The international journal of science / 8 July 2021

nature

Big-data studies of human behaviour need a common language

Computational social science will benefit from stronger bridges between its disciplines.

What are the causes of vaccine hesitancy? How can people be encouraged to exercise more? Social scientists researching these questions observe how people behave and record data on those behaviours, then augment this knowledge by interviewing and/or polling those whom they are studying. Carrying out research in this way is a time-consuming and manual process. Moreover, it is difficult to obtain large amounts of data simultaneously.

But now, researchers have access to an unprecedented amount of social data, generated every second by continuous interactions on digital devices or platforms. These include data that trace people's movements, purchases and online social interactions — which are all proving extraordinarily powerful for research. As a result, work that weaves together large data analysis and social questions, known as computational social science, has witnessed huge growth.

During the course of the coronavirus pandemic alone, researchers have been able to access millions of mobile-phone records to study how people's movements changed during the pandemic, and the impact of those changes on how SARS-CoV-2 spread. They have been able to access anonymized credit-card purchase histories to study how people are spending money during the pandemic — information that can be used to understand how COVID-19 is affecting various sectors of the economy.

Power and responsibility

At the same time, researchers need to remember that gathering and sharing such personal data — practices that are currently largely unregulated — pose many challenges to society. These include risks from increased surveillance, and the danger that people could be reidentified from otherwise anonymized data.

There are also concerns that people whose data are being used have not fully consented to this — and wider worries about the economic monopoly of tech corporations that own the majority of the data. These digital traces tend to be left disproportionately by relatively wealthy people in high-income countries, biasing attempts to draw global conclusions. Acknowledging and working on these issues is key to the development of ethical computational social science that promotes real societal progress.

> **"Trans-disciplinary co-working is essential for better decisions and robust outcomes."**

The need to blend expertise in the social sciences with the skills required to collect, clean and analyse large data sets means that computational social science requires teams of researchers with a remarkably diverse set of expertise and skills. But with collaborations across disciplines come other challenges.

This week, *Nature* is publishing a special collection of articles with the objective of bridging the research disciplines and perspectives on doing science that underpin computational social science. We're highlighting ways in which communities of social, natural and computational scientists can learn to better work together, to complement each other and to overcome shared challenges.

Stronger bridges

To begin with, the varied disciplines need to overcome language barriers in which the same terms have different meanings. For example, in many of the social sciences, 'prediction' often refers to a correlation; in the physical sciences, it usually means a forecast. True transdisciplinary research requires scientists to first learn each other's languages, and then develop a shared understanding of terms.

But the divide can run deeper than language, into how to curate, analyse and interpret data to explain a phenomenon. Jake Hofman at Microsoft Research in New York City and colleagues argue on page 181 that computational social science could most effectively answer research questions by combining complementary approaches. For example, researchers building a numerical forecast on, say, the causes of traffic jams would assemble data on traffic flows, with insights from drivers on their reasons for taking particular routes.

The results of any study are determined by not only the analytical strategies used, but also the quality of the data — and this becomes particularly delicate when dealing with social data. The vast amounts of available data that make computational social science possible — such as tweets or location data from phones — are usually not gathered for research purposes and so can easily be misinterpreted.

That is why, as David Lazer at Northeastern University in Boston, Massachusetts, and colleagues write on page 189, researchers who work with large data sets must resist drawing conclusions from just the trends or patterns seen in the numbers — and should account for factors that could affect a result. To extract real meaning from data, researchers need to ensure that they carefully define the objects of their measurement according to theory, validate them and interpret them appropriately.

The widespread influence of algorithms is another source of potential error, as Claudia Wagner at the Leibniz Institute for the Social Sciences in Mannheim, Germany, and colleagues explain on page 197. They note that the algorithms that pervade our societies influence individual and group behaviour in many ways — meaning that any observations describe not just human behaviour, but also the effects of algorithms on how people behave. They argue that the theories that inform social science need to be updated to acknowledge these influences; without

© 2021 Springer Nature Limited. All rights reserved.

nature

these theories and a clear understanding of the impact of algorithms on the available data, researchers will not be able to draw meaningful conclusions.

Yet another complicating factor for computational social science is that large data sets are often the private property of commercial enterprises. Academic scientists need to liaise with corporations to obtain access, and this might introduce even more bias. This is partly because, for companies, data are valuable — and therefore sharing data is a risk to their bottom line. That is among the reasons why firms tend to restrict what they share, as Jathan Sadowski at Monash University in Melbourne, Australia, and colleagues highlight on page 169. But in light of the potential of these data to provide societal benefits, companies — together with academic researchers and public bodies — need to engage with these questions and set standards for quality, access and data ownership.

Ways forward

There are ways to obtain useful and reliable data, as Mirta Galesic at the Santa Fe Institute in New Mexico and colleagues describe on page 214 in an article on 'human social sensing'. This is the study of how individuals gather information on others in their social networks. For instance, researchers could predict a swing in political opinions by interviewing people and asking them what their friends are talking about. Gathering data about people from other people can help to avoid some of the biases seen in self-reported data, and has the added benefit of generating anonymous data: the researchers never need to know any personal or sensitive details about the people on whom they are receiving information.

Another area ripe for growth lies at the intersection of infectious-disease modelling and behavioural science. As Caroline Buckee at the Harvard T. H. Chan School of Public Health in Boston and colleagues argue on page 205, an accurate model of contagion and infection requires researchers to understand the cultures and behaviours of people who have been — or might be — infected. It is hard to predict a disease's path without considering these and other social aspects of transmission. Structured and widespread collaborations cutting across disciplines are key to achieving this.

The pandemic has shown that lives can be saved when large-scale data sets are harnessed for science. This potential is only starting to be realized as researchers with backgrounds in computer science or applied mathematics join with social scientists. These relationships must deepen and expand to encompass researchers in more fields — such as ethics, responsible research, and science and technology studies — to ensure that we avoid known pitfalls and that we use these data in a way that maximizes gained knowledge and minimizes potential harm.

Transdisciplinary co-working is rarely easy, but it is essential for both better decisions and robust outcomes. *Nature* is committed to fostering this conversation and helping scientists to learn each other's languages so that, together, researchers can make more progress on some of societies' most pressing problems.

> **"Research managers are becoming advocates for responsible research."**

Research managers are researchers, too

Academic administrators have a key part to play in improving research culture.

In the space of three decades, academic research management has become an attractive career prospect for researchers around the world. Once focused principally on helping academics to manage funding, research managers and administrators (RMAs) are now part of a globally recognized profession that spans the research spectrum. There are some 20,000 RMAs working in universities; most are in high-income countries, but expansion is under way in lower-income nations, particularly in Africa.

The role has evolved as research has become more complex, and this, in turn, is attracting more candidates with research-level qualifications and experience. Today's managers and administrators need knowledge and experience of open science, equality and diversity, ethics and public engagement — as well as of more conventional areas such as accounting, project management and research policy.

RMA courses and qualifications are now offered by universities and by some of the 20 national and regional professional associations belonging to the International Network of Research Management Societies (INORMS).

But, as we report in this issue (see page 321), tensions between RMAs and the researchers they work with are not uncommon. There are still those who regard the academic as 'king' and the RMA as little more than research support. Meanwhile, at some institutions, university leaders expect RMAs to monitor academics' performance metrics — such as targets for publishing and research income — which can be stressful for both researchers and managers.

As a result, RMAs and their professional organizations are becoming advocates for responsible research. And they are embracing the academic study of research management and administration. This is helping to establish good practice, as well as professional standards that can be used to hold universities and publishers to account.

For example, members of INORMS are taking a lead in addressing how university league tables might be improved to make them fairer and more transparent. And the UK research managers association, ARMA, has been involved in an independent review on the use of metrics in research evaluation, a project called the Metric Tide. This year also saw the launch by management professionals of the *Journal of Research Management and Administration*.

These are welcome developments. RMAs are crucial to the research enterprise. Moreover, their involvement in active scholarship is essential to achieving the aims set out above. Researchers and managers must work collegially and respectfully to make the research environment happier and more productive.

© 2021 Springer Nature Limited. All rights reserved.