search to traverse the set of retrieved results. In contrast to our work, an adversarial retrieval setting where incentivized publishers change their documents so as to have them ranked high was not explored. Furthermore, the Probability Ranking Principle and the novel non-deterministic ranking approaches we propose, as well as the various search settings we explore, were not studied in this work (Eliaz & Spiegler, 2016). We note that addressing ambiguous queries and diversifying search results is outside the scope of this paper.

Balancing relevance and monetization criteria (specifically, cost per click for advertisements), by a search engine that plays the role of a mediator, when creating search-results sets in response to queries was studied by Eliaz and Spiegler (2011). In contrast, we assume a fixed ranking function per search setting, and our analysis is based on coupling relevance and monetization. Furthermore, the adversarial search settings we explore, which rise from different assumptions about publishers'

6. https://www.google.com/trends/

strategies, were not addressed by Eliaz and Spiegler (2011), nor were the ranking mechanisms we study (i.e., the Probability Ranking Principle and the novel non-deterministic ranking functions).

3. Preliminaries

Let $[n]$ denote the set $\{1, 2, \ldots, n\}$, and let $\Delta(S)$ denote the set of all probability distributions defined over a set S.

Given a distribution $d \in \Delta(S)$, we use $d(i)$ to denote the probability of an event $i \in S$, and $\mathrm{SUP}\,(d)$ denotes the set of all elements $i \in S$ for which $d(i) > 0$.

3.1 Game Theory

We start with a standard definition of a game.

Definition 1 *An n-player game is a tuple $G = \left(\{S_i\}_{i\in[n]}, \{U_i\}_{i\in[n]}\right)$. For each player $i \in [n]$, S_i is the set of* **actions** *of the player. The set $S = S_1 \times S_2 \times \cdots \times S_n$ is called the set of* **pure strategy profiles**. *For each $i \in [n]$, $U_i : S \to \mathbb{R}$ is the* **utility function** *of player i.*

In a game, each player tries to maximize the (expected) value of his utility function. Note that the utility of a player depends on actions taken by all players, not only his own. The case of a two-player games can be represented by a bimatrix — a matrix in which each cell contains a 2-tuple: the first value is the payoff for the first player, and the second value is the payoff for the second player. Each row in the matrix represents an action of the first player, and each column represents an action of the second player; i.e., the matrix has $|S_1|$ rows and $|S_2|$ columns. For example, consider the payoff matrix

$(5, 2)$	$(0, 4)$
$(3, 6)$	$(1, 1)$

.

According to the matrix, if the players chose actions 2 and 1 respectively, the first (row) player payoff will be 3, while the second (column) player payoff will be 6. In a game, each player may choose an action randomly:

Definition 2 *A* **mixed strategy** *for player i is a distribution $s_i \in \Delta(S_i)$.*

Using this definition, we extend the notion of a strategy profile so that $\Delta S \triangleq \Delta(S_1) \times \ldots \times \Delta(S_n)$ is the set of **mixed strategy profiles**.

We assume that each player i chooses his action according to $\Delta(S_i)$ *independently* of the actions selected by the other players. This allows a straight-forward generalization of the player's utility to the *expected* payoff over the joint distribution. That is, let $s = (s_1, \ldots, s_n) \in \Delta S$ be a mixed strategy profile, then

$$U_i(s) = \sum_{(a_1,\ldots,a_n)\in S} \left(U_i(a_1, \ldots, a_n) \prod_{i=1}^{n} s_i(a_i) \right).$$

We denote $\Delta S_{-i} \triangleq \Delta(S_1) \ldots \times \Delta(S_{i-1}) \times \Delta(S_{i+1}) \times \ldots \times \Delta(S_n)$ — the strategy profiles of everyone but i, and use $s_{-i} = (s_1, \ldots, s_{i-1}, s_{i+1}, \ldots, s_n) \in \Delta S_{-i}$ to denote the strategies of the other players.

A useful notion is the concept of **best response**. This is the set of actions a player may take to maximize his utility given the actions played by all other players:

Definition 3 *Given a strategy profile s_{-i}, the best response of player i is defined as* $BR_i(s_{-i}) \triangleq \mathrm{argmax}_{a_i \in S_i} U_i(a_i, s_{-i})$.

Finally, this allows defining a "stable" state of a game: a strategy profile for which no player has an incentive to **deviate**, i.e., change his strategy. This concept is often referred to as the solution of a game.

	Football	Opera
Football	$(2,1)$	$(0,0)$
Opera	$(0,0)$	$(1,2)$

Table 1: Battle of the Sexes' payoff matrix.

Definition 4 *A mixed strategy profile* $s = (s_1, \ldots, s_n) \in \Delta S$ *is a (mixed)* **Nash equilibrium**, *if for every player* $i \in [n] : SUP(s_i) \subseteq BR_i(s_{-i})$. *Equivalently,* s *is an equilibrium if* $\forall i \in [n], \forall a'_i \in S_i : U_i(a'_i, s_{-i}) \leq U_i(s_i, s_{-i})$. *If* s *is a pure strategy profile, the equilibrium is called* pure.

For example, consider the two-player game known as "Battle of the Sexes", where a couple has agreed to go on a date this evening, but forgot to agree on a venue. The husband prefers going to a football match, while the wife wants to see the opera, and they cannot communicate. Both would prefer meeting at any place rather than not to meet at all. The game payoffs are given in Table 1. The game has two pure equilibria: jointly watching the match or attending the opera together. This game also has a mixed Nash equilibrium, where the husband goes to the match with probability 2/3 while the wife attends the opera with the same probability. Note that if each were to pick a venue uniformly, they would have met with a higher probability. However, the different preferences of the two incentivize each to try his/hers place of choice with a higher probability. A fundamental result in game theory is Nash Theorem:

Theorem 1 *(Nash, 1950) Every game with a finite number of players having finite action sets has a mixed strategies equilibrium.*

A common metric for measuring the "quality" of a given mixed strategy profile is called **social welfare**. The social welfare is defined as the sum of (expected) utilities received by the players playing it; that is, for $s \in \Delta S$ we define $SW(s) \triangleq \sum_{i \in [n]} U_i(s)$.

3.2 Price of Anarchy

Given a game, an interesting question is "how inefficient" the system could get in a stable state, due to the selfish behavior of the players. Specifically, an equilibrium of the game is compared with an idealized scenario where all players selflessly collaborate to maximize the profit of the group which is measured by social welfare. A popular measure of this inefficiency is called the **Price of Anarchy** (PoA)(Koutsoupias & Papadimitriou, 1999; Roughgarden & Tardos, 2002). Formally, the price of anarchy is the ratio between the sum of utilities achieved by an optimal strategy profile (i.e., $OPT = \max_{s \in \Delta S} SW(s)$), and the social welfare of the *worst* equilibrium, i.e., the one with the lowest social welfare. Going back to the example from Table 1, when both are going together to either the opera or the football match, we have an optimal social welfare of 3, while the mixed equilibrium only provides a social welfare of 4/3. Thus, the price of anarchy for the game is $\frac{3}{4/3} = 2.25$.

3.3 Safety Level Strategies

In a Nash equilibrium, each player is assumed to be playing the equilibrium strategy as no player has an incentive to deviate. However, the utility for a player might be affected if other players choose not to follow the equilibrium. That is, if we cannot assume that players are rational in the sense that each tries to maximize his payoff, playing the equilibrium strategy can lead to sub-optimal results. Intuitively, a Safety Level Strategy is designed to guarantee a payoff to the player, regardless of the actions taken by others. Formally, given a game $G = \left(\{S_i\}_{i \in [n]}, \{U_i\}_{i \in [n]}\right)$, the Safety Level Strategy of player i is defined as:

$$SLS_i(G) = \text{argmax}_{s_i \in \Delta(S_i)} \min_{s_{-i} \in \Delta(S_{-i})} U_i\left(s_i, s_{-i}\right).$$

An important observation is that the utility of a player playing in an equilibrium is always at least as large as the payoff he may guarantee playing the Safety Level Strategy, as the equilibrium strategy assumes all players follow it, while the Safety Level Strategy does not assume anything about the other players.

4. Adversarial Ad Hoc Retrieval

As already noted, in adversarial retrieval settings such as the Web, the goal of many owners of documents, henceforth referred to as *publishers*, is to have their documents ranked as high as possible in response to queries of interest. Specifically, a ranking induced for a query by a search engine might incentivize publishers of documents that are not highly ranked to change their documents for future rankings. Naturally, this is an on-going process.

We model the dynamics just mentioned as a game. The publishers are players whose actions are document production; i.e., a publisher decides which document to produce at any point in time. The search engine plays the role of a *mediator*. The choice of a ranking function affects the publishers' utilities by the virtue of the induced rankings.

We make the assumption that publishers' utilities and users' utilities are aligned. That is, the payoff a publisher gets for a document with respect to a user query reflects the utility the user attains from reading the document in response to the query. In other words, the premise is that user satisfaction will lead to publisher satisfaction. In that respect, the social welfare of publishers (as players) attained in a game is considered to be the social welfare of the search engine's users. Analysis of more complicated situations wherein the publishers' and users' utilities are not aligned is left for future work.

Our goal here is to study the optimality, or lack thereof, of ranking functions that obey the Probability Ranking Principle (PRP) in adversarial retrieval setting; that is, functions that rank documents in response to a query by the relevance (probability) of the documents. To that end, we describe games that are defined by the choice of the ranking function and the assumptions about the type of documents the publishers can produce. We assume that in each game the ranking function we want to analyze is *fixed*. Accounting for changes of the ranking function throughout the game is outside the scope of this paper.

To analyze a ranking function, we examine the equilibria of the game it entails. Specifically, we focus on the price of anarchy of the game (PoA; see Section 3.2), as is common in work on algorithmic game theory. Indeed, this practice allows to reason about the worst case steady state of the game (equilibrium), in terms of the social welfare attained, given the selfish behavior of players — publishers in our case.

4.1 Single-topic Documents

Here and after, we assume that each query is about a single topic. We use the term "topic" for the mere convenience of discussing the connection (or more exactly the "relevance match") between a query and a document. We assume that a relevance event happens for a pair of a document and query if and only if the two discuss the same topic. We do not engage in estimating relevance, but rather assume that the fact of whether a relevance event happened is known. We examine the social welfare that results from applying different ranking paradigms that use this knowledge. In that respect, a more strict (and a somewhat less realistic) interpretation of a "topic" is information need. That is, assuming a universe of information needs, the premise is that the query is about a single (uniquely defined) information need. A relevance event for a document happens in that case if it includes information satisfying the information need.

We start our analysis in this section with a basic model where every document is also assumed to be about a single topic. (In the more strict interpretation of "topics" mentioned above, each document can satisfy a single information need.) The model assumes, in addition, that all publishers have equal writing quality for all topics and that each publisher may write on a topic of his choice. We note that in many practical cases, publishers will not change their documents so as to discuss completely different topics, but will rather fine tune — via search engine optimization (SEO) — their pages for queries of interest. However, considering topics in a more refined way, specifically, as specific aspects of more general topics, such changes actually do happen quite often. For example, a publisher describing a product in her document might decide to change the document content to discuss a completely different aspect of the product with the goal being attracting currently trending queries. In the strict view of topics described above, the publisher essentially changed the information need that the document can satisfy.

We show that in the single-topic documents model just described, the Probability Ranking Principle (PRP) is optimal; i.e., using a ranking function that obeys the PRP results in an optimal price of anarchy. In sections 4.2 and 4.3 we extend this model by allowing multi-topic documents, and differential writing qualities, respectively, and show that the PRP is no longer optimal in these cases. We note that assuming that a publisher has different writing qualities for topics is one way to address the standard SEO process that documents go through. That is – the more distant the document is from the best-quality document the publisher can, and intended to, produce (regardless of SEO considerations) – the higher the likelihood is that the quality is lower. A case in point, excessive SEO attempts (such as keyword stuffing) often hurt the quality of documents and consequently the utility that can be attained from them.

4.1.1 Definitions

A **Publishers Game** $PG = (n, m, \mathcal{D}; \mathcal{R})$ is an n **publishers** game over m **topics**. Each publisher selects a topic to write on; i.e., the possible actions are topics. In addition, **queries** are posted, each of which on a specific topic, according to a public **query distribution** $\mathcal{D} \in \Delta([m])$. Throughout the paper, we assume without loss of generality that $\mathcal{D}(1) \geq \mathcal{D}(2) \geq \ldots \geq \mathcal{D}(m)$. That is, the first topic is always the most frequently queried, then the second, etc.

For the sake of simplifying the analysis, we assume that the user utility with respect to his query is determined solely based on the relevance of the first ranked document. While in practice users examine also documents in lower ranks, it is a well known fact that in the Web setting users pay most of their attention to the most highly ranked documents (Joachims et al., 2005). In fact, users often prefer to reformulate their queries rather than spend additional time examining lower ranked documents (Liu & Wei, 2016). Furthermore, it was shown that the precision at rank 1 (i.e., 0 or 1) is highly correlated (with Pearson correlation above 0.5) with average precision at cutoff 1000, which is a standard measure for quantifying the effectiveness of a retrieved list (Butman et al., 2013). Thus, we assume that the utility the user attains from the top-most document is highly correlated with the utility the user will attain by examining additional documents further down the ranked list.

Given a query and the set of topics selected by the publishers, the **Ranking Function** $\mathcal{R} : [m] \times [m]^{[n]} \to \Delta([n])$, selects a probability distribution over the documents (i.e., over the players) using which the documents (players) will be ranked; specifically, the probability assigned to a document corresponds to the probability that it will be ranked first. We note that this definition is different than that of standard ranking functions in two respects. First, it does not specify how to rank documents for ranks 2 and below. Second, the ranking function is allowed to be *probabilistic*, randomizing over the set of documents in the corpus. In contrast, ranking functions used in the information retrieval literature are deterministic. While at first glance randomization may not seem to provide additional value, we show that it could actually result in an improved social welfare even if it hurts retrieval effectiveness for a specific query. We assume that ranking functions are oblivious to the players' identity and thus if $i, i' \in [n]$ chose to write about the same topic, they have an equal probability of being ranked first for that query (i.e., $\mathcal{R}_i(s) = \mathcal{R}_{i'}(s)$), regardless of other the players actions.

As an example of the notation used for ranking functions, consider $\mathcal{R}(1, \langle 1, 1, 2\rangle) = \langle 0.4, 0.4, 0.2\rangle$. Given a query on the first topic, if players 1 and 2 wrote on topic 1, and player 3 created a document

	1	2	...	m
1	$\left(\frac{\mathcal{D}(1)}{2}, \frac{\mathcal{D}(1)}{2}\right)$	$(\mathcal{D}(1), \mathcal{D}(2))$		$(\mathcal{D}(1), \mathcal{D}(m))$
2	$(\mathcal{D}(2), \mathcal{D}(1))$	$\left(\frac{\mathcal{D}(2)}{2}, \frac{\mathcal{D}(2)}{2}\right)$		$(\mathcal{D}(2), \mathcal{D}(m))$
$\vdots$			$\ddots$	
m	$(\mathcal{D}(m), \mathcal{D}(1))$	$(\mathcal{D}(m), \mathcal{D}(2))$		$\left(\frac{\mathcal{D}(m)}{2}, \frac{\mathcal{D}(m)}{2}\right)$

Table 2: Payoff matrix for the game $(2, m, \mathcal{D}; R_{PRP})$

on topic 2, then each of players 1 and 2 will be ranked first with probability 0.4, while player 3 will be ranked first with probability 0.2. We use $\mathcal{R}_i : [m] \times [m]^{[n]} \to [0, 1]$ to denote the probability assigned to player i for the given input; in our example, $\mathcal{R}_3(1, \langle 1, 1, 2 \rangle) = 0.2$.

We define the payoff for publisher i as 1 for queries he is ranked first on, and 0 for all other queries. His expected payoff is then $U_i(a_i, a_{-i}) = \mathcal{D}(a_i) \cdot \mathcal{R}_i(a_i, a_{-i})$; i.e., the probability that a query on his selected topic arrives times the probability he will be ranked first. We define the Probability Ranking Principle ranking, **PRP ranking** in short, denoted R_{PRP}, to be the function which always ranks first the document most relevant to a query. Indeed, the PRP (Robertson, 1977) states that documents should be ranked by their relevance probabilities, and we care here about the highest ranked document. In this model, the PRP translates to choosing uniformly a document written on the query topic for the first rank. For example, for the two publishers game $(2, m, \mathcal{D}; R_{PRP})$, where the PRP serves for ranking, the payoffs can be expressed by the matrix in Table 2.

4.1.2 Connection to Congestion Games

We show a connection to the extensively studied class of congestion games (Rosenthal, 1973; Monderer & Shapley, 1996), and specifically to singleton games (Ieong et al., 2005). As congestion games of this form are known to allow relatively efficient computation of an equilibrium, as well as relatively fast convergence to an equilibrium, showing a reduction to such games imply that these are applicable here too. For self containment, we provide the definition of a singleton congestion game.

Definition 5 *Let R be a set of* resources *and let $n \in \mathbb{N}$. Assume that each player $i \in [n]$ has an action set $S_i \subseteq R$ and that each resource $r \in R$ has an associated profit function $p_r : [n] \to \mathbb{R}$. A* **Singleton Congestion Game** *$SCG = \langle n, R, \{S_i\}_{i \in [n]}, \{p_r\}_{r \in R} \rangle$ is a n-players game where each selects a single resource from R. Given a (pure) strategy profile a, the payoff function for player i depends only on the resource and the number of players that chose the same resource. Specifically, the utility is defined as*

$$U_i(a) \triangleq p_{a_i}\left(\left| \{ j \in [n] \mid a_j = a_i \} \right| \right).$$

Next, we prove that a Publishers Game is in fact a Singleton Congestion Game.

Theorem 2 *For any $n, m \in \mathbb{N}^+$, $\mathcal{D} \in \Delta(m)$ the Publishers Game $PG = (n, m, \mathcal{D}; R_{PRP})$ is a Singleton Congestion Game.*

Proof. *Let $R = [m]$, $\forall i \in [n] : S_i = [m]$, and $\forall j \in [m], a \in [m]^{[n]} : p_j(a) = \frac{\mathcal{D}(j)}{\{i \in [n] | a_i = j\}}$. By the definition of R_{PRP}, if a player i wrote about topic j, it will be ranked first for queries about j with probability $(1/\#j)$, where $\#j$ is the number of players that wrote about topic j. This means that for a given strategy profile a, her payoff is indeed $U_i(a) = \frac{\mathcal{D}(j)}{\{i \in [n] | a_i = j\}}$.* □

4.1.3 Price of Anarchy

We focus on two-player games, where a couple of rival publishers compete for the same incoming queries flow, each trying to maximize his payoff by being ranked first. Our goal in this section is showing that in this basic model, the PRP remains optimal. We start with a theorem showing that the PRP ranking yields a game in which every equilibrium is at most 1.5 times worse than the optimal scenario in terms of social welfare. The optimal scenario is having each of the two publishers write about a different topic among the two most queried topics.

Theorem 3 *Let $\mathcal{D} \in \Delta([m])$ be a query distribution and let $PG = (2, m, \mathcal{D}; R_{PRP})$ be a two player Publishers Game, then the price of anarchy of PG is at most* 1.5.

The proof, presented in Appendix A, is based on analysis of two cases: when publishers write deterministically on a single topic, and when they choose to randomize over several topics. We show that in any case, the worst equilibrium is reached when both publishers choose a topic only from the two most frequently queried topics.

We complement Theorem 3 by proving that no ranking function is able to achieve price of anarchy smaller than 1.5, for any query distribution. Together, the two theorems show that the PRP is optimal in the case where each document (and query) is about a single topic, and all publishers have the same writing quality for all topics. As noted above, optimality refers to the price of anarchy unless otherwise noted.

Theorem 4 *For any ranking function $\mathcal{R}$, there exists a distribution $\mathcal{D} \in \Delta([m])$, such that the Publishers Game $PG = (2, m, \mathcal{D}; \mathcal{R})$ has a price of anarchy of* at least 1.5.

Intuitively, we show that for the query distribution in which the first topic is queried two-thirds of the time, while the second is queried otherwise, $\mathcal{R}$ cannot decrease the price of anarchy below 1.5. The full proof is in Appendix B.

For the case of $m = 2$ topics, we strengthen the result of Theorem 4 by showing that for *any* distribution, one cannot improve the social welfare obtained by R_{PRP} using a different ranking function.

Theorem 5 *For any distribution $\mathcal{D} \in \Delta([2])$ and ranking function $\mathcal{R}$, the social welfare of $(2, 2, \mathcal{D}; \mathcal{R})$ is at most the social welfare of $(2, 2, \mathcal{D}; R_{PRP})$.*

The proof of the theorem is provided in Appendix C. Thus, we get that in the two topics setting, the PRP ranking is optimal in a stronger sense than that considered above (i.e., price of anarchy): there exists *no* query distribution for which an alternative function could achieve higher social welfare than that attained by the PRP.

While we showed that the PRP is optimal in this model, it is not the case in the models considered next. Section 4.2 discusses a setting in which publishers may produce multi-topic documents. Section 4.3 presents a model which accounts for publishers with differential writing qualities.

4.2 Multi-topic Documents

We next discuss an extension of the Publishers Game that allows publishers to compose documents about multiple topics. Formally, a **Multiple Topics Publishers Game** $MTPG = (n, m, c, \mathcal{D}; \mathcal{R})$ is similar to Publishers Game, except that publishers may now choose an action from $[m] \cup \binom{[m]}{2}$; i.e., they can either write a document on a single topic or on two topics. Accordingly, a ranking function $\mathcal{R} : [m] \times \left([m] \cup \binom{[m]}{2}\right)^{[n]} \to \Delta([n])$ is generalized to consider multi-topic documents. We assume that a multiple topics document has equal proportions of the two topics. When a publisher writes on two topics, an additional parameter, $c \in [0, 1]$, is the payoff he receives when ranked first on either of these topics. We note that $(1 - c)$ can be interpreted as the amount of effort a user needs to put into distilling the topic of her interest from the two-topic document. The higher the effort, the lower the utility the user – and therefore the publisher – attain. (Recall that we assume that the publisher and user utilities are aligned.)

We next have to define what a ranking that obeys the PRP is. Recalling that the PRP states that documents should be ranked according to their probability of being relevant, we use the following definition. Assuming that documents are of equal length, we assume that the PRP amounts to determining relevance level (grade), or probability of relevance, based on the prevalence of relevance information in the document. This definition was also applied in work on focused retrieval (Geva, Kamps, & Schenkel, 2012) where the extent a document was relevant to a query was based on the prevalence of relevant information in the document. Specifically, a document dedicated entirely to the query topic is assumed to be more relevant than any document that discusses two topics, given that all documents are of the same length. For two documents that discuss two topics, the one which discusses the query topic is assumed to be more relevant than the one which does not; if both documents discuss the query topic, one of the two is randomly selected to be ranked first.

It is important to note that the original PRP definition, and accordingly that which we apply here, define only the ranking of documents and not a utility function. It was originally shown that if utility is defined solely based on relevance, and that the relevance of one document is independent of others, and that all users have the same utility function, then the PRP results in optimal utility (Robertson, 1977). Here we do not alter the PRP definition. Rather we define a relevance grade based on the amount of relevant information in the document, and independently introduce a utility definition that accounts for the different potential levels of effort required to distill information about one topic out of the two a document discusses. Thus, we allow for a finer connection between relevance and utility in this setting. Conceptually, a similar practice was also used originally to demonstrate the sub-optimality of the PRP when different users have different utility functions; specifically, the functions differed in the way they depended on relevance (Robertson, 1977).

Next, we show that for any positive constant $c \in (0,1) \setminus \{0.5\}$, the Multiple Topics Publishers Game under the PRP ranking is not a congestion game. This result means that realistic variants of the games originating from information retrieval, which one may view as similar in spirit to congestion games popular in the computer science literature (where topics are resources, and one's payoff is proportional to the number of people selecting his resource), go beyond the scope of such games. The proof of the next theorem appears in Appendix D.

Theorem 6 *Let $n, m \in \mathbb{N}^+$, $c \in (0,1) \setminus \{0.5\}$, and $\mathcal{D} \in \Delta([m])$, then $MTPG = (n, m, c, \mathcal{D}; R_{PRP})$ is not a congestion game.*

We focus on the special case of two documents and two publishers, showing that even in this restricted scenario, the PRP is not always optimal. We start by showing that if the multi-topic document payoff c is small, the PRP ranking remains optimal.

Theorem 7 *Let $c \in [0, 1/2]$. For any distribution $\mathcal{D} \in \Delta([m])$, the Multiple Topics Publishers Game $MTPG = (2, m, c, \mathcal{D}; R_{PRP})$ achieves a price of anarchy of at most* 1.5.

The idea behind the proof, presented in Appendix E, is showing that when multi-topic documents provide just a small benefit, the publishers will never write such documents; in this case, we can use Theorem 3 for the price of anarchy bound.

The next theorem shows that for small c values, the PRP is indeed optimal as any ranking function will incur a price of anarchy of 1.5 for some query distribution.

Theorem 8 *Let $c \in [0, 1/3]$. There exists a distribution $\mathcal{D} \in \Delta([m])$, such that for* any *ranking function $\mathcal{R}$, the Multiple Topics Publishers Game $MTPG = (2, m, c, \mathcal{D}; \mathcal{R})$ has a price of anarchy of* at least 1.5.

Similarly to Theorem 4, we consider a query distribution where the first topic is queried two-thirds of the time, and otherwise the second topic is queried. In the proof, which appears in Appendix F, we show that for this distribution, no ranking function could reduce the price of anarchy when c is small.

Next, when c is sufficiently large, we show an example in which the anarchy entailed by using the PRP ranking could be completely avoided by using an alternative ranking function which does not adhere to the PRP; that is, a social welfare that approaches the optimal can be attained.

Example 1 *Consider the distribution $\mathcal{D}(1) = 2/3, \mathcal{D}(2) = 1/3$. For any c, the price of anarchy of $MTPG_{PRP} = (2, 2, c, \mathcal{D}; R_{PRP})$ is 1.5, while there exists a ranking function, R_{prob}, such that the price of anarchy of $MTPG_{prob} = (2, 2, c, \mathcal{D}; R_{prob})$ could be arbitrarily close to 1 as $c \to 1$.*

For the example, we propose a non-deterministic, probabilistic, ranking function[7] R_{prob}. The function randomizes the relative ranking of a multi-topic document and a pure document written on the first topic, whenever a query on the first topic is posted. The complete proof is provided in Appendix G.

We conclude this section by showing a large family of games in which the PRP is not optimal, and propose a probabilistic (randomized) ranking function instead. This finding implies that introducing randomness to the *ranking* of documents may result in a more content enriched corpus in the long run, by providing incentives for publishers to diversify the topics they write about.

Theorem 9 *Let $c \in (1/2, 1]$ and $p \in [2/3, \frac{2c}{c+1})$. Consider the distribution $\mathcal{D}(1) = p, \mathcal{D}(2) = 1 - p$. There exists a ranking function R_{prob}, such that the price of anarchy of $(2, 2, c, \mathcal{D}; R_{PRP})$ is strictly larger than that of $(2, 2, c, \mathcal{D}; R_{prob})$.*

The core idea underlying the proof of the theorem, which can be found in Appendix H, is to carefully construct a probabilistic ranking function which occasionally ranks a multi-topic document higher than a document written entirely on the query topic.

4.3 Publishers with Differential Writing Qualities

Heretofore, we assumed that every publisher can write about every topic with the same quality. We now turn to analyze a setting wherein this is not the case. Specifically, we explore an extension of the Publishers Game to a setting where publishers write on a single topic, but have different writing qualities. We assume that the writing quality of a publisher per topic represents the expertise and knowledge the publisher has about this topic. Assuming that the writing quality of a publisher per topic can be set by the publisher, e.g., by having a cost incurred by reaching a desired quality level, results in an interesting different setting than that we explore here, and we leave its exploration for future work.

Formally, a **Different Qualities Publishers Game** $DQPG = (n, m, \mathcal{Q}; \mathcal{R})$ is similar to the Publishers Game, except that publishers now have individual payoff functions representing the utility of the document they create for a queried topic. This setting amounts to assuming, for example, that the utility (for the user, and consequently for the publisher, per our working assumption) attained from the same document produced by two different publishers is different and depends on their writing quality for the topic. Indeed, the less the publisher is authoritative about the topic, the more effort the user will have to put into verifying the content (e.g., by consulting other sources of information), thereby decreasing her utility. Along these lines, we essentially assume that the more distant a document is from the original "genuine" document the publisher is authoritative to write, the higher the likelihood of the document to be of lower quality. This assumption corresponds to the fact that excessive SEO attempts can hurt the quality of a document (e.g., excessive keyword stuffing), and therefore decrease the utility attained from it.

7. As explained in Section 4.1. By probabilistic ranking function we mean a non-deterministic ranking function that applies randomization. This type of a function should be differentiated from probabilistic retrieval methods that are deterministic and which rank documents by relevance probabilities (Sparck Jones et al., 2000).

The game has a **Quality Matrix** parameter, $\mathcal{Q} \in [0,1]^{[n]\times[m]}$, where $\mathcal{Q}_{i,j}$ is the payoff for publisher i if he is ranked first on topic j. Note that the query distribution parameter has been omitted as we may normalize the Quality Matrix to factor the topic frequencies. For example, if the first topic is queried 2/3 of the time, and otherwise the second topic is queried, then we multiply the first column of $\mathcal{Q}$ by 2/3 and the second column by 1/3. In this setting, we assume that in the PRP ranking, R_{PRP}, the *highest quality* document written on the queried topic is ranked first. In case multiple documents have the same quality, it uniformly selects a top-quality document. That is, originally, the PRP was shown to optimize utility by *implicitly* assuming equal writing qualities of publishers; thus, relevance amounted to utility for a given query. Here, we stick to the original goal of the PRP, optimizing utility, and account for the fact that relevance is not necessarily coupled with utility.

Similarly to the Multiple Topics Publishers Game, we show that the Different Qualities Publishers Game is not a congestion game. We start by examining two-player games; the complete proof of the lemma appears in Appendix I.

Lemma 1 *Let $\mathcal{Q} \in [0,1]^{[2]\times[m]}$ whose entries are positive and distinct, then DQPG $= (2, m, \mathcal{Q}; R_{PRP})$ is not a congestion game.*

Next, we show that this argument generalizes to any number of players. For simplicity, we show the proof for $m \geq 3$, although the result holds for two-topic games as well.

Theorem 10 *Let $\mathcal{Q} \in [0,1]^{[n]\times[m]}$ whose entries are positive and distinct, then DQPG $= (n, m, \mathcal{Q}; R_{PRP})$ is not a congestion game.*

Proof. Let $i, j \in [n]$ be the two players with the highest writing quality on topic 1, and assume without loss of generality that $\mathcal{Q}_{i,1} > \mathcal{Q}_{j,1}$. By considering action profile in which all of the remaining players write about the third topic (i.e., $a_{-ij} = \langle 3, 3, \ldots, 3\rangle$), we get the same utility quantities as in the above lemma. This means that the game is not a congestion game for any number of players. $\square$

We continue with a tight bound for the price of anarchy achieved by the PRP ranking.

Theorem 11 *Let $\mathcal{Q} \in [0,1]^{[n]\times[m]}$ be a Quality Matrix such that no two players have the same writing quality on the same topic and no player has equal writing quality on two different topics. Then the Different Qualities Publishers Game DQPG $= (n, m, \mathcal{Q}; R_{PRP})$ achieves a price of anarchy of at most 2.*

The essence of the proof, provided in Appendix J, is reducing the problem to finding a matching in the bipartite graph which has the publishers on one partite and the topics on the other. We show that the game has a unique equilibrium that corresponds to a matching which may be generated by the greedy algorithm, and that the approximation ratio of the algorithm provides a bound on the price of anarchy.

Complementing the theorem, we show that the analysis for R_{PRP} is tight.

Theorem 12 *For any $\epsilon > 0$ and $n, m \geq 2$, there exists a Quality Matrix $\mathcal{Q}^\epsilon \in [0,1]^{[n]\times[m]}$ whose entries are positive and distinct, such that the price of anarchy of the Different Qualities Publishers Game $(n, m, \mathcal{Q}^\epsilon; R_{PRP})$ is $2 - \epsilon$.*

The result is obtained for the simple case where one publisher has writing quality for multiple topics, while another publisher can write mainly on a specific topic; the rest of the players can only produce low quality documents, regardless of the topic they choose. If the first player has even a slight writing quality advantage over the second on that specific topic, he may prefer writing about it, and no publisher would satisfy the information need for the remaining topics. The complete proof is found in Appendix K.

It is important to note that if a ranking function could act arbitrarily, a search engine could abuse this model to force the optimal strategy, which is the *maximum weight matching* of the complete bipartite graph over $[n] \cup [m]$, where the weight of an edge (i, j) is $\mathcal{Q}_{i,j}$. For example, consider the ranking function which given a query on topic j, ranks first the publisher i, if (i, j) is a part of the maximum weight matching, and i wrote a document on topic j. Otherwise, it places a completely irrelevant document at the top position. In this case, every publisher realizes that the only way users with information need relevant to his page reach his document is by cooperation with the engine, writing on the topic assigned to him in the matching.

In order to avoid ranking functions which dictate for each player the topic to write on, we restrict the discussion to **Fair Ranking Functions**. Intuitively, a Fair Ranking Function is a function which given a query, assigns first-rank probabilities (i.e., probability to be ranked first) to documents written on the topic, without considering the writing qualities of the publishers on *other* topics. Formally, a Fair Ranking Function is a *set* of functions $\mathcal{R} = \{\mathcal{R}^k : [0,1]^{[k]} \to \Delta([k])\}_{k \in [n]}$, such that given a query on a topic that k publishers wrote about, $\mathcal{R}^k$ is applied on the qualities of these documents to determine the ranking, and the ranker is not allowed to consider the writing qualities of publishers on topics they did not write on.

Theorem 12 shows that for *any* values of n, m (i.e., number of publishers and number of topics, respectively) the price of anarchy of the PRP is 2. We complement this by showing that for the special case of two-player games, R_{PRP} is not optimal, even when compared only with Fair Ranking Functions.

Theorem 13 *There exists a Fair Ranking Function R_{uni}, such that for any Quality Matrix $\mathcal{Q} \in [0,1]^{[2]\times[2]}$, the Different Qualities Publishers Game DQPG $= (2, 2, \mathcal{Q}; R_{uni})$ achieves a price of anarchy of at most* 1.683.

The proof, which appears in Appendix L, considers a *probabilistic* ranking function R_{uni}. The function makes a fair coin flip to select the top ranked document, whenever the difference in quality of given documents is "small enough". We show that this mechanism could improve the effectiveness of the retrieval, in means of price of anarchy. Next, we discuss the properties of a safety-level strategy profiles in this game.

Lemma 2 *There exists a Fair Ranking Function R_{uni}, such that for any Quality Matrix $\mathcal{Q} \in [0,1]^{[2]\times[2]}$, the social welfare of every safety-level strategy profile in DQPG $= (2, 2, \mathcal{Q}; R_{uni})$ is at least $\frac{1}{1.683}$ times the social welfare of any strategy profile.*

The proof the lemma, given in Appendix M, is done by an extensive case analysis over the possible supports of the players' strategies. Next, we use the lemma for generalizing Theorem 13 to the more general m-topics game.

Theorem 14 *For any $m \geq 2$, there exists a Fair Ranking Function R_{uni}, such that for any Quality Matrix $\mathcal{Q} \in [0,1]^{[2]\times[m]}$, the Different Qualities Publishers Game DQPG $= (2, m, \mathcal{Q}; R_{uni})$ achieves a price of anarchy of at most* 1.683.

The proof, appearing in Appendix N, uses Lemma 2 for a modified version of the game in which players may only write about 2 topics, the ones which provide the optimal social welfare. It then argues that every player in only do better in the original game, hence proving the theorem.

Next, we show a lower bound on the price of anarchy of any Fair Ranking Function.

Theorem 15 *For any $n, m \geq 2$, there exists a quality matrix $\mathcal{Q} \in [0,1]^{[n]\times[m]}$ such that for every Fair Ranking Function $\mathcal{R}$, the price of anarchy of the Different Qualities Publishers Game DQPG $= (n, m, \mathcal{Q}; \mathcal{R})$ is at least* $4/3$.

The proof, deferred to Appendix N, shows that if two players are indistinguishable, no Fair Ranking Function can achieve a price of anarchy lower than $4/3$. However, in reality, it might be unlikely that two publishers have the same writing quality. For this, we say that $\mathcal{Q} \in [0,1]^{[n]\times[m]}$ is a

Distinguishable Quality Matrix if for every topic, all of the writing qualities of the publishers are distinct. We now show that for 2 players, a non-trivial lower bound exists even when considering only distinguishable quality matrices.

Theorem 16 *For any Fair Ranking Function $\mathcal{R}$ and $m \geq 2$, there exists a Distinguishable Quality Matrix $\mathcal{Q} \in [0,1]^{[2]\times[m]}$, such that the price of anarchy of the Different Qualities Publishers Game $DQPG = (2, m, \mathcal{Q}; \mathcal{R})$ is at least* 1.5.

In Appendix P, we present a proof which constructs two games with different quality matrices. We show that any ranking function achieves social welfare which is at most two thirds of the optimum, at least for one of the games.

5. Conclusions and Future Work

We presented a novel game theoretic analysis of the ad hoc document retrieval task in adversarial settings. The analysis accounts for the incentives of authors to have their documents ranked high in response to queries. Thus, the analysis provides formal grounds for modeling the dynamic nature of the adversarial retrieval setting that results from authors consistently changing their documents so as to promote them in rankings. We performed the analysis for different types of documents (namely, single topic versus multi topic) and by using different assumptions about the writing qualities of authors. One of our most important findings is that the Probability Ranking Principle (PRP), which is the theoretical foundation of most ad hoc retrieval methods, can be sub-optimal in the adversarial setting. Specifically, we showed that in some cases, introducing randomization into the document ranking function can result in user utility higher than that attained by applying the PRP.

For future work we plan to devise additional non-deterministic ranking functions that can improve social welfare in the games we presented. Empirical study of such functions is also an important future venue.

Additional future venue which we intend to explore, as mentioned above, is studying a setting where publishers have explicit costs of changing their documents (e.g., as a function of the extent of change). One of the goals of these changes can be improving the quality of a document for topics of interest.

Acknowledgments

We thank the reviewers for their comments. Early part of this project has been supported and carried out at the Technion-Microsoft Electronic Commerce Research Center. Ran Ben Basat was partially funded by the Technion-HPI research school. Moshe Tennenholtz is supported by funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement 740435).

Appendix A. Proof of Theorem 3

We here prove the following:

Theorem *Let $\mathcal{D} \in \Delta([m])$ be a query distribution and let $PG = (2, m, \mathcal{D}; R_{PRP})$ be a two player Publishers Game, then the price of anarchy of PG is at most 1.5.*

Proof. The optimal scenario we compare to, is when some publisher writes about topic 1, while the other composes a document on topic 2. The optimal social welfare is therefore $OPT = \mathcal{D}(1)+\mathcal{D}(2) \leq 1$. Next, we cover all possible equilibria:

1. Each player writes on a different topic.
   When such equilibrium exists, players must write about topics 1 and 2 (or other topics with the same frequency, if such exists). The social welfare of the equilibrium is then $\mathcal{D}(1) + \mathcal{D}(2)$, and the players reach optimal welfare.

2. Both players write about topic 1 with probability 1.
   In this scenario, $\mathcal{D}(1) \geq 2/3$, as the publisher have no incentive to deviate to topic 2. This means the social welfare of the equilibrium is $\mathcal{D}(1)$, and the price of anarchy is at most $\frac{1}{2/3} = 1.5$.

3. The players play a symmetric equilibrium (i.e., both have the same mixed strategy), randomizing their actions over the first two topics.
   As both players have topic 2 in their support, we know that $\mathcal{D}(2) \geq \mathcal{D}(1)/2$. Assume that each player writes about the first topic with probability p. The payoff of writing about topic 1, assuming that the other publishers follows this strategy, is $\mathcal{D}(1)/2$ with probability p (as the ranking function would rank each of the publishers randomly), or $\mathcal{D}(1)$ if the other player chose to write about topic 2, which happens with probability $1 - p$. Therefore the utility is $U(1) = p \cdot \mathcal{D}(1)/2 + (1 - p) \cdot \mathcal{D}(1)$. Similarly, the utility of writing about the second topic is $U(2) = p \cdot \mathcal{D}(2) + (1-p) \cdot \mathcal{D}(2)/2$. As both topics are in the support of the players' strategy, the utility of writing on each of them must be equal, i.e. $U(1) = U(2)$. This give us $p = \frac{2\mathcal{D}(1)-\mathcal{D}(2)}{\mathcal{D}(1)+\mathcal{D}(2)}$. Since both publishers play the same mixed strategy, the social welfare of the equilibrium is $P = 2 \cdot U(1) = 2 \cdot U(2) = \frac{3\mathcal{D}(1)\mathcal{D}(2)}{\mathcal{D}(1)+\mathcal{D}(2)}$. Finally, we can express the price of anarchy as

   $$\frac{OPT}{P} = \frac{\mathcal{D}(1) + \mathcal{D}(2)}{\frac{3\mathcal{D}(1)\mathcal{D}(2)}{\mathcal{D}(1)+\mathcal{D}(2)}} = \frac{(\mathcal{D}(1) + \mathcal{D}(2))^2}{3\mathcal{D}(1)\mathcal{D}(2)}.$$

   In the domain $0 \leq \mathcal{D}(1)/2 \leq \mathcal{D}(2) \leq \mathcal{D}(1) \leq 1 - \mathcal{D}(2)$, this function is bounded by $3/2$, which is reached on the line $\mathcal{D}(1) = 2\mathcal{D}(2)$.

4. One of the players does not have the second topic in the support of his strategy.
   Without loss of generality, we assume that player 2 is the one who never writes about topic 2. In this case, since we already covered all equilibria in which players restrict themselves only to the first two topics, the second player is must have the third topic in his support (or a different topic with the same query frequency). The first player then must earn at least $\mathcal{D}(2)$, as he would be the only publisher to writing on the second topic, should he choose to do so. Player 2 may randomize between the first and third topic, but since he is the only publisher to (occasionally) write on topic 3, his utility is $\mathcal{D}(3)$. Also, since the players can always earn $\mathcal{D}(1)/2$ by writing on the first topic, we get that $\mathcal{D}(2), \mathcal{D}(3) \geq \mathcal{D}(1)/2$. Putting it together, we

get a social welfare of $\mathcal{D}(2)+\mathcal{D}(3)$, which means price of anarchy of

$$\begin{aligned}\frac{OPT}{P} \le \frac{\mathcal{D}(1)+\mathcal{D}(2)}{\mathcal{D}(2)+\mathcal{D}(3)} &\le \frac{\mathcal{D}(1)+\mathcal{D}(2)}{\mathcal{D}(2)+\mathcal{D}(1)/2} \\ &= \frac{\mathcal{D}(1)+\mathcal{D}(2)}{2\mathcal{D}(2)/3+\mathcal{D}(2)/3+\mathcal{D}(1)/2} \\ &\le \frac{\mathcal{D}(1)+\mathcal{D}(2)}{2\mathcal{D}(2)/3+\mathcal{D}(1)/6+\mathcal{D}(1)/2} = \frac{3}{2}.\end{aligned}$$

5. Both players are playing an equilibrium in which the support of each contains topics 1 and 2. In this case, we argue that each player must earn at least as much as he did in equilibrium in which both players only write on the first two topics (case 3). Let $p = \frac{2\mathcal{D}(1)-\mathcal{D}(2)}{\mathcal{D}(1)+\mathcal{D}(2)}$ be the 2-topics symmetric equilibrium probability computed in 3. Notice that for an equilibrium $< s_1, s_2 >$, we get that the utilities received by the players are:

$$U_1 = \mathcal{D}(1)(1-s_2(1)/2) = \mathcal{D}(2)(1-s_2(2)/2),$$

$$U_2 = \mathcal{D}(1)(1-s_1(1)/2) = \mathcal{D}(2)(1-s_1(2)/2).$$

Now if $\{1,2\} \subsetneq SUP(s_i)$ then either $s_i(1) < p$ or $s_i(2) < 1-p$. This means that the profit for the *other* player has to increase. Since the arguments holds for both players, we get that each gains at least $\frac{3\mathcal{D}(1)\mathcal{D}(2)}{2(\mathcal{D}(1)+\mathcal{D}(2))}$ and the bound from 3 give us the required 3/2 bound on the price of anarchy.

Since we covered all possible equilibria, we established that the price of anarchy is 1.5. □

Appendix B. Proof of Theorem 4

We here prove the following:

Theorem *For any ranking function $\mathcal{R}$, there exists a distribution $\mathcal{D} \in \Delta([m])$, such that the Publishers Game $PG = (2, m, \mathcal{D}; \mathcal{R})$ has a price of anarchy of* at least 1.5.

Proof. Consider the simple scenario when queries about the first topic arrive with probability 2/3, while the second topic is queried 1/3 of the time; i.e., $\mathcal{D}(1) = 2/3, \mathcal{D}(2) = 1/3, \forall i \in [m] \setminus \{1,2\} : \mathcal{D}(i) = 0$. In the PRP-based game, $PG = (2, m, \mathcal{D}; R_{PRP})$, the players' payoffs are given by

$(1/3, 1/3)$	$(2/3, 1/3)$
$(1/3, 2/3)$	$(1/6, 1/6)$

.

The ranking function $\mathcal{R}$ is allowed to choose the first-rank probabilities when the players write on different topics. That is, the probability of ranking a non-relevant document first, rather than one written on the queried topic. This means that $\mathcal{R}$ may modify the game to

$(1/3, 1/3)$	(α, β)
(β, α)	$(1/6, 1/6)$

,

for any values of α, β satisfying $0 \le \alpha \le 2/3, 0 \le \beta \le 1/3$. In the new game, regardless of the values of α and β, there exists an equilibrium in which both players write on topic 1. In this equilibrium, each player payoff is 1/3, while the utility of deviating to the second topic is $\beta \le 1/3$. This means that the price of anarchy of $\mathcal{R}$ is at least $\frac{1}{2/3} = 1.5$. □

Appendix C. Proof of Theorem 5

Here, we prove the following theorem:

Theorem *For any distribution $\mathcal{D} \in \Delta([2])$ and ranking function $\mathcal{R}$, the social welfare of $(2, 2, \mathcal{D}; \mathcal{R})$ is at most the social welfare of $(2, 2, \mathcal{D}; R_{PRP})$.*

Proof. When R_{PRP} is used, the game is a special case of the one presented in Table 2, and is given by the matrix

$\left(\frac{\mathcal{D}(1)}{2}, \frac{\mathcal{D}(1)}{2}\right)$	$(\mathcal{D}(1), \mathcal{D}(2))$
$(\mathcal{D}(2), \mathcal{D}(1))$	$\left(\frac{\mathcal{D}(2)}{2}, \frac{\mathcal{D}(2)}{2}\right)$

.

If $\mathcal{D}(1) \geq 2/3$, the game has a unique equilibrium in which both players write about the first topic. This means that the social welfare of the game will be $\mathcal{D}(1)$. When $\mathcal{D}(1) \in [1/2, 2/3)$, the game has an optimal social welfare equilibrium, in which publishers write on different topics. The game also has a mixed strategies equilibrium, in which each writes about the first topic with probability $p = 3\mathcal{D}(1) - 1$. This equilibrium has a social welfare of $3\mathcal{D}(1)\mathcal{D}(2)$. Similarly to Theorem 4, the ranking function $\mathcal{R}$ may rank a document on a different topic before a relevant topic, hence it may modify the game to

$\left(\frac{\mathcal{D}(1)}{2}, \frac{\mathcal{D}(1)}{2}\right)$	(α, β)
(β, α)	$\left(\frac{\mathcal{D}(2)}{2}, \frac{\mathcal{D}(2)}{2}\right)$

,

for any values of α, β such that $0 \leq \alpha \leq \mathcal{D}(1), 0 \leq \beta \leq \mathcal{D}(2)$. In the case where $\alpha \geq \frac{\mathcal{D}(2)}{2}$ and $\beta \geq \frac{\mathcal{D}(1)}{2}$, this game have a symmetric mixed strategy equilibrium where each player writes about the first topic with probability $p \triangleq \frac{2\alpha+\mathcal{D}(1)-1}{2\alpha+2\beta-1}$. In all other cases, the game has only pure-strategy equilibrium where both surely write about topic 1. The pure strategy equilibrium gives a social welfare of $\mathcal{D}(1)$. This is strictly lower than the $3\mathcal{D}(1)\mathcal{D}(2)$ achieved by R_{PRP} in the $\mathcal{D}(1) \in [1/2, 2/3)$ case, and identical to it when $\mathcal{D}(1) \geq \frac{2}{3}$. The mixed strategies equilibrium gives a social welfare of $\mathcal{D}(1) \cdot p + (\alpha + \beta) \cdot (1 - p) = \frac{4\alpha\beta - \mathcal{D}(1)\mathcal{D}(2)}{2\alpha+2\beta-1}$. Fixing $\mathcal{D}(1)$ and $\mathcal{D}(2)$, analysis of the social welfare as a function of α and β reveals it is maximized at $\alpha = \mathcal{D}(1), \beta = \mathcal{D}(2)$. This shows that in any case, no ranking function outperforms R_{PRP}. □

Appendix D. Proof of Theorem 6

In this appendix we prove the following:

Theorem *Let $n, m \in \mathbb{N}^+$, $c \in (0, 1) \setminus \{0.5\}$, and $\mathcal{D} \in \Delta([m])$, then MTPG= $(n, m, c, \mathcal{D}; R_{PRP})$ is not a congestion game.*

Proof. To prove the theorem we use a result from (Monderer & Shapley, 1996) (Theorem 2.8) that shows that a game Γ is a congestion game if and only if for every two players i, j, actions $a_i, a_i' \in S_i$, $a_j, a_j' \in S_j$ and action profile for the remaining players a_{-ij} we have:

$$\begin{aligned} &U_i(a_i', a_j, a_{-ij}) - U_i(a_i, a_j, a_{-ij}) \\ +&U_j(a_i', a_j', a_{-ij}) - U_j(a_i', a_j, a_{-ij}) \\ +&U_i(a_i, a_j', a_{-ij}) - U_i(a_i', a_j', a_{-ij}) \\ +&U_j(a_i, a_j, a_{-ij}) - U_j(a_i, a_j', a_{-ij}) = 0. \end{aligned}$$

Consider the action profile in which every player other than $1, 2$ writes on topic 3.[8] Next, consider $a_1 = a_2 = 1$ and $a'_1 = 2, a'_2 = \{1, 2\}$. By switching from a_1 to a'_1, the utility of the first player changes by

$$U_1(a'_1, a_2, a_{-1,2}) - U_1(a_1, a_2, a_{-1,2}) = 1 - 3\mathcal{D}(1)/2.$$

Next, player 2 changes from writing on topic 1 to a mixed document, thereby changing its utility by

$$U_2(a'_1, a'_2, a_{-1,2}) - U_2(a'_1, a_2, a_{-1,2}) = c \cdot \mathcal{D}(1) - \mathcal{D}(1).$$

Similarly, the next two changes impact the utility functions by

$$U_1(a_1, a'_2, a_{-1,2}) - U_1(a'_1, a'_2, a_{-1,2}) = 2\mathcal{D}(1) - 1,$$

and

$$U_2(a_1, a_2, a_{-1,2}) - U_2(a_1, a'_2, a_{-1,2}) = \mathcal{D}(1)/2 - c \cdot (1 - \mathcal{D}(1)).$$

Summing over the four utility changes we get $2c \cdot \mathcal{D}(1) - \mathcal{D}(1)$, which is non-zero for $c \neq 1/2$. □

Appendix E. Proof of Theorem 7

We now prove the following theorem:

Theorem *Let $c \in [0, 1/2]$. For any distribution $\mathcal{D} \in \Delta([m])$, the Multiple Topics Publishers Game $MTPG = (2, m, c, \mathcal{D}; R_{PRP})$ achieves a price of anarchy of at most 1.5.*

Proof. Let $s_2 \in \Delta([m])$ be a mixed strategy for the second player and let $j, \ell \in [m]$ be two distinct topics. The first publisher payoffs for writing on the topics are:

$$U_1(j, s_2) = \mathcal{D}(j) \cdot (1 - s_2(j)/2)\,,$$
$$U_1(\ell, s_2) = \mathcal{D}(\ell) \cdot (1 - s_2(\ell)/2)\,.$$

Notice that if player 1 chooses to write on the multi-topic document $\{j, \ell\}$, she will be ranked first on each of these topics unless the other player also writes about the queried topic. Now if player 2 wrote on topic j, for example, then player 1 can only be ranked first on topic ℓ. If the second player also wrote on a mix, the identity of the document that is ranked first, for each of these topics, will be determined by a coin flip. Therefore, the utility of writing the multi topic document is

$$\begin{aligned}
U_1(\{j, \ell\}, s_2) =& c \cdot \Big(\mathcal{D}(j) \cdot (1 - s_2(\{j, \ell\})/2 - s_2(j)) + \mathcal{D}(\ell) \cdot (1 - s_2(\{j, \ell\})/2 - s_2(\ell))\Big) \\
=& c \cdot \Big(\mathcal{D}(j) \cdot (1 - s_2(j)/2) - \mathcal{D}(j) \cdot (s_2(\{j, \ell\})/2 + s_2(j)/2) \\
& + \mathcal{D}(\ell) \cdot (1 - s_2(\ell)/2) - \mathcal{D}(\ell) \cdot (s_2(\{j, \ell\})/2 + s_2(\ell)/2)\Big) \\
=& c \cdot \Big(U_1(j, s_2) + U_1(\ell, s_2)\Big) \\
& - c \cdot \Big(\mathcal{D}(j) \cdot (s_2(\{j, \ell\})/2 + s_2(j)/2) + \mathcal{D}(\ell) \cdot (s_2(\{j, \ell\})/2 + s_2(\ell)/2)\Big). \quad (1)
\end{aligned}$$

Now if we assume that writing about a mix of topics j and ℓ is a part of player 1's equilibrium strategy, we must have:

$$U_1(\{j, \ell\}, s_2) \geq U_1(j, s_2) \qquad \text{and}$$
$$U_1(\{j, \ell\}, s_2) \geq U_1(\ell, s_2).$$

8. For simplicity we assume $m \geq 3$, the result holds also for $m = 2$.

Putting the two inequalities together, we also get that

$$U_1(\{j, \ell\}, s_2) \geq (U_1(j, s_2) + U_1(\ell, s_2))/2.$$

Combining this with (1), we get that:

1. $c = 1/2$, as the subtrahend of (1) is non-negative and $c \leq 1/2$.

2. The subtrahend itself must be zero, i.e.,

$$c \cdot \Big(\mathcal{D}(j) \cdot (s_2(\{j,\ell\})/2 + s_2(j)/2) + \mathcal{D}(\ell) \cdot (s_2(\{j,\ell\})/2 + s_2(\ell)/2)\Big) = 0.$$

3. As the player can always ensure a positive payoff by writing on the most queried topic, we have that the probability of getting a query about j *or* ℓ, $\mathcal{D}(j) + \mathcal{D}(\ell)$, must be strictly positive. Now if we assume by contradiction that one of the topics has probability zero, writing about the second topic is a dominating strategy over writing a multi-topic document, as it will only give us half the utility. We can thus conclude that *both* $\mathcal{D}(j), \mathcal{D}(\ell)$ must be strictly positive.

4. Combining the previous two points with the fact that s_2 is a probability distribution, we get that $s_2(j) = s_2(\ell) = s_2(\{j,\ell\}) = 0$.

5. The player's payoff is $U_1(\{j,\ell\}, s_2) = (U_1(j, s_2) + U_1(\ell, s_2))/2$, while $U_1(j, s_2) = \mathcal{D}(j)$ and $U_1(\ell, s_2) = \mathcal{D}(\ell)$. This allows us to conclude that $U_1(j, s_2) = U_1(\ell, s_2)$.

This means that player 1 has the same payoff as of writing about a single topic (let it be j or ℓ). We can thus assume that without loss of generality, to compute a lower bound on the social welfare, it never chooses to write about a topic mix. Finally, we can repeat this argument for player 2 and conclude that no equilibrium in which the players write about mixed topics yield lower social-welfare than equilibrium in which they only write about pure topics. This allows us to use the result of Theorem 3 and conclude that the price of anarchy is at most 1.5. ☐

Appendix F. Proof of Theorem 8

We now prove the following:

Theorem *Let* $c \in [0, 1/3]$. *There exists a distribution* $\mathcal{D} \in \Delta([m])$, *such that for* any *ranking function* $\mathcal{R}$, *the Multiple Topics Publishers Game* $MTPG = (2, m, c, \mathcal{D}; \mathcal{R})$ *has a price of anarchy of* at least 1.5.

Proof. Consider the query distribution $\mathcal{D}$ in which $\mathcal{D}(1) = 2/3$, $\mathcal{D}(2) = 1/3$, and $\forall j \in [m] \setminus \{1, 2\}$: $\mathcal{D}(j) = 0$. Consider the case in which both players chose to write on the first topic. This means that the payoff for each will be 1/3. We will show that this is an equilibrium. Clearly, neither player has an incentive to deviate to topic 2, as they could make at most 1/3 by doing so. We will analyze the case in which player 1 deviates to the mixed topic $\{1, 2\}$; similar analysis holds for player 2. By such deviation, the payoff for player 1 becomes

$$\begin{aligned} U_1(\{1,2\}, 1) &= c \cdot (\mathcal{D}(1) R_1(1, \{1,2\}, 1) + \mathcal{D}(2) R_1(2, \{1,2\}, 1)) \\ &= c/3 \cdot (2R_1(1, \{1,2\}, 1) + R_1(2, \{1,2\}, 1)) \leq c \leq 1/3. \end{aligned}$$

This means that no player incentive to deviate from $\langle 1, 1 \rangle$, whose social welfare is 2/3. The optimal social welfare is 1 and is achieved when both write about different topics, and thus the price of anarchy is at least 1.5. ☐

Appendix G. Proof of Example 1

We now prove the following:

Example *Consider the distribution $\mathcal{D}(1) = 2/3, \mathcal{D}(2) = 1/3$. For any c, the price of anarchy of $MTPG_{PRP} = (2, 2, c, \mathcal{D}; R_{PRP})$ is 1.5, while there exists a ranking function, R_{prob}, such that the price of anarchy of $MTPG_{prob} = (2, 2, c, \mathcal{D}; R_{prob})$ could be arbitrarily close to 1 as $c \to 1$.*

Proof. The payoff matrix of the $MTPG_{PRP}$ game is

$(1/3, 1/3)$	$(2/3, c/3)$	$(2/3, 1/3)$
$(c/3, 2/3)$	$(c/2, c/2)$	$(2c/3, 1/3)$
$(1/3, 2/3)$	$(1/3, 2c/3)$	$(1/6, 1/6)$

,

where for each player the first action is associated with writing a document on topic 1, the second is writing a multiple topic document, and the last action is writing on the second topic. For any value of c, there exists an equilibrium where both players write on topic 1. This equilibrium has a social welfare of 2/3, while an optimum of 1 is reached if both players write on different topics. This means that for any value of c, the price of anarchy of R_{PRP} is 1.5.

For $c > 1/2$, consider the ranking function R_{prob}, which is identical to R_{PRP}, except for the case where one publisher wrote about topic 1, the other about a multiple topics document, and a query on the first topic was posted. In this case, R_{prob} ranks the multi-topic document first with probability $1 - \frac{3c}{2(1+c)}$. The probabilistic ranking changes the utility of the players which are aiming to maximize their expected payoff, and the resulting payoff matrix is

$(1/3, 1/3)$	$\left(\frac{c}{1+c}, \frac{c}{1+c}\right)$	$(2/3, 1/3)$
$\left(\frac{c}{1+c}, \frac{c}{1+c}\right)$	$(c/2, c/2)$	$(2c/3, 1/3)$
$(1/3, 2/3)$	$(1/3, 2c/3)$	$(1/6, 1/6)$

.

This game has an optimal pure-strategies equilibria when players write on different documents, and a single mixed strategies equilibrium. In this equilibrium, each player creates a document about the first topic with probability $\frac{3(c-1)c}{3c^2-7c+2}$, and a multi topic document otherwise. The resulting social welfare in this case is $\frac{2c(c^2-4c+1)}{(c+1)(3c^2-7c+2)}$, which approaches 1 when c is almost 1. □

Appendix H. Proof of Theorem 9

Here, we prove the theorem below.

Theorem *Let $c \in (1/2, 1]$ and $p \in [2/3, \frac{2c}{c+1})$. Consider the distribution $\mathcal{D}(1) = p, \mathcal{D}(2) = 1 - p$. There exists a ranking function R_{prob}, such that the price of anarchy of $(2, 2, c, \mathcal{D}; R_{PRP})$ is strictly larger than that of $(2, 2, c, \mathcal{D}; R_{prob})$.*

Proof. In $MTPG_{PRP} = (2, 2, c, \mathcal{D}; R_{PRP})$, there exists a dominant-strategy equilibrium in which both players write on the first topic. Playing it, the publishers earn $U(1) = p/2$, while the utility of writing about multiple topics is $U(\{1, 2\}) = c \cdot (1 - p) \le 1 - p \le p/2$, and the utility of writing about the second topic is $U(2) = 1 - p \le p/2$. This means that the social welfare achieved by R_{PRP} is p. Next, we define R_{prob} as follows:

- Consider a query on the first topic, a document written on the topic, and a multi-topic document. In this case, R_{prob} will rank the multi-topic document first with probability $r \triangleq 1 - \frac{c}{p\cdot(1+c)}$. Notice that since $c \in (1/2, 1]$ and $p \in [2/3, \frac{2c}{c+1})$, we have $r \in [\frac{1}{4}, \frac{1}{2}]$ and this is a legal ranking function.

- On all other cases, R_{prob} behaves just like R_{PRP}.

The introduction of probabilistic ranking in R_{prob} changes the game payoffs when one player writes on topic 1, while the other on multiple topics. In this case, the utility of writing on topic 1 drops to $U(1,\{1,2\}) = p(1-r) = \frac{c}{1+c}$. Similarly, the utility of writing a mix rises- $U(\{1,2\},1) = c(1-p+p\cdot r) = \frac{c}{1+c}$. The game payoff matrix is therefore:

$(p/2, p/2)$	$\left(\frac{c}{1+c}, \frac{c}{1+c}\right)$	$(p, 1-p)$
$\left(\frac{c}{1+c}, \frac{c}{1+c}\right)$	$(c/2, c/2)$	$(c \cdot p, 1-p)$
$(1-p, p)$	$(1-p, c \cdot p)$	$\left(\frac{1-p}{2}, \frac{1-p}{2}\right)$

.

First, we notice that the scenario in which both players write on the first topic is no longer an equilibrium, as players better off deviating to the multi-topic document, earning $\frac{c}{1+c} > p/2$. Next, the strategy of writing on topic 1 still dominates creating a document about the second topic. Finally, we observe that the pure-strategies equilibria ($(1,2)$ or $(2,1)$) gives a social welfare of $\frac{2c}{1+c} > p$. The only mixed strategies equilibrium in this game is reached when each player writes about topic 1 with probability $x = \frac{c\cdot(c-1)}{c^2+c\cdot(p-3)+p}$ and otherwise writes a multi-topic document. This gives a social welfare of

$$\begin{aligned} P = 2U(2) &= 2x \cdot \frac{c}{1+c} + (1-x)\cdot c \\ &= \frac{c}{1+c} \cdot \frac{c^2p + 2cp - 4c + p}{c^2 + c\cdot(p-3) + p}, \end{aligned}$$

which on the domain of $c \in (1/2, 1]$ and $p \in [2/3, \frac{2c}{c+1})$ is strictly larger than p. □

Appendix I. Proof of Lemma 1

We now prove the following:

Lemma *Let $\mathcal{Q} \in [0,1]^{[2]\times[m]}$ whose entries are positive and distinct, then $DQPG = (2, m, \mathcal{Q}; R_{PRP})$ is not a congestion game.*

Proof. Assume the qualities of the two players on the first two topics are $Q_{1,1} = a$, $Q_{1,2} = b$, $Q_{2,1} = c$ and $Q_{2,2} = d$. Without loss of generality, assume that

$$c < a \leq 1.$$

We continue with a case analysis:

- $\mathbf{b > d}$.
  In this case, player 1 will be ranked first on the topic she choose to write on, regardless of the action taken by the second player. This means that the players' utility satisfy

$$\begin{aligned} &U_1(1,1) = a, U_2(1,1) = 0, \\ &U_1(1,2) = a, U_2(1,2) = d, \\ &U_1(2,1) = b, U_2(2,1) = c, \\ &U_1(2,2) = b, U_2(2,2) = 0. \end{aligned}$$

Next, consider the actions $\forall i \in \{1,2\} : a_i = 1, a'_i = 2$. Similarly to the proof of Theorem 6, we get that

$$\begin{aligned} &U_1(a'_1, a_2) - U_1(a_1, a_2) \\ +&U_2(a'_1, a'_2) - U_2(a'_1, a_2) \\ +&U_1(a_1, a_2) - U_1(a'_1, a'_2) \\ +&U_2(a_1, a_2) - U_2(a_1, a'_2) = d + c \neq 0. \end{aligned}$$

- $\mathbf{b} < \mathbf{d}$.
  In this case, player 1 will be ranked first on the first topic if he writes about it, but will only rank second if both write about topic 2. The relevant utility quantities then become:

$$\begin{aligned} U_1(1,1) = a, U_2(1,1) = 0, \\ U_1(1,2) = a, U_2(1,2) = d, \\ U_1(2,1) = b, U_2(2,1) = c, \\ U_1(2,2) = 0, U_2(2,2) = d. \end{aligned}$$

  Once again, consider $\forall i \in \{1,2\} : a_i = 1, a'_i = 2$. This gives us

$$\begin{aligned} &U_1(a'_1, a_2) - U_1(a_1, a_2), \\ +&U_2(a'_1, a'_2) - U_2(a'_1, a_2), \\ +&U_1(a_1, a_2) - U_1(a'_1, a'_2), \\ +&U_2(a_1, a_2) - U_2(a_1, a'_2) = c - b \neq 0. \end{aligned}$$

We have established that in all cases DQPG is not a congestion game. □

Appendix J. Proof of Theorem 11

In this appedix, we prove the following:

Theorem *Let $\mathcal{Q} \in [0,1]^{[n]\times[m]}$ be a Quality Matrix such that no two players have the same writing quality on the same topic and no player has equal writing quality on two different topics. Then the Different Qualities Publishers Game DQPG $= (n, m, \mathcal{Q}; R_{PRP})$ achieves a price of anarchy of at most* 2.

Proof. For simplicity, the proof assumes $n \leq m$, although this is not required, as otherwise the utility of $n - m$ players is 0, and they may be removed to obtain a valid matching as we show below. We start by proving that in this case, there exists only a single equilibrium. This is done using *dominant strategies elimination* and induction over the number of players.

- **Basis:** In a single player game, the only publisher will surely write on the topic for which she is most qualified. Since the writing quality for every topic is distinct, there is a single topic which yields the greatest payoff.
- **Induction Step:** Let $i, j \in \text{argmax}\,\{Q_{i,j}\}$ (notice that there may be several such pairs); then player i can profit $Q_{i,j}$ by writing on topic j, and will surely gain less by choosing a different topic, since $Q_{i,j} > Q_{i,k}$ for every other topic k. Thus, the player will surely write on this topic while no other player would, as player i will surely rank first because for every other player ℓ, $Q_{i,j} > Q_{\ell,j}$ and thus ℓ will never rank first for topic j. Using the induction hypothesis, we get that for the residual game in which player i and topic j are removed, there exists a single equilibrium. Finally, since player i is sure to write on topic j, we conclude that only a single equilibrium exists.

Next, we prove the theorem by showing that the equilibrium is associated with a matching in the bipartite graph which has the publishers on one partite and the topics on the other. Formally, let $G = ([n] \cup [m], [n] \times [m])$ be a full bipartite graph. Let $w : [n] \times [m] \to [0, 1]$ be a weight function such that $w(i, j) = Q_{i,j}$. We prove that an equilibrium for the game corresponds to a matching that could be generated by the greedy algorithm for graph matching. The algorithm iteratively expands the matching by simply adding the maximum-weight edge connecting unmatched vertices to the matching. We complete the proof by induction over the number of players in the game.

- **Basis:** In a single publisher game, the only equilibrium is when the player writes a document on the topic j maximizing $w(1, j)$, and so does the greedy matching.
- **Induction Step:** Let $i, j \in \operatorname{argmax}\{Q_{i,j}\}$ be a top-quality publisher-topic pair. Since no other player has the same writing quality on topic j, player i writes about j, knowing he will surely rank first. The remaining players then know that no gain would come from writing on topic j, hence we are left with a game on $n-1$ players and $m-1$ topics. In the graph, this pair represents a maximum weight edge and therefore may be selected by the greedy algorithm. Using the induction hypothesis, we conclude the remaining players will also follow the greedy matching algorithm.

We conclude the proof by using the well-known result stating that the greedy algorithm achieves a 2-approximation for the maximum weight matching of the graph. □

Appendix K. Proof of Theorem 12

We now prove the following theorem:

Theorem *For any $\epsilon > 0$ and $n, m \in \mathbb{N}$, there exists a Quality Matrix $\mathcal{Q}^\epsilon$ whose entries are positive and distinct, such that the price of anarchy of $(n, m, \mathcal{Q}^\epsilon; R_{PRP})$ is $2 - \epsilon$.*

Proof. Denote $q \triangleq \frac{\epsilon}{2n(n+1)}$ and use $I(i, j)$ as an indicator for whether $i \neq j$, i.e.,

$$I(i,j) \triangleq \begin{cases} 1 & \text{if } i \neq j \\ 0 & \text{otherwise} \end{cases}.$$

Consider the following Quality Matrix:

players/topics	1	2	$\ldots$	j	$\ldots$	m
1	1	$1-\epsilon/6$				
2	$1-\epsilon/3$					
$\vdots$			$\ddots$			
i				$q \cdot i \cdot 2^{-j \cdot n \cdot m \cdot I(i,j)}$		
$\vdots$					$\ddots$	
n						

Every entry i, j in Quality Matrix except for $(1, 1), (1, 2), (2, 1)$ has a value of $q \cdot i \cdot 2^{-j \cdot n \cdot m \cdot I(i,j)}$.

Notice that in our matrix all entries are positive and distinct. First, observe that there exists a strategy profile with a social welfare of at least $2 - \epsilon/2$. We choose the matrix value such that every player i has a dominant strategy of writing about topic i (if $m < n$, then players $m + 1, \ldots, n$ are never ranked first on any topic and do not contribute to the social welfare). Thus, the *only* equilibrium in our game provides a social welfare of at most

$$1 + \sum_{i=2}^{n} q \cdot i < 1 + q \cdot n(n + 1)/2 = 1 + \epsilon/4.$$

Thus, we get that the social welfare of the game satisfies

$$\frac{OPT}{P} > \frac{2 - \epsilon/2}{1 + \epsilon/4} = 2 - \frac{\epsilon}{1 + \epsilon/4} > 2 - \epsilon.$$

□

Appendix L. Proof of Theorem 13

Here, we give a proof for the following theorem:

Theorem *There exists a Fair Ranking Function R_{uni}, such that for any Quality Matrix $\mathcal{Q} \in [0,1]^{[2]\times[2]}$, the Different Qualities Publishers Game $DQPG = (2, 2, \mathcal{Q}; R_{uni})$ achieves a price of anarchy of at most* 1.683*.*

Proof. Since price of anarchy analysis concerns with the *ratio* between the optimal strategy and an equilibrium strategy, it stays the same if $\mathcal{Q}$ is multiplied by a constant. Therefore, we assume that $\mathcal{Q}_{1,1} = 1$, i.e., the first publisher writes about the first topic perfectly. Therefore, a general Quality Matrix can be described using three parameters, $a, b, c \in [0, 1]$:

players/topics	1	2
1	1	a
2	b	c

The Quality Matrix $\mathcal{Q}$.

Notice that these notation mean that the optimal solution for price of anarchy analysis is $OPT = \max\{1 + c, a + b\}$. Denote by r the real root of $f(r) = r^3 + r - 1$; we define the ranking function R_{uni} as the function which gives equal probability to all documents written on the topic whose quality is at least an r fraction of the best document's quality. We note that $r \approx 0.682$ satisfies $r \geq 0.5$ and thus such a ranking is indeed a Fair Ranking Function. We continue the proof by an exhaustive case-analysis showing that for all possible values of a, b, c, the price of anarchy is at most $1 + r$. We denote the social welfare achieved by R_{uni} by P.

1. Case $a < 1/2$.
   Notice that in this case, player 1 will always prefer writing on topic 1, as R_{uni} will rank him first with probability of at least 1/2. Next, we consider two sub-cases:

   a) $b < r$.
   By the definition of R_{uni}, player 1 will always be ranked first on query 1. Therefore, player two could only earn c by writing on topic 2, and thus the social welfare achieved is $P = 1+c$. Assume $\max\{1 + c, a + b\} = a + b$, then:

   $$\frac{OPT}{P} = \frac{a + b}{1 + c} < \frac{1/2 + r}{1 + c} < 1 + r.$$

   b) $b \geq r$.
   For analyzing this case, we examine two sub-cases:

      I. $b \geq 2c$.
      This scenario leads player 2 to write about topic 1 (or a mixed strategy in case of $b = 2c$). This means that the social welfare will be $P \geq \frac{1+b}{2}$. Once again, we consider two sub-cases of this case as well:

- $\max\{1+c, a+b\} = a+b$.
In this case we get

$$\begin{aligned}\frac{OPT}{P} &= \frac{a+b}{\frac{1+b}{2}} = \frac{2(a+b)}{1+b} \leq_{(a<1/2)} \frac{1+2b}{1+b} \\ &= 1 + \frac{b}{1+b} \leq_{(b\leq 1)} 1.5 \leq 1+r.\end{aligned}$$

- $\max\{1+c, a+b\} = 1+c$.
In this case we get

$$\begin{aligned}\frac{OPT}{P} &= \frac{1+c}{\frac{1+b}{2}} = \frac{2\,(1+c)}{1+b} \leq_{(c\leq 1/2b)} \frac{2+b}{1+b} \\ &= 1 + \frac{1}{1+b} \leq_{(b>r)} \frac{2+r}{1+r}.\end{aligned}$$

Notice that our selection of r satisfies $\frac{2+r}{1+r} < 1+r$ and thus we obtain $\frac{OPT}{P} \leq 1+r$.

II. $b < 2c$.
For these parameters, player 2 will be writing on topic 2 as he will only rank first topic 1 with probability 1/2. Our profit is therefore $P = 1+c$. If $OPT = a+b$, then

$$\begin{aligned}\frac{OPT}{P} &= \frac{a+b}{1+c} \leq_{(b<2c, a<1/2)} \frac{1/2+2c}{1+c} \\ &= 1 + \frac{c-1/2}{1+c} \leq_{(c\leq 1)} 1.25 \leq 1+r.\end{aligned}$$

2. $b < r$.
This means that player 2 will surly write on topic 2 since player 1 will write on the first topic, knowing he will be ranked first. This means that the social welfare is $P = 1+c$. If $OPT = a+b$ then we have

$$\frac{OPT}{P} = \frac{a+b}{1+c} \leq_{(a\leq 1, b<r)} \frac{1+r}{1+c} \leq 1+r.$$

3. $a > 1/2, b > r$ and $c < b/2$.
In this case, player 2 knows he gets ranked first on topic 1 with probability of at least 1/2 and therefore chooses to write about topic 1. Player 1, on the other hand, anticipates player 2 action and therefore writes on topic 2. The social welfare is then $P = a+b$. If $OPT = 1+c$ then

$$\begin{aligned}\frac{OPT}{P} &= \frac{1+c}{a+b} \leq_{\left(\substack{c<b/2,\\ a>1/2}\right)} \frac{1+b/2}{1/2+b} \\ &\leq_{(b>r>1/2)} \frac{1+1/2}{1/2+1/2} = 1.5 < 1+r.\end{aligned}$$

4. $a > 1/2, b > r, c > b/2, c < r\cdot a$.
In this case, if both players write on topic 1 a coin flip will decide the ranking. For topic 2, the ranking will place player 1 first. The payoff matrix of the players is therefore:

$(1/2, b/2)$	$(1, c)$
(a, b)	$(a, 0)$

This game has a single mixed-strategy equilibrium in which the first player writes about topic 1 with probability $x \triangleq \frac{b}{c+b/2}$, while player two writes about the first topic with probability $y \triangleq 2(1-a)$. The expected utility for player 1 is then $U_1 = a$, while the second player payoff is $U_2 = x \cdot c = \frac{b \cdot c}{c+b/2}$. The expected social welfare is then $P = U_1 + U_2 = a + \frac{b \cdot c}{c+b/2}$. We consider the two possible cases for OPT:

- $OPT = a + b$.
  In this case we can bound the price of anarchy as follows:

$$\begin{aligned}\frac{OPT}{P} &= \frac{a+b}{a+\frac{b\cdot c}{c+b/2}} \leq_{(c>b/2)} \frac{a+b}{a+\frac{b^2/2}{b}} \\ &= \frac{a+b}{a+b/2} \leq_{\binom{a>1/2,}{b<1}} 1.5 < 1+r.\end{aligned}$$

- $OPT = 1 + c$.
  In this case we get:

$$\begin{aligned}\frac{OPT}{P} &= \frac{1+c}{a+\frac{b\cdot c}{c+b/2}} \leq_{(a>c/r)} \frac{1+c}{c/r+\frac{b\cdot c}{c+b/2}} \\ &\leq_{(b>r)} \frac{1+c}{c/r+\frac{c\cdot r}{c+r/2}} \\ &= \frac{r\cdot(1+c)\cdot(2c+r)}{2c^2+2cr^2+cr} \\ &\leq_{(r<0.7)} \frac{0.7\cdot(1+c)\cdot(2c+0.7)}{2c^2+2cr^2+cr} \\ &\leq_{(r>0.68)} \frac{0.7\cdot(1+c)\cdot(2c+0.7)}{2c^2+2c\cdot 0.68+c\cdot 0.68} \\ &= \frac{0.7c^2+0.945c+0.245}{c^2+0.8024c}.\end{aligned}$$

  The function

$$f(c) \triangleq \frac{0.7c^2+0.945c+0.245}{c^2+0.8024c}$$

  is monotonically decreasing for $c > 0$; we have $c > b/2 > r/2 > 0.34$ and thus we conclude that $\frac{OPT}{P} \leq f(0.34) \approx 1.666 < 1+r$.

5. $a > 1/2, b > r, c > b/2, \quad r \cdot a < c < a/r$.
   For these parameters, R_{uni} will rank first each player with probability 1/2 on the second topic. This means that the payoff matrix for the players is:

$(1/2, b/2)$	$(1, c)$
(a, b)	$(a/2, c/2)$

   The mixed-strategy equilibrium of this game is reached when the first players writes about topic 1 with probability $x = \frac{2b-c}{b+c}$, while player 2 writes about it with probability $y = \frac{2-a}{a+1}$. The expected social welfare is then

$$\begin{aligned}P &= U_1 + U_2 = 1 - y/2 + b(1 - x/2) \\ &= \frac{3}{2}\cdot\left(\frac{a}{a+1}+\frac{bc}{b+c}\right).\end{aligned}$$

   We now consider the two possible values of OPT:

- $OPT = a + b$.
In this case we can bounded the price of anarchy as follows:
$$\begin{aligned}\frac{OPT}{P} &= \frac{a+b}{\frac{3}{2}\cdot\left(\frac{a}{a+1}+\frac{bc}{b+c}\right)} \leq_{(c>b/2)} \frac{a+b}{\frac{3}{2}\cdot\left(\frac{a}{a+1}+\frac{b^2\cdot 1/2}{b+b/2}\right)} \\ &= \frac{a+b}{\frac{3a}{2(a+1)}+\frac{b}{2}}.\end{aligned}$$
The function $g(a,b) \triangleq \frac{a+b}{\frac{3a}{2(a+1)}+\frac{b}{2}}$ is monotonically increasing in both a and b. Since $a, b \leq 1$ we conclude that $\frac{OPT}{P} \leq g(1,1) = 1.6 < 1 + r$.
- $OPT = 1 + c$.
We have to consider these sub cases here:
  - ⋆ $c < \frac{1}{2r}$.
  In which cases we bound the price of anarchy:
  $$\begin{aligned}\frac{OPT}{P} &= \frac{1+c}{\frac{3}{2}\cdot\left(\frac{a}{a+1}+\frac{bc}{b+c}\right)} \\ &\leq_{(a>1/2)} \frac{1+c}{\frac{3}{2}\cdot\left(1/3+\frac{bc}{b+c}\right)} \\ &= \frac{2\cdot(1+c)\cdot(b+c)}{3bc+b+c}.\end{aligned}$$
  We define the function $h(b,c) \triangleq \frac{2\cdot(1+c)\cdot(b+c)}{3bc+b+c}$. Over the domain
  $$\mathcal{D}_h \triangleq \left\{(b,c) \mid r \leq b \leq 1, b/2 \leq c \leq \frac{1}{2r}\right\},$$
  the function h satisfies
  $$\forall (b,c) \in \mathcal{D}_h : h(b,c) \leq h(r, 1/2r) = \frac{2r^2+1}{r^2+r}.$$
  Our choice of r suffice $\frac{2r^2+1}{r^2+r} = 1 + r$, and thus $\frac{OPT}{P} \leq 1 + r$.
  - ⋆ $c \geq \frac{1}{2r}$.
  Similarly, we bound the price of anarchy:
  $$\begin{aligned}\frac{OPT}{P} &= \frac{1+c}{\frac{3}{2}\cdot\left(\frac{a}{a+1}+\frac{bc}{b+c}\right)} \leq_{(a>r\cdot c)} \frac{1+c}{\frac{3}{2}\cdot\left(\frac{r\cdot c}{r\cdot c+1}+\frac{bc}{b+c}\right)} \\ &= \frac{2\cdot(1+c)}{\frac{3r\cdot c}{1+r\cdot c}+\frac{3bc}{b+c}}.\end{aligned}$$
  We then define the function $j(b,c) \triangleq \frac{2\cdot(1+c)}{\frac{3r\cdot c}{1+r\cdot c}+\frac{3bc}{b+c}}$. Over the domain
  $$\mathcal{D}_j \triangleq \left\{(b,c) \mid r \leq b \leq 1, \frac{1}{2r} \leq c \leq 1\right\},$$
  the function j satisfies
  $$\forall (b,c) \in \mathcal{D}_j : j(b,c) \leq j(r, 1/2r) = 1 + r,$$
  and thus $\frac{OPT}{P} \leq 1 + r$.

6. $a > 1/2, b > r, c > b/2, \quad c > a/r$.
   In this case, R_{uni} will prefer player 2 over player 1 if both write on topic 2. The payoff matrix of the game is therefore:

$(1/2, b/2)$	$(1, c)$
(a, b)	$(0, c)$

   .

   In this case the mixed-strategy equilibrium is reached when player 1 writes on the first topic with probability $x = \frac{2(b-c)}{b}$, and player 2 with probability $y = \frac{1}{a+1/2}$. The resulting social welfare is then

   $$\begin{aligned} P = U_1 + U_2 &= ay + c \\ &= \frac{a}{a+1/2} + c. \end{aligned}$$

   We complete the proof by considering the two sub cases:

   - $OPT = a + b$.
     We bound the the price of anarchy:

     $$\begin{aligned} \frac{OPT}{P} = \frac{a+b}{\frac{a}{a+1/2} + c} &\leq_{(c>a/r)} \frac{a+b}{\frac{a}{a+1/2} + a/r} \\ &\leq_{(b\leq 1)} \frac{a+1}{\frac{a}{a+1/2} + a/r} \leq_{(a>1/2)} 1.5 < 1 + r. \end{aligned}$$

   - $OPT = 1 + c$.
     Here, we get the bound as follows;

     $$\frac{OPT}{P} = \frac{1+c}{\frac{a}{a+1/2} + c} \leq_{(a>1/2)} \frac{1+c}{1/2 + c} \leq_{(c>a/r>1/3)} 1.6 < 1 + r.$$

As our case analysis covers all possible values for a, b and c, this concludes the proof. □

Appendix M. Proof of lemma 2

We now prove the following lemma:

Lemma *There exists a Fair Ranking Function R_{uni}, such that for any Quality Matrix $\mathcal{Q} \in [0,1]^{[2]\times[2]}$, the social welfare of every safety-level strategy profile in $DQPG = (2, 2, \mathcal{Q}; R_{uni})$is at least $\frac{1}{1.683}$ times the social welfare of any strategy profile.*

Proof. We prove the lemma by an extensive case analysis over all possible quality matrices. Let $\mathcal{Q} \in [0,1]^{[2]\times[2]}$ be the Quality Matrix of the game:

players/topics	1	2
1	$Q_{1,1}$	$Q_{1,2}$
2	$Q_{2,1}$	$Q_{2,2}$

The Quality Matrix $\mathcal{Q}$.

We consider the ranking function R_{uni} described in Theorem 13, which uniformly chooses a top-rank document when the quality ratio of the competing documents is at least r, where r is the real root of $f(r) = r^3 + r - 1$. We note that $r \approx 0.682$ satisfies $r \geq 0.5$ and thus such a ranking is indeed a Fair

Ranking Function. Without loss of generality, we assume that the optimal social welfare is achieved when player 1 writes on topic 1 and the second player on the other topic, i.e., $OPT = Q_{1,1} + Q_{2,2}$. Let $s = (s_1, s_2)$ be a safety-level strategy profile for the game. Using these notations, the lemma claims that

$$\frac{OPT}{SW(s)} = \frac{Q_{1,1} + Q_{2,2}}{U_1(s) + U_2(s)} \leq 1 + r.$$

We next consider the possible supports of s_1 and s_2:

1. Both players are writing deterministically on the same topic; without loss of generality, assume that

$$\text{SUP}\,(s_1) = \text{SUP}\,(s_2) = \{1\}.$$

In this case, both players are writing on the first topic in order to maximize their worst-case utility. The ranking function R_{uni} only top-ranks a document if its quality is at least r that of the highest-quality document written on the topic. This means player 2 has to have comparable writing quality on this topic, i.e. $Q_{2,1} \geq r \cdot Q_{1,1}$. The payoffs for the players are then $U_1(s) = Q_{1,1}/2$, and $U_2(s) = Q_{2,1}/2$. Additionally, since second player chose not to ever write on topic 2 we have $Q_{2,2} \leq Q_{2,1}/2$, it has to be at least the payoff of the second topic, i.e. $Q_{2,1} \geq 2Q_{2,2}$. Finally, we bound the ratio as follows:

$$\begin{aligned}\frac{OPT}{SW(s)} &= \frac{Q_{1,1} + Q_{2,2}}{U_1(s) + U_2(s)} = \frac{Q_{1,1} + Q_{2,2}}{Q_{1,1}/2 + Q_{2,1}/2} \\ &= \frac{2\,(Q_{1,1} + Q_{2,2})}{Q_{1,1} + \frac{3}{8} \cdot Q_{2,1} + \frac{5}{8} \cdot Q_{2,1}} \\ &\leq \frac{2\,(Q_{1,1} + Q_{2,2})}{Q_{1,1} + \frac{3}{8} \cdot r \cdot Q_{1,1} + \frac{5}{8} \cdot 2Q_{2,2}} \\ &\leq_{(r \leq 2/3)} \frac{2\,(Q_{1,1} + Q_{2,2})}{Q_{1,1} + \frac{3}{8} \cdot \frac{2Q_{1,1}}{3} + \frac{5}{8} \cdot 2Q_{2,2}} = 1.6.\end{aligned}$$

This shows that if both players play a safety-level strategy of writing on the same topic, the anarchy is at most 1.6. The case where both players write about the second topic, $\text{SUP}\,(s_1) = \text{SUP}\,(s_2) = \{2\}$, is symmetric and the result holds for it as well.

2. $\text{SUP}\,(s_1) = \{1\}, \text{SUP}\,(s_2) = \{2\}$. In this case, the optimal social welfare is obtained.

3. $\text{SUP}\,(s_1) = \{2\}, \text{SUP}\,(s_2) = \{1\}$. Since player 1 writes on topic 2 as a safety level strategy, and player 2 on the first topic, we get that $Q_{1,2} \geq 2Q_{1,1}$ and $Q_{2,1} \geq 2Q2,2$. But then $SW(s) \geq OPT$, which is a contradiction.

4. One of the players randomizes over both topics while the other writes on a single topic, which is *not* the one he writes on in the optimal strategy profile; without loss of generality, assume that

$$\text{SUP}\,(s_1) = \{1, 2\}, \text{SUP}\,(s_2) = \{1\}.$$

As both players write about the first topic, their quality ratio has to be at most r, i.e.

$$Q_{2,1} \geq Q_{1,1}/r > \frac{2Q_{1,1}}{3}. \tag{2}$$

Also, as the player 2 always writes about the first topic, we have

$$Q_{2,1} \geq 2Q_{2,2}. \tag{3}$$

Player 1 randomizes over both topics, which implies that

$$Q_{1,2} \geq Q_{1,1}/2. \tag{4}$$

In this case, we can explicitly compute s_1 strategy. Being a safety-level strategy, we get that:

$$s_1 \in \operatorname{argmax}_{s_1 \in \Delta(2)} \min_{s_2 \in \Delta(2)} U_1(s_1, s_2).$$

Using the notations $p \triangleq s_1(1), q \triangleq s_2(1)$, define the function:

$$\begin{aligned} f(q) \triangleq U_1(s) = p \cdot q \cdot Q_{1,1}/2 + p \cdot (1-q) \cdot Q_{1,1} \\ + (1-p) \cdot q \cdot Q_{1,2} + (1-p) \cdot (1-q) \cdot Q_{1,2}/2. \end{aligned}$$

Minimizing over q, we get that

$$q = \begin{cases} 0 & \text{if } p < \frac{Q_{1,2}}{Q_{1,1}+Q_{1,2}} \\ 1 & \text{otherwise} \end{cases}.$$

Therefore, the utility of the first player for a given value of p is

$$g(p) \triangleq \min_q U_1(s) = \begin{cases} p \cdot Q_{1,1} + (1-p) \cdot Q_{1,2}/2 & \text{if } p < \frac{Q_{1,2}}{Q_{1,1}+Q_{1,2}} \\ p \cdot Q_{1,1}/2 + (1-p) \cdot Q_{1,2} & \text{otherwise} \end{cases}.$$

By choosing $p = s_1(1) = \frac{Q_{1,2}}{Q_{1,1}+Q_{1,2}}$, player 1 maximizes $g(p)$ and therefore this is his safety level strategy. The utility of player 2 is then:

$$\begin{aligned} U_2(s) &= Q_{2,1}/2 \cdot s_1(1) + Q_{2,1} \cdot s_1(2) \\ &= Q_{2,1} \cdot (1 - s_1(1)/2) \\ &= Q_{2,1} \cdot \frac{2Q_{1,1} + Q_{1,2}}{2Q_{1,1} + 2Q_{1,2}}. \end{aligned}$$

As player 1 is the only one writing on topic 2, his payoff is $U_1(s) = Q_{1,2}$. Finally, the social welfare ratio is

$$\begin{aligned} \frac{OPT}{SW(s)} &= \frac{Q_{1,1} + Q_{2,2}}{Q_{1,2} + Q_{2,1} \cdot \frac{2Q_{1,1}+Q_{1,2}}{2Q_{1,1}+2Q_{1,2}}} \\ &\overset{(4)}{\leq} \frac{Q_{1,1} + Q_{2,2}}{Q_{1,1}/2 + Q_{2,1} \cdot \left(1 - \frac{Q_{1,2}}{2Q_{1,1}+2Q_{1,2}}\right)} \\ &\leq \frac{Q_{1,1} + Q_{2,2}}{\frac{Q_{1,1}}{2} + Q_{2,1} \cdot \frac{1}{2}} \\ &\overset{(2),(3)}{\leq} \frac{2(Q_{1,1} + Q_{2,2})}{Q_{1,1} + \frac{3}{8} \cdot \frac{2Q_{1,1}}{3} + \frac{5}{8} \cdot 2Q_{2,2}} = 1.6. \end{aligned}$$

5. $\mathrm{SUP}(s_1) = \{1, 2\}, \mathrm{SUP}(s_2) = \{2\}$. As both players write about the second topic, their quality ratio has to be at most r, i.e.

$$Q_{1,2} \geq Q_{2,2}/r > \frac{2Q_{2,2}}{3}. \tag{5}$$

As player 1 is the only one to write on the first topic, his payoff is $U_1(s) = Q_{1,1}$. Similarly to the previous case, player 1 plays a safety level strategy of $s_1(2) = \frac{Q_{1,1}}{Q_{1,1}+Q_{1,2}}$. As player 2 writes on the second topic, his payoff is then

$$U_2(s) = Q_{2,2} \cdot (1 - s_1(2)/2) = Q_{2,2} \cdot \frac{Q_{1,1} + 2Q_{1,2}}{2Q_{1,1} + 2Q_{1,2}}.$$

Next, we bound the social welfare ratio:

$$\begin{aligned}
\frac{OPT}{SW(s)} &= \frac{Q_{1,1} + Q_{2,2}}{Q_{1,1} + Q_{2,2} \cdot \frac{Q_{1,1}+2Q_{1,2}}{2Q_{1,1}+2Q_{1,2}}} \\
&= \frac{Q_{1,1} + Q_{2,2}}{Q_{1,1} + Q_{2,2} - \frac{Q_{1,1}Q_{2,2}}{2Q_{1,1}+2Q_{1,2}}} \\
\overset{(5)}{\leq} &\frac{Q_{1,1} + Q_{2,2}}{Q_{1,1} + Q_{2,2} - \frac{Q_{1,1}Q_{2,2}}{2Q_{1,1}+2\cdot\frac{2Q_{2,2}}{3}}} \\
&\leq \frac{Q_{1,1} + Q_{2,2}}{Q_{1,1} + Q_{2,2} - \frac{Q_{1,1}Q_{2,2}}{Q_{1,1}+Q_{2,2}}} \\
&= \frac{(Q_{1,1} + Q_{2,2})^2}{Q_{1,1}^2 + Q_{1,1}Q_{2,2} + Q_{2,2}^2} \\
&= \frac{(Q_{1,1} + Q_{2,2})^2}{\frac{3}{4}(Q_{1,1} + Q_{2,2})^2 + \frac{1}{4}(Q_{1,1} - Q_{2,2})^2} \leq \frac{4}{3}.
\end{aligned}$$

6. $\mathrm{SUP}(s_1) = \{1, 2\}, \mathrm{SUP}(s_2) = \{1, 2\}$.
   Tennenholtz (2002) showed that in a two-player game in which every player has two strategies, a safety-level strategy which randomizes over both topics has the same expected payoff for each of the players as in a Nash equilibrium of that game. In our case, as both player play a mixed safety-level strategy, its social welfare is the same as the social welfare obtained by an equilibrium. Finally, Theorem 13 guarantees that any equilibrium achieves a price of anarchy of at most 1.683, and therefore this bound also applies here.

We showed that for any support of s the lemma holds, thereby establishing its correctness. □

Appendix N. Proof of Theorem 14

We now prove the following.

Theorem *There exists a Fair Ranking Function R_{uni}, such that for any Quality Matrix $\mathcal{Q} \in [0,1]^{[2]\times[m]}$, the Different Qualities Publishers Game $DQPG = (2, m, \mathcal{Q}; R_{uni})$ achieves a price of anarchy of at most* 1.683.

Proof. For a general m-topics game, consider two topics i, j such that the maximum social welfare strategy is reached when player 1 writes on topic i while player 2 creates a document on topic j; i.e.,

$$i, j \in \mathrm{argmax}_{i \neq j} \{\mathcal{Q}_{1,i} + \mathcal{Q}_{2,j}\}.$$

This means that the optimal social welfare reached for any strategy profile is

$$OPT = \mathcal{Q}_{1,i} + \mathcal{Q}_{2,j}.$$

Let $\mathcal{Q}' \in [0,1]^{[2]\times[2]}$ be the following quality matrix:

players/topics	1	2
1	$\mathcal{Q}_{1,i}$	$\mathcal{Q}_{1,j}$
2	$\mathcal{Q}_{2,i}$	$\mathcal{Q}_{2,j}$

The Quality Matrix $\mathcal{Q}'$.

Denote by DQPG'=$(2, 2, \mathcal{Q}'; R_{uni})$ the 2×2 game created when restricting players' strategies to topics i, j. Lemma 2 shows that the ratio between the social welfare reached by a safety-level strategy in DQPG' and OPT is at most 1.683. DQPG is different from DQPG' by allowing players to write on additional topics. Since expansion of the action set cannot reduce the players' utility, the safety level strategy for DQPG also achieves 1/1.683 of the optimal social welfare. Finally, we recall that the equilibrium strategy profile provides utility which is at least as high as the Safety Level Strategy of the players, therefore the price of anarchy of DQPG is at most 1.683. □

Appendix O. Proof of Theorem 15

Here, we provide a full proof for the theorem below.

Theorem *For any $n, m \geq 2$, there exists a quality matrix $\mathcal{Q} \in [0,1]^{[n]\times[m]}$ such that for every Fair Ranking Function $\mathcal{R}$, the price of anarchy of the Different Qualities Publishers Game DQPG=$(n, m, \mathcal{Q}; \mathcal{R})$ is at least $4/3$.*

Proof. Consider the quality matrix:

players/topics	1	2	...	m
1	1	1		0
2	1	1		0
$\vdots$			$\ddots$	
n	0	0		0

All zeros Quality Matrix except for $(1,1), (1,2), (2,1)$, and $(2,2)$.

We denote by p the probability that player 1 gets ranked first when all players write on the first topic. Similarly, we use q for the probability of seeing player 1 ranked first, when everyone except the second player write on topic 1. Being a Fair Ranking Function, we are guaranteed that:

1. The probability that the second player ranks first on topic 1 if all players write about it is p.

2. The probability that player 2 ranks first on topic 1 if every player except the first write about it is q.

3. If both players 1 and 2 write on the second topic and all other players write on the first, they will each rank first with probability $1/2$.

4. If exactly one of $\{1, 2\}$ writes on topic 2 while the remaining players write on the first topic, she will rank first (on the second topic) with probability 1.

This means that if all players except the first two writes about the first topic, the induced game for players $1, 2$ has the following payoff matrix:

(p, p)	$(q, 1)$	0	...	0
$(1, q)$	$(1/2, 1/2)$	0	...	0

The payoffs of players $1, 2$ when all of the other players write on topic 1.

This game has a mixed strategies equilibrium in which:

- players $1, 2$ each write on the first topic with probability $x \triangleq \frac{2q-1}{-2p+2q+1}$ and on the second topic otherwise.
- players $3, \ldots, n$ write on topic 1.

The player payoffs are then $U_1 = U_2 = x + (1-x)/2 = \frac{p-2q}{2p-2q-1}$ and $U_3, \ldots, U_n = 0$. This means that the social welfare of the game is

$$SW(p,q) \triangleq \sum_{i=1}^{n} U_i = \frac{2p-4q}{2p-2q-1}.$$

Maximizing over the domain[9] $\{(p,q) \mid 0 \leq p \leq 1/2, 0 \leq q \leq 1\}$, we get that in all cases $SW(p,q) \leq 1.5$. Next, note that a social welfare of 2 is easily achievable if player 1 writes on the first topic while the second player write on topic 2, and the ranking function never ranks a 0-quality document first. This means that for every Fair Ranking Function the price of anarchy is at least $\frac{2}{1.5} = 4/3$. □

Appendix P. Proof of Theorem 16

We now prove the following theorem:

Theorem *For any Fair Ranking Function $\mathcal{R}$ and $m \geq 2$, there exists a Distinguishable Quality Matrix $\mathcal{Q} \in [0,1]^{[2]\times[m]}$, such that the price of anarchy of the Different Qualities Publishers Game $DQPG = (2, m, \mathcal{Q}; \mathcal{R})$ is at least* 1.5.

Proof. Intuitively, we show that $\mathcal{R}$ has to be at least 1.5 worst than the optimum in one of two cases. If it prefers a "$\frac{2}{3}$-quality" document over a perfect document with probability larger than third, then both players might write on the topic they are less qualified to write on, thereby incurring a 1.5 anarchy. Alternatively, if $\mathcal{R}$ do not rank the worse document with that probability, players might selfishly write on their topic of choice, and no one would satisfy the information need of the remaining topics.

We denote $p \triangleq \mathcal{R}(2/3, 1)(1)$, i.e. the probability $\mathcal{R}$ ranks a document with quality $2/3$ higher than a document of quality 1 . Next, we consider two cases:

- $\mathbf{p \geq 1/3}$. In this case, consider the Distinguishable Quality Matrix:[10]

players/topics	1	2
1	1	2/3
2	2/3	1

.

This means that the payoff for the players can be represented as:

$(1-p, 2p/3)$	$(1,1)$
$(2/3, 2/3)$	$(2p/3, 1-p)$

.

Since $1/3 \leq p \leq 1$, there exists an equilibrium in which player 1 writes about topic 2 while player 2 writes about topic 1. In this equilibrium, the payoff for each player is $2/3$, while the optimum is 2. This means that the price of anarchy of $\mathcal{R}$ is at least $\frac{OPT}{P} = \frac{2}{4/3} = 1.5$.

- $\mathbf{p < 1/3}$. When $\mathcal{R}$ has a high probability of ranking the second best document first, it will pay higher in the following situation. Consider the Distinguishable Quality Matrix:

9. p is bounded by $1/2$ since a Fair Ranking Function have to give equal probability to players $1, 2$.
10. If the number of topics is larger than 2, we can simply use a quality of close to zero for all topics except the first two and obtain the same result.

players/topics	1	2
1	1	$1-p$
2	2/3	0

.

The following payoff matrix is then:

$(1-p, 2p/3)$	$(1, 0)$
$(1-p, 2/3)$	$(1-p, 0)$[11]

.

Notice that both players writing on the first topic is an equilibrium for the game, while the optimum is reached at $2, 1$, thus $\frac{OPT}{P} = \frac{1-p+2/3}{1-p+\frac{2p}{3}} = \frac{5-3p}{3-p} \geq_{(p<1/3)} 1.5$.

□

References

(Ed.)., AIRWeb (2005–2009). *AIRWeb — International Workshop on Adversarial Information Retrieval on the Web.*

Amati, G., & van Rijsbergen, C. J. (2002). Probabilistic models of information retrieval based on measuring the divergence from randomness. *ACM Transactions on Information Systems*, *20*(4), 357–389.

Ashlagi, I., Monderer, D., & Tennenholtz, M. (2008). On the value of correlation. *J. Artif. Intell. Res. (JAIR)*, *33*, 575–613.

Ashlagi, I., Monderer, D., & 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., & Kravi, E. (2016). The ranking game. In *Proceedings of the 19th International Workshop on Web and Databases (WebDB)*, pp. 7:1–7:6.

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

Benczúr, A. A., Csalogány, K., Sarlós, T., & Uher, M. (2005). Spamrank – fully automatic link spam detection. In *Proceedings of AIRWeb*, pp. 25–38.

Bendersky, M., Croft, W. B., & Diao, Y. (2011). Quality-biased ranking of web documents. In *Proceedings of WSDM*, pp. 95–104.

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

Carbonell, J. G., & Goldstein, J. (1998). The use of MMR, diversity-based reranking for reordering documents and producing summaries. In *Proceedings of SIGIR*, pp. 335–336.

Cormack, G. V. (2007). TREC 2007 spam track overview. In *Processing of TREC*.

Cormack, G. V., Smucker, M. D., & Clarke, C. L. A. (2010). Efficient and effective spam filtering and re-ranking for large web datasets. *CoRR*, *abs/1004.5168*.

Dalvi, N. N., Domingos, P., Mausam, Sanghai, S. K., & Verma, D. (2004). Adversarial classification. In *Proceedings of KDD*, pp. 99–108.

11. This implicitly assumes that $\mathcal{R}$ always ranks a positive-quality document over one with quality 0. The theorem holds in any case, as the payoff of this strategy profile is irrelevant.

El-Yaniv, R., & Nisenson, M. (2010). On the foundations of adversarial single-class classification. *CoRR*, *abs/1010.4466*.

Eliaz, K., & Spiegler, R. (2011). A simple model of search engine pricing. *The Economic Journal*, *121*(556), F329–F339.

Eliaz, K., & Spiegler, R. (2016). Search design and broad matching. *American Economic Review*, *106*(3), 563–586.

Elsas, J. L., & Dumais, S. T. (2010). Leveraging temporal dynamics of document content in relevance ranking. In *Proceedings of WSDM*, pp. 1–10.

Fetterly, D., Manasse, M., & Najork, M. (2004). Spam, damn spam, and statistics: Using statistical analysis to locate spam web pages. In *Proceedings of WebDB*, pp. 1–6.

Fuhr, N. (2008). A probability ranking principle for interactive information retrieval. *Information Retrieval*, *11*(3), 251–265.

Geva, S., Kamps, J., & Schenkel, R. (Eds.). (2012). *Focused Retrieval of Content and Structure, 10th International Workshop of the Initiative for the Evaluation of XML Retrieval, INEX 2011, Saarbrücken, Germany, December 12-14, 2011, Revised Selected Papers*, Vol. 7424 of *Lecture Notes in Computer Science*. Springer.

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

Hopcroft, J., & Sheldon, D. (2008). Network reputation games. Tech. rep., Cornell University.

Ieong, S., McGrew, R., Nudelman, E., Shoham, Y., & Sun, Q. (2005). Fast and compact: A simple class of congestion games. In *Proceedings of AAAI*, Vol. 5, pp. 489–494.

Izsak, P., Raiber, F., Kurland, O., & Tennenholtz, M. (2014). The search duel: a response to a strong ranker. In *Proceedings of SIGIR*, pp. 919–922.

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

Jones, T., Hawking, D., & Sankaranarayana, R. (2007). A framework for measuring the impact of web spam. In *Proceedings of ADCS*.

Koutsoupias, E., & Papadimitriou, C. (1999). Worst-Case Equilibria. In *STACS*.

Krishnan, V., & Raj, R. (2006). Web spam detection with Anti-Trust rank. In *Proceedings of AIRWeb*, pp. 37–40.

Liu, C., & 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).

Liu, T.-Y. (2009). Learning to rank for information retrieval. *Foundations and Trends in Information Retrieval*, *3*(3).

Mas-Colell, A., Whinston, M., & Green, J. (1995). *Microeconomic Theory*. Oxford University Press.

Monderer, D., & Shapley, L. S. (1996). Potential games. *Games and economic behavior*, *14*(1), 124–143.

Monderer, D., & Tennenholtz, M. (2004). K-implementation. *J. Artif. Intell. Res. (JAIR)*, *21*, 37–62.

Monderer, D., & Tennenholtz, M. (2006). Strong mediated equilibrium. In *Proceedings of AAAI*, pp. 691–696.

Nash, J. F. (1950). Equilibrium points in n-person games.. *Proc. of the National Academy of Sciences of the United States of America*, *36.1*, 48–49.

Nisan, N., & Ronen, A. (2001). Algorithmic mechanism design. *Games and Economic Behavior*, *35*, 166–196.

Ponte, J. M., & Croft, W. B. (1998). A language modeling approach to information retrieval. In *Proceedings of SIGIR*, pp. 275–281.

Radinsky, K., & Bennett, P. N. (2013). Predicting content change on the web. In *Proceedings of WSDM*, pp. 415–424.

Raiber, F., Collins-Thompson, K., & Kurland, O. (2013). Shame to be sham: Addressing content-based grey hat search engine optimization. In *Proceedings of SIGIR*, pp. 1013–1016.

Richardson, S., & Cox, I. J. (2014). Estimating global statistics for unstructured P2P search in the presence of adversarial peers. In *Proceedings of SIGIR*, pp. 203–212.

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.

Roughgarden, T., & Tardos, E. (2002). How bad is selfish routing?. *Journal of the ACM*, *49*(2), 236–259.

Rozenfeld, O., & Tennenholtz, M. (2007). Routing mediators. In *Proceedings of IJCAI*, pp. 1488–1493.

Salton, J., Wong, A., & Yang, C. S. (1975). A vector space model for automatic indexing. *Communications of the ACM*, *18*(11), 613–620.

Santos, A. S. R., Pasini, B., & Freire, J. (2016). A first study on temporal dynamics of topics on the web. In *Proceedings of WWW*, pp. 849–854.

Shoham, Y., & Tennenholtz, M. (1995). Social Laws for Artificial Agent Societies: Off-line Design. *Artificial Intelligence*, *73*.

Sloan, M., & Wang, J. (2012). Dynamical information retrieval modelling: a portfolio-armed bandit machine approach. In *Proceedings WWW*, pp. 603–604.

Sparck Jones, K., Walker, S., & Robertson, S. E. (2000). A probabilistic model of information retrieval: development and comparative experiments - part 1. *Information Processing and Management*, *36*(6), 779–808.

Tennenholtz, M. (2002). Competitive safety analysis: Robust decision-making in multi-agent systems. *J. Artif. Intell. Res. (JAIR)*, *17*, 363–37.

Wang, H., Xing, W., Asif, K., & Ziebart, B. D. (2015). Adversarial prediction games for multivariate losses. In *Proceedings of NIPS*, pp. 2728–2736.

WICOW (2012). WICOW/AIRWeb workshop on web quality (webquality)..

Wu, B., & Davison, B. D. (2005). Identifying link farm spam pages. In *Proceedings of WWW*, pp. 820–829.

Yang, G. H., Sloan, M., & Wang, J. (2016). *Dynamic Information Retrieval Modeling*. Synthesis Lectures on Information Concepts, Retrieval, and Services. Morgan & Claypool Publishers.

Zhang, Y., & Zhai, C. (2015). Information retrieval as card playing: A formal model for optimizing interactive retrieval interface. In *Proceedings of SIGIR*, pp. 685–694.

Zuccon, G., & Azzopardi, L. (2010). Developing the quantum probability ranking principle. In *IIR 2010 - Proceedings of the First Italian Information Retrieval Workshop, Padua, Italy, January 27-28, 2010*, pp. 21–22.