

**ArCOV-19:  
The First Arabic COVID-19 Twitter Dataset  
with Propagation Networks**

By

**Dr. Tamer Elsayed  
Associate Professor  
Computer Science & Engineering Dept.  
Qatar University**



## Bio

Dr. Tamer Elsayed is an associate professor of Computer Science at the CSE department at Qatar University (QU). He received his B.Sc. and M.Sc. degrees from Alexandria University in Egypt, and Ph.D. degree from University of Maryland (UMD) in the United States in 2009. He also had an internship at Google in 2007. After getting his Ph.D., he worked as a researcher at UMD, KAUST, and Microsoft, before joining QU in 2012. His main research interests are information retrieval and big data analytics. He published more than 70 publications in top-tier journals and conferences. He received 2 best paper awards and his research team at QU was ranked first at TREC 2016 and second at TREC 2019 international research competitions. He co-organized several workshops and evaluation labs in international conferences, and he is the co-chair of ACM Africa Summer School on Machine Learning for Data Mining and Search (AFIRM). He led 3 NPRP projects funded by QNRF and leads an active research group (bigIR) at QU. Personal webpage:

<http://qufaculty.qu.edu.qa/telsayed/>

**Date:** Mon. 20 Apr. 2020

**Time:** 12:30 -13:30 (Doha time)

**Venue:** Online – [Register here](#)

## Abstract

While COVID-19 has severely affected the lives of billions of people around the globe (in one way or another), it opens up new frontiers of research directions that need prompt exploration. To that end, in this webinar, I present ArCOV-19, an Arabic COVID-19 Twitter dataset that covers the period from 27th of January till 31st of March 2020 (and still ongoing). ArCOV-19 is the first publicly-available Arabic Twitter dataset covering COVID-19 pandemic that includes around 748k popular tweets (according to Twitter search criterion) alongside the propagation networks of the most-popular subset of them. The propagation networks include both retweets and conversational threads (i.e., threads of replies). ArCOV-19 is designed to enable research under several domains including natural language processing, data science, and social computing, among others. Preliminary analysis shows that ArCOV-19 captures rising discussions associated with the first reported cases of the disease as they appeared in the Arab world. In addition to the source tweets and the propagation networks, we also release the search queries and the language-independent crawler used to collect the tweets to encourage the curation of similar datasets. The webinar sheds the light on why we collected ArCOV-19, how we designed it, what it really covers, how researchers in several different domains can make use of it, and what future directions we can take from there.