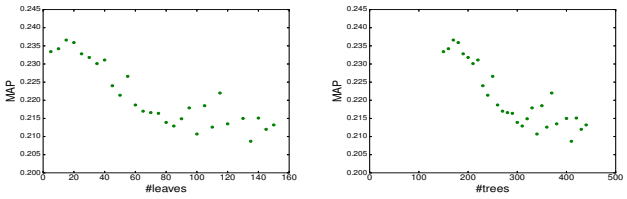

Figure 3: The connection between #leaves and #trees used in LambdaMART and its MAP effectiveness over ClueWeb.

The changes of a ranked document list that result from synthetic random noise introduced to documents were used to predict query performance [28]. We address a different setting — adversarial document changes intended to promote documents in rankings — and present a different analysis.

Much of the work on adversarial information retrieval focuses on identifying and addressing different types of spam (e.g., content-based and hyperlink-based) [1, 2]. In contrast to our work, there was no formal and empirical analysis of ranking robustness with respect to (adversarial) documents' manipulations.

Recent work analyzes the strategies employed by documents' authors in ranking competitions [17]. In contrast, we analyze the robustness of ranking functions. We use the datasets of ranking competitions organized in this work [17] in our empirical analysis.

The probability ranking principle [18] was shown to be sub-optimal in competitive retrieval settings, where authors manipulate their documents so as to have them highly ranked. However, the robustness of ranking functions was not studied as in our work.

There is a growing body of work on adversarial and robust classification; e.g., [4, 6–10, 14, 20, 21, 25]. The focus is on improving classifier's robustness to adversarial (often, minuscule) manipulations of objects and their feature values. In contrast to our work, the robustness of document ranking functions was not studied; specifically, the pairwise robustness notions we analyze, which are a core aspect of ranking robustness, were not studied.

The connection between neural network regularization and the stability of classification decisions has recently been demonstrated [20]. We demonstrate the connection between regularization of linear ranking functions and stability of retrieval scores, and more importantly, ranking robustness.

5 CONCLUSIONS AND FUTURE WORK

We presented a formal and empirical analysis of the robustness of rankings induced by feature-based relevance-ranking functions to (adversarial) manipulations of documents. We formally showed that increased regularization of linear ranking functions results in increased ranking robustness. Accordingly, we conjectured that increased variance of any learned ranking function results in decreased ranking robustness. We provided empirical support to our formal findings and the conjecture by analyzing ranking competitions where authors introduced adversarial changes to documents.

We plan to further study and improve the robustness of non-linear learning-to-rank functions. We also intend to extend the robustness analysis to sets of queries; e.g., those representing the same information needs.

Acknowledgments We thank the reviewers for their comments. This work was supported by funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement 740435).

REFERENCES

[1] 2005–2009. *AIRWeb — International Workshop on Adversarial Information Retrieval on the Web.*
[2] Carlos Castillo and Brian D. Davison. 2010. Adversarial Web Search. *Foundations and Trends in Information Retrieval* 4, 5 (2010), 377–486.
[3] Gordon V. Cormack, Mark D. Smucker, and Charles L. A. Clarke. 2011. Efficient and effective spam filtering and re-ranking for large web datasets. *Informaltiom Retrieval Journal* 14, 5 (2011), 441–465.
[4] Nilesh Dalvi, Pedro Domingos, Mausam, Sumit Sanghai, and Deepak Verma. 2004. Adversarial Classification. In *Proc. of KDD*. 99–108.
[5] Fernando Diaz. 2007. Regularizing query-based retrieval scores. *Information Retrieval* 10, 6 (2007), 531–562.
[6] Alhussein Fawzi, Omar Fawzi, and Pascal Frossard. 2015. Analysis of classifiers' robustness to adversarial perturbations. *CoRR* abs/1502.02590 (2015).
[7] Alhussein Fawzi, Seyed-Mohsen Moosavi-Dezfooli, and Pascal Frossard. 2016. Robustness of classifiers: from adversarial to random noise. In *Proc. of NIPS*. 1624–1632.
[8] Alhussein Fawzi, Seyed-Mohsen Moosavi-Dezfooli, and Pascal Frossard. 2017. The Robustness of Deep Networks: A Geometrical Perspective. *IEEE Signal Processing Magazine* 34, 6 (2017), 50–62.
[9] Amir Globerson and Sam T. Roweis. 2006. Nightmare at test time: robust learning by feature deletion. In *Proc. of ICML*. 353–360.
[10] Ian J. Goodfellow, Jonathon Shlens, and Christian Szegedy. 2015. Explaining and Harnessing Adversarial Examples. In *Proc. of ICLR*.
[11] Zoltán Gyöngyi and Hector Garcia-Molina. 2005. Web Spam Taxonomy. In *Proc. of AIRWeb 2005*. 39–47.
[12] Thorsten Joachims. 2006. Training linear SVMs in linear time. In *Proc. of KDD*. 217–226.
[13] Tie-Yan Liu. 2011. *Learning to Rank for Information Retrieval*. Springer. I–XVII, 1–285 pages.
[14] Daniel Lowd and Christopher Meek. 2005. Adversarial Learning. In *Proc. of SIGKDD*. 641–647.
[15] Donald Metzler and W. Bruce Croft. 2007. Linear feature-based models for information retrieval. *Information Retrieval* 10, 3 (2007), 257–274.
[16] Constantinos Panagiotakopoulos and Petroula Tsampouka. 2013. The Stochastic Gradient Descent for the Primal L1-SVM Optimization Revisited. In *Proc. of ECML PKDD*. 65–80.
[17] Nimrod Raifer, Fiana Raiber, Moshe Tennenholtz, and Oren Kurland. 2017. Information Retrieval Meets Game Theory: The Ranking Competition Between Documents' Authors. In *Proc. of SIGIR*. 465–474.
[18] Stephen E. Robertson. 1977. The Probability Ranking Principle in IR. *Journal of Documentation* (1977), 294–304. Reprinted in K. Sparck Jones and P. Willett (eds), *Readings in Information Retrieval*, pp. 281–286, 1997.
[19] Grace S.Shieh. 1998. A weighted Kendall's tau statistic. *Statistics & Probability Letters* 39, 1 (1998).
[20] Christian Szegedy, Wojciech Zaremba, Ilya Sutskever, Joan Bruna, Dumitru Erhan, Ian J. Goodfellow, and Rob Fergus. 2014. Intriguing properties of neural networks. In *Proc. of ICLR*.
[21] Thomas Tanay and Lewis D. Griffin. 2016. A Boundary Tilting Persepective on the Phenomenon of Adversarial Examples. *CoRR* abs/1608.07690 (2016).
[22] C. J. van Rijsbergen. 1979. *Information Retrieval* (second ed.). Butterworths.
[23] Ellen M. Voorhees. 2004. Overview of the TREC 2004 Robust Retrieval Track. In *Proc. of TREC*.
[24] Lidan Wang, Paul N. Bennett, and Kevyn Collins-Thompson. 2012. Robust ranking models via risk-sensitive optimization. In *Proc. of SIGIR*. 761–770.
[25] Yizhen Wang, Somesh Jha, and Kamalika Chaudhuri. 2017. Analyzing the Robustness of Nearest Neighbors to Adversarial Examples. *CoRR* abs/1706.03922 (2017).
[26] William Webber, Alistair Moffat, and Justin Zobel. 2010. A Similarity Measure for Indefinite Rankings. *ACM Transactions on Information Systems* 28, 4 (2010), 20:1–20:38.
[27] Qiang Wu, Christopher J. C. Burges, Krysta Marie Svore, and Jianfeng Gao. 2010. Adapting boosting for information retrieval measures. *Information Retrieval* 13, 3 (2010), 254–270.
[28] Yun Zhou and Bruce Croft. 2006. Ranking robustness: a novel framework to predict query performance. In *Proc. of CIKM*. 567–574.