Learning on Social Networks

Social media and online social networking sites contain many opinionated, inaccurate or false facts that are often refuted over time. Spread of misinformation may have confused and misled voters in the last U.S. presidential election or the Brexit referendum. To overcome this problem, online platforms deploy evaluation mechanisms for their users to further curate information within these platforms. For example, users can remove inaccurate contents from *Wikipedia*, mark a correct answer as verified in *Stack Overflow* and flag a story as misinformation in *Facebook* and *Twitter*.

We developed [158] a unified computational framework that leverages the temporal traces left by the aforementioned examples of noisy evaluations to estimate robust, unbiased and interpretable notions of information reliability and source trustworthiness. The key idea is that unreliable contents are often removed quickly while reliable contents remain on platforms such as *Wikipedia* for a long time. Similarly, information contributed by sources which systematically spread misinformation are often removed quickly on a wide range of entries, while contents contributed by thrustworthy sources remain on the platform for a long time. By applying our framework to *Wikipedia* data, we are able to answer questions such as whether bbc.co.uk provides more reliable information compared to newyorktimes.com in *Wikipedia* entries related to the U.S politics, and at which point in time a particular *Wikipedia* entry, such as *Barack H. Obama*, was unreliable due to ongoing controversies.

Next, we focused on developing a machine learning method to detect and reduce the spread of harmful misinformation in online social networking sites through the power of the crowd and fact-checking [105]. Given limited reviewing resources, the main question is how to prioritize growing amounts of questionable contents. Some cases of misinformation are not identified until a large number of users have been already exposed to it. However, many cases may only have a limited impact on people and unnecessarily consume reviewing resources for fact-checking. To address these challenges, we developed a robust methodology with provable guarantees to minimize the impact of potentially harmful contents on a large number of people. Results of applying this algorithm on datasets from *Twitter* and *Weibo* suggest that our method can identify cases of misinformation earlier than alternative methods and also uses fact-checking resources more efficiently.

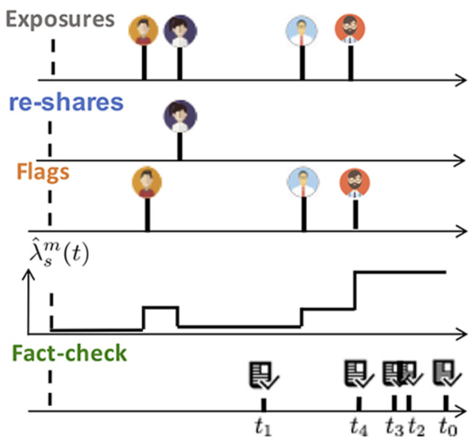

Figure 1.7: Fact checking of contents on social networks using observed exposure, reshare and flag events. The rate of fact checking, $\hat{\lambda}_s^m(t)$, is updated after every observed event.

More recently, we applied some of the above techniques to human learning. In this project, due to appear in PNAS [17], we develop a methodology to optimize spaced repetition algorithms, a widely used procedure to memorize new information, e.g., vocabulary when learning a foreign language in online learning platforms such as *Duolingo*. The promise of online platforms such as *Duolingo* is that automated fine-grained monitoring and a greater degree of control will result in more effective spaced repetition algorithms. However, the algorithms used in these services tend to be simple rule-based heuristics. We introduce a principled and data-driven method to characterize spaced repetition algorithms. Results from applying this framework to a dataset from *Duolingo* indicate that this method help learners memorize more effectively and efficiently.

More information: https://ei.is.mpg.de/project/learning-on-social-networks