

# What's in a name? Big Data reveals distinctive patterns in higher education systems

July 3 2017

---



Credit: CC0 Public Domain

Using lists of names collected from publicly available websites, two University of Chicago researchers have revealed distinctive patterns in higher education systems, ranging from ethnic representation and gender

imbalance in the sciences, to the presence of academic couples, and even the illegal hiring of relatives in Italian universities.

"This study was an exercise in exploiting bare-bones techniques," said author Stefano Allesina, PhD, professor of ecology & evolution and a member of the Computation Institute at the University of Chicago. "We wanted to analyze the simplest form of data you could imagine: lists of [names](#). That's all we had. We wondered what kinds of information we could extract from such a meager source of data. We also asked: how could we use this to explore real-world problems?"

For the study—"Last name analysis of mobility, [gender imbalance](#), and nepotism across academic systems," published July 3, 2017 in the *Proceedings of the National Academy of Sciences*—Allesina and postdoctoral scholar Jacopo Grilli, PhD, acquired lists of the surnames of all Italian academics in the four years 2000, 2005, 2010 and 2015. For comparison, they also gathered lists of all researchers currently working at the Centre National de la Recherche Scientifique (CNRS) in France, and those working at research-intensive public institutions in the United States.

Then they counted the number of professors in each department who shared last names and contrasted that to the number expected by chance. They found three possible explanations for an overabundance of identical last names. An unusually high proportion of name sharing could be due to geography; certain names are typical of a region. Or, immigration could have an impact, for example, the influx of Asian faculty to the United States in disciplines such as in mathematics and computer science.

If the clustering of names cannot be explained by these two factors—which was the case in certain disciplines and regions in Italy—then the data point to nepotistic hires: professors who recruit their

relatives for academic positions.

The Allesina laboratory is not new to this type of analysis. In a 2011 paper published in PLoS One, Allesina demonstrated that certain disciplines (law, medicine, engineering) in Italian universities displayed a severe scarcity of last names, raising the suspicion of nepotism.

That study caused "quite a stir in Italy," Allesina said. The publication followed a complete overhaul of the nation's academic system. The reform, passed in late 2010, included a provision intended to prevent professors from recruiting relatives by shifting hiring and funding decisions away from the universities to independent panels. The perception at the time was that "promotions and funding were often awarded on the basis of connections rather than merit, providing mediocre and unproductive professors with jobs for life while pushing many of the country's brightest minds abroad," Allesina said.

Grilli and Allesina decided to take a closer look at the law's impact since 2010 and to compare the prevalence of nepotism in Italy with other countries. They found that nepotism in Italy appears to have declined somewhat over the period from 2000 to 2015. In 2000, seven of the 14 fields measured showed clear signs of nepotism. That fell to five fields in 2010, and only two, chemistry and medicine, by 2015.

The 2010 law, they point out, was not the only factor in the decrease of apparent nepotism. Much of the decline, the researchers point out, could be traced to an increase in faculty retirements and a dearth of new hires.

The Italian university system has been "virtually butchered over the last decade," Allesina said, with a staggering 10 percent overall loss of faculty, and losses of 20 to 30 percent of the faculty at several leading universities. "This had a strong effect on new hires," he said, "but only a limited impact on favoritism over the whole university system."

The researchers' focus on last names illuminates some recent changes in U.S. academics as well. When faculty last names were randomized by field, the huge impact of immigration on U.S. universities became obvious. More than half of the 5.2 million immigrant scientists, mathematicians and engineers currently working in the United States were born in Asia.

"Certain names are associated with specific academic fields and certain heritages tend to target preponderantly science and engineering," said Grilli. Zhang, for example, is now the most common last name in the U.S. in the fields of chemistry and mathematics. It ranks third in agriculture, geology and physics, but falls to 115th in humanities. Smith, on the other hand, is among the top three in humanities, sociology and medicine, but 20th in chemistry and 47th in geology.

"Sometimes using very simple data can get you expected and unexpected results," Allesina said. First names can reveal a field's gender imbalance. They can also fluctuate wildly. The most common first name in the past decade for boys in Italy was Francesco, but that increased by 40 percent following the election of Pope Francis. "It was declining," Grilli said, "but it bounced back."

"The good and bad of Italy is the family," Allesina said. "It protects you from collapse, but it also prevents growth. This really becomes a weight on the shoulders of young people, especially in the South, where many talented students have no choice but to emigrate."

**More information:** Jacopo Grilli et al., "Last name analysis of mobility, gender imbalance, and nepotism across academic systems," *PNAS* (2017). [www.pnas.org/cgi/doi/10.1073/pnas.1703513114](http://www.pnas.org/cgi/doi/10.1073/pnas.1703513114)

Provided by University of Chicago Medical Center

Citation: What's in a name? Big Data reveals distinctive patterns in higher education systems (2017, July 3) retrieved 19 April 2024 from <https://phys.org/news/2017-07-big-reveals-distinctive-patterns-higher.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.