

# A New Science for Describing Unhealthy Online Environments

A theory derived from nonlinear fluid dynamics is able to reproduce the formation dynamics of online hate communities—offering insights that could inform public policies.

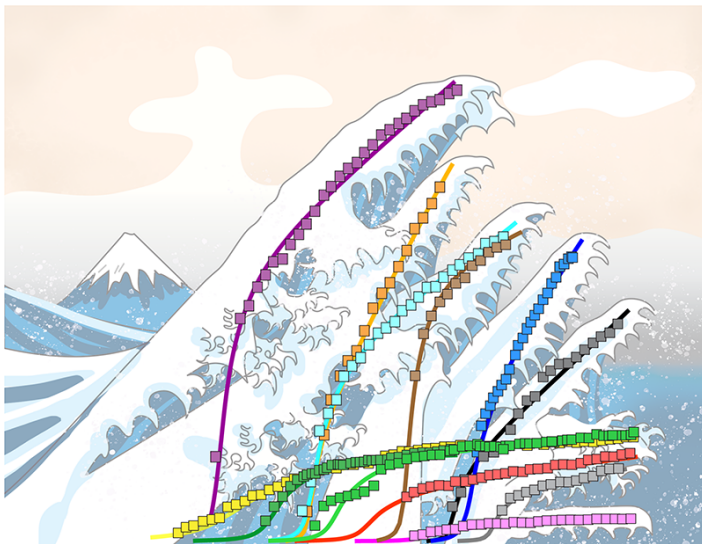
By Elisabeth Wilhelm

The diffusion of harmful online content is becoming one of the most pressing problems of our society—a trend that the COVID-19 pandemic has dramatically accelerated. Such content drives and amplifies dangerous forms of behaviors. In public health practice, for instance, we have observed that online misinformation impacted people’s trust in health authorities and in measures including social

distancing, masking recommendations, vaccines, and therapies. Despite the urgency, a thorough description of online information dynamics, which would be essential to inform the formulation of public policies, is missing. Now Pedro Manrique and colleagues at George Washington University in Washington, DC, have tackled one important aspect of such dynamics [1] (Fig. 1). They presented a “first principles” theory, derived from nonlinear fluid dynamics and nonequilibrium statistical physics, that captures the formation of online communities supporting “anti-X” hate, where X can stand for categories related to religion, science, ethnicity, race, and more. The theory not only explains how harmful online activity develops but also gives hints on how this development might be slowed or even prevented by adjusting what the researchers call the online collective chemistry.

The COVID-19 pandemic has renewed interest in systematic approaches that account for risks related to online information. Since 2020, initiatives led by the World Health Organization have brought together practitioners, policymakers, and experts in scientific fields as diverse as law, behavioral science, digital health, user experience, information science, and physics with the goal of guiding the global “infodemiology” research agenda. These initiatives have identified ways for assessing the impact of health misinformation but have also suggested promising mitigation strategies, some of which have been vetted in small-scale tests involving a few hundred to a few thousand online users. But the problem entails an enormous challenge of scale: 5 billion or so people are online, and new users keep adding to this tally every day.

A new science that rigorously describes anti-X activity at the



**Figure 1:** Manrique and colleagues have developed a new theory whose predictions (solid line) match data on the dynamics of online “anti-X” communities (square symbols). The theory, derived from fluid dynamics, shows that the growth of these communities is reminiscent of the growth of nonlinear waves in a liquid (sketched in the background).

Credit: APS/Carin Cain

required scale must account for the special features of online dynamics. In online spaces, people don't just exchange information, they interact with each other and form communities in support of or in opposition to certain shared likes, values, or opinions. The formation dynamics of anti-X groups can be extremely complex. These groups may appear out of nowhere and grow extremely quickly, but also suddenly disappear when shut down by moderators. Multiple groups may undergo "fusion" and form larger groups, while "fission" may lead some groups to split into smaller entities. Individuals and groups may also have "personalities" that evolve with time. Backed up by a large, empirical dataset, Manrique and colleagues have now provided a scientific description of these complex dynamics.

The researchers' key insight is that online communities can be treated like a fluid. The correct description of boiling water does not come from considering water molecules one at a time but from taking into account the correlated pockets (bubbles) that the molecules form. Similarly, the correct science of how an online system "boils" is based on a proper description of the correlated "pockets of users," that is, the online communities and how these communities interconnect. Building on this analogy, Manrique and co-workers apply nonlinear fluid dynamics to describe mathematically how collections of different individuals aggregate into communities and how communities then aggregate and evolve within internet platforms or across multiple platforms. The intriguing picture emerging from this model is that the characteristic features of anti-X dynamics are reminiscent of that of shockwaves in a fluid—those that can produce abrupt changes in macroscopic properties such as pressure, temperature, or density.

While prior studies have tackled this problem, they haven't been able to account for some of the complex features of these dynamical systems. In particular, they failed to capture the cumulative impact that the reverberation and amplification of information may have on people's behaviors and on online-community formation. Even seemingly benign and inconsequential pieces of information can have dramatic effects because of the social dynamics they trigger—an aspect that the new theory successfully captures. What's more, the analytical description put forward by Manrique and co-workers can in principle tackle problems at any scale.

The researchers show a remarkable agreement between their model's predictions and a huge dataset that they have been collecting since 2014, covering platforms ranging from Facebook to the Russian platform VK. Their formalism successfully reproduces, for instance, the empirical shape of the growth curves of pro-ISIS communities on VK and of anti-government communities on Facebook related to the US Capitol riot. In their model, a significant fraction of the total population can abruptly condense into a single large cluster, a shockwave.

Importantly, the model shows that the community dynamics can be controlled by acting on a parameter, dubbed online collective chemistry, that quantifies the average probability of fusion between different groups. This observation allowed the team to evaluate two general mitigation scenarios as well as to explain why in some cases removing individuals or organizations from online conversations or even whole communities from online platforms doesn't prevent them from reforming. These insights will be useful in informing approaches to address health misinformation in future emergencies.

Manrique and colleagues' rigorous mathematical description, grounded in fluid dynamics and vetted by empirical data, offers a general model that might be applicable to a wide range of online threats. Similar approaches will become even more important in the future with the emergence of new online platforms and services, of gaming technology that parents aren't able to supervise, and of novel AI tools—from ChatGPT to content promotion and moderation algorithms—that will disrupt the information environment.

So just as we demand a robust scientific basis when discussing nuclear power or climate science, we must do the same for online-information problems that can have severe consequences for our society. The work of Manrique and colleagues is an encouraging step toward the development of a scientific language that could describe these phenomena and guide science-based strategies for tackling them.

### Acknowledgement

The author acknowledges the contribution of Tina D. Purnat, with the Department of Epidemic and Pandemic Preparedness and Prevention of the World Health Organization, to the conceptualization of this article.

---

**Elisabeth Wilhelm:** Information Futures Lab, School of Public Health, Brown University, Providence, Rhode Island

#### REFERENCES

1. P. D. Manrique *et al.*, “Shockwavelike behavior across social media,” *Phys. Rev. Lett.* **130**, 237401 (2023).