

Facebook 'likes' the scientific method

Social-networking giant may allow researchers to check the data underlying its studies.

Zoë Corbyn

25 July 2012

Social scientists hungry for Facebook's data may be about to get a taste of it. *Nature* has learned that the social-networking website is considering giving researchers limited access to the petabytes of data that it has amassed on the preferences and behaviour of its almost one billion users.

Outsiders will not get a free run of the data, but the move could quell criticism from social scientists who have complained that the company's own research on its users cannot be verified. Facebook's in-house scientists have been involved in publishing more than 30 papers since 2009, covering topics from what drives the spread of information and ideas¹ to the relationship between social-networking activity and loneliness². However, because the company fears breaching its users' privacy, it does not release the underlying raw data.

Facebook is now exploring a plan that could allow external researchers to check its work in future by inspecting the data sets and methods used to produce a particular study. A paper currently submitted to a journal could prove to be a test case, after the journal said that allowing third-party academics the opportunity to verify the findings was a condition of publication.

"We want to participate in the scientific process and we believe that there should be a way to have other researchers validate [our studies] without infringing on the policies that we have set with our users," says Cameron Marlow, head of Facebook's data-science team.

Restricted access

If the scheme were to go ahead, it would apply to papers after publication. Scholars would have to travel to the company's headquarters in Menlo Park, California, because Facebook would not risk sending the data electronically, and they would have access to aggregated data only, and no personally identifiable information. The company would also allow access for only a limited period — and contingent upon researchers signing a non-disclosure agreement. Marlow says, however, that these conditions should not keep researchers from being openly critical about matters related to the published paper such as technique or data processing.

External scholars would not be allowed to conduct their own studies on the data sets.

The alternative — publicly releasing anonymized raw data sets — is not likely to be an option, says Facebook. Internet company AOL, based in New York, and film rental and streaming firm Netflix, based in Los Gatos, California, have both done this in the past, only for researchers to show that individuals could be identified in the anonymized data. "It is hard to really guarantee that it is anonymous," says Marlow.

Facebook's proposals are a step in the right direction, say researchers. "Their intentions are very good," agrees Bernardo Huberman, director of the social-computing group at Hewlett-Packard Laboratories in Palo Alto, California. Huberman has voiced concerns in *Nature* about the lack of researcher access to 'big data' at private companies³. Facebook "wants to get closer to something that is the scientific method", he says.

But Huberman and others have practical concerns. The requirement for on-site visits will hinder many researchers, with few likely to receive funding to travel to merely validate a completed study. Furthermore, it is unclear whether Facebook will allow researchers to validate research by running their own programs on the data. If scientists are restricted to repeating Facebook researchers' own analyses, says Anatoliy Gruzd, director of the social-media lab at Dalhousie University in Halifax, Canada, "they may be unknowingly repeating the same errors inherent in a technique".



References

1. Ugander, J., Backstrom, L., Marlow, C. & Kleinberg, J. *Proc. Natl Acad. Sci. USA* **109**, 5962–5966 (2012).
2. Burke, M., Marlow, C. & Lento, T. *Proc. 28th Int. Conf. Human Factors Computing Systems* 1909–1912 <http://dx.doi.org/10.1145/1753326.1753613> (ACM, 2012).
3. Huberman, B. A. *Nature* **482**, 308 <http://dx.doi.org/10.1038/482308d> (2012).