Reader Report

Prestige drives epistemic inequality in the diffusion of scientific ideas

Claudia Shi (js5334)

April 11, 2019

This is not really related to causal inference but related to the topic I am working on.

This paper challenges the traditional view that the intrinsic fitness of an idea causes it to widespread. Furthermore, citation counts correlate with importance and impact. This paper established a theoretical trade-offs between university prestige and the quality of ideas necessary for efficient circulation, and argued that faculty hiring is an underlying mechanism that drove such so call "epsitemic advantage".I am less interested in the conclusion this paper reached, (not just because it undermined a lot of the assumptions I am making) than the mechanism the paper used, since it shares a level of "vagueness" in defining outcome and covariates that other sciences do not usually encounter. It is also more question driven than data driven.

It defined "prestige" as a department's ability to place its graduates as faculty at other institutions. It justified their definition, instead of department ranking, on the ground that it more accurately predicts faculty placement. This seems an odd argument to make for justification. Of course "ability to produce faculty placement" can more accurately predict "faculty placement", since they essentially the same variable excepts one is at $t-\infty 0$ and the other one is at t_0, but is it actually a good measurement for prestige? One could easily argue otherwise.

It also vaguely defined the "the spread of idea" in this paper. Specifically, it pulled publication dataset from corresponding authors, and matched them to 5 different topics. Similar to the "tag words" in my dataset, this paper decided to do keyword search among all the $200,476$ publications in their dataset, and then manually verify them. One of their topic is "topic modeling", it is quite surprising that there are only 217 paper among that dataset, showing it is likely that it is drastically under classified.

Finally, it did some hypothesis testing that verified their claims. Overall, I think this paper is asking an interesting question. It also used a quite clever network design to map out the relationship, and produced very cool figured in the papaer (less relevant, granted). But I think it is trying to answer a causal question without thinking in that framework. Perhaps mechanism design and hypothesis testing can an alternative? I would imagine most people in this class would disagree.