this second excited state: 29.19 keV. Finally, the nucleus decayed directly to the isomeric state. The approach of Masuda *et al.* could enable this state to be produced more efficiently than was previously possible.

In Seiferle and colleagues' experiment, a beam of thorium-229 ions was generated from the natural decay of uranium-233 ions. About 2% of the thorium ions were in the isomeric state. These ions were then neutralized to allow them to decay to the ground state through a process called internal conversion. In this process, a nuclear decay that would typically produce a γ-ray instead causes the neutral atom to emit an electron (Fig. 1). However, internal conversion is complicated, because the electron can originate from many different energy levels in the neutral atom.

To observe the ejected electrons from internal conversion, Seiferle and co-workers used a magnetic field to bend the trajectory of these particles towards an electron detector. They applied an electric field to the electrons until the voltage associated with this field was large enough to stop the electrons. The final voltage was equal to the initial energy of the electrons. Seiferle *et al.* then used a theoretical model to interpret the electron energy spectrum, which is the first energy spectrum observed from the decay products of the isomeric state. Their analysis indicated that the energy of the isomeric state is 8.28 ± 0.17 eV.

Although the ultimate and groundbreaking goal of directly observing the thorium-229 isomeric transition remains elusive, substantial progress continues to be made. The results of Masuda *et al.* and Seiferle *et al.* are key steps forward. Hopefully, the observation is not too far off, as teams of scientists race to make the world's first nuclear clock, which would offer unprecedented precision. This finding would enable a whole host of experiments and discoveries in the decades to follow. For instance, a nuclear clock could have applications in dark-matter research[8] and in the observation of possible variations in the fundamental constants of physics[9]. ■

Jason T. Burke *is in the Nuclear and Particle Physics Group, Nuclear and Chemical Sciences Division, Physical and Life Sciences Directorate, Lawrence Livermore National Laboratory, Livermore, California 94550, USA.*
e-mail: burke26@llnl.gov

1. Masuda, T. *et al. Nature* **573**, 238–242 (2019).
2. Seiferle, B. *et al. Nature* **573**, 243–246 (2019).
3. Lyons, H. *Instruments* **22**, 133–135 (1949).
4. Brewer, S. M. *et al. Phys. Rev. Lett.* **123**, 033201 (2019).
5. Campbell, C. J. *et al. Phys. Rev. Lett.* **108**, 120802 (2012).
6. Peik, E. & Tamm, C. *Europhys. Lett.* **61**, 181–186 (2003).
7. Beck, B. R. *et al. Phys. Rev. Lett.* **98**, 142501 (2007).
8. Derevianko, A. & Pospelov, M. *Nature Phys.* **10**, 933–936 (2014).
9. Flambaum, V. V. *Phys. Rev. Lett.* **97**, 092502 (2006).

SOCIAL SCIENCE

The dynamics of online hate

An analysis of the dynamics of online hate groups on social-media platforms reveals why current methods to ban hate content are ineffective, and provides the basis for four potential strategies to combat online hate. SEE LETTER P.261

NOEMI DERZSY

How does the online hate ecosystem persist on social-media platforms, and what measures can be taken to effectively reduce its presence? On page 261, Johnson *et al.*[1] address these questions in a captivating report on the behaviour of online hate communities that reside on multiple social-media platforms. The authors shed light on the structure and dynamics of online hate groups and, informed by the results, propose four policies to reduce hate content on online social media.

We live in an age of high social interconnectedness, whereby opinions shared in one geographical region do not remain spatially localized, but can spread rapidly around the globe thanks to online social media. The high speed of such diffusion poses problems for those policing hate speech, and creates opportunities for nefarious organizations to share their messages and expand their recruiting efforts globally. When the policing of social media is inefficient, the online ecosystem can become a powerful radicalizing instrument[2]. Understanding the mechanisms that govern hate-community dynamics is thus crucial to proposing effective measures to combat such organizations in this online battleground.

Johnson *et al.* examined the dynamics of hate clusters on two social-media platforms, Facebook and VKontakte, over a period of a few months. Clusters were defined as online pages or groups that organized individuals who shared similar views, interests or declared purposes, into communities. These pages and groups on social-media platforms contain links to other clusters with similar content that users can join. Through these links, the authors established the network connections between clusters, and could track how members of one cluster also joined other clusters. Two clusters (groups or pages) were considered connected if they contained links to one another. The authors' approach had the advantage of not requiring individual-level information about users who are members of clusters.

Johnson *et al.* show that online hate groups are organized in highly resilient clusters. The users in these clusters are not geographically localized, but are globally interconnected by 'highways' that facilitate the spread of online hate across different countries, continents and languages. When these clusters are attacked — for example, when hate groups are removed by social-media platform administrators (Fig. 1) — the clusters rapidly rewire and repair themselves, and strong bonds are made between clusters, formed by users shared between them, analogous to covalent chemical bonds. In some cases, two or more small clusters can even merge to form a large cluster, in a process the authors liken to the fusion of two atomic nuclei. Using their mathematical model, the authors demonstrated that banning hate content on a single platform aggravates online hate ecosystems and promotes the creation of clusters that are not detectable by platform policing (which the authors call 'dark pools'), where hate content can thrive unchecked.

Online social-media platforms are challenging to regulate, and policymakers have struggled to suggest practicable ways of reducing hate online. Efforts to ban and remove hate-related content have proved ineffective[3,4]. Over the past few years, the incidence of reports of hate speech online has been rising[5], indicating that the battle against the diffusion of hateful content is being lost, an unsettling direction for the well-being and safety of our society. Furthermore, exposure to and engagement with online hate on social media has been suggested to promote offline aggression[6], with some perpetrators of violent hate crimes reported to have engaged with such content[7].

Previous studies (for example, ref. 8) have considered hate groups as individual networks, or considered the interconnected clusters together as one global network. In their fresh approach, Johnson and colleagues studied the interconnected structure of a community of hate clusters as a 'network of networks'[9–11], in which clusters are networks that are interconnected by highways. Moreover, they propose four policies for effective intervention that are informed by the mechanisms their study revealed govern the structure and dynamics of the online-hate ecosystem.

Currently, social-media companies must decide which content to ban, but often have to contend with overwhelming volumes of content and various legal and regulatory

GORDON WELTERS/NYT/REDUX/EYEVINE

Figure 1 | Facebook moderators removing hate-related content. Johnson *et al.*[1] examined the dynamics of online hate groups on Facebook and another social-media platform, VKontakte, and used their results to propose four policies to tackle online hate.

constraints in different countries. Johnson and co-workers' four recommended interventions — policies 1 to 4 — take into account the legal considerations associated with banning groups and individual users. Notably, each of the authors' suggested policies could be implemented independently by individual platforms without the need for sharing sensitive information between them, which in most cases is not legally allowed without explicit user consent.

In policy 1, the authors propose banning relatively small hate clusters, rather than removing the largest online hate cluster. This policy leverages the authors' finding that the size distribution of online hate clusters follows a power-law trend, such that most clusters are small and only very few are large. Banning the largest hate cluster would be predicted to lead to the formation of a new large cluster from the myriad small ones. By contrast, small clusters are highly abundant — meaning that they are relatively easy to locate — and eliminating them prevents the emergence of other large clusters.

Banning whole groups of users, regardless of the size of the groups, can result in outrage in the hate community and allegations against social-media platforms that rights to free speech are being suppressed[12]. To avoid that, policy 2 instead recommends banning a small number of users selected at random from online hate clusters. This random-targeting approach does not require users to be spatially located or the use of sensitive user-profile information (which cannot be applied to target specific users), thus avoiding potential violations of privacy regulations. However, the effectiveness of this approach depends heavily on the structure of the social network, because the topological characteristics of networks strongly shape their resilience to random failures or targeted attacks.

Policy 3 leverages the finding that clusters self-organize from an initially disordered group of users; it recommends that platform administrators promote the organization of clusters of anti-hate users, which could serve as a 'human immune system' to fight and counteract hate clusters. Policy 4 exploits the fact that many hate groups online have opposing views. The policy suggests that the platform administrators introduce an artificial group of users to encourage interactions between hate clusters that have opposing views, with a view to the hate clusters subsequently battling out their differences among themselves. The authors' modelling demonstrated that such battles would effectively remove large hate clusters that have opposing views. Once put into action, policies 3 and 4 would require little direct intervention by the platform administrators; however, setting opposing clusters against each other would require meticulous engineering.

The authors recommend caution in assessing the advantages and disadvantages of adopting each policy, because the feasibility of implementing a policy will rely on available computational and human resources, and legal privacy constraints. Moreover, any decisions about whether to implement one policy over another must be made on the basis of empirical analysis and data obtained by closely monitoring these clusters.

Over the years, it has become apparent that effective solutions to dealing with online hate and the legal and privacy issues that arise from online social-media platforms cannot arise solely from individual industry segments, but instead will require a combined effort from technology companies, policymakers and researchers. Johnson and colleagues' study provides valuable insights, and their proposed policies can serve as a guideline for future efforts. ■

Noemi Derzsy *is in the Data Science and AI Research Organization, AT&T Labs, New York, New York 10007, USA.*
e-mail: noemiderzsy@gmail.com

1. Johnson, N. F. *et al. Nature* **573**, 261–265 (2019).
2. Hernandez, D. & Olson, P. *Wall Street J.* (5 July 2019); available at go.nature.com/2oxoqdw
3. Wakefield, J. BBC News (15 March 2019); available at go.nature.com/2kabi2p
4. O'Brien, S. A. CNN Business (28 February 2019); available at go.nature.com/31c2nny
5. BBC News (17 March 2018); available at go.nature.com/2h04gci
6. SELMA (23 April 2019); available at go.nature.com/2ygkhs1
7. Benner, K. & Spencer, H. *New York Times* (27 June 2019).
8. Mathew, B., Dutt, R., Goyal, P. & Mukherjee, A. *Proc. 10th ACM Conf. Web Sci.* 173–182 (2019).
9. Havlin, S., Kenett, D. Y., Bashan, A., Gao, J. & Stanley, H. E. *Eur. Phys. J. Spec. Top.* **223**, 2087–2106 (2014).
10. Palla, G., Barabasi, A. L. & Vicsek, T. *Nature* **446**, 664–667 (2007).
11. Jarrett, T. C., Ashton, D. J., Fricker, M. & Johnson, N. F. *Phys. Rev. E* **74**, 026116 (2006).
12. Coaston, J. *Vox* (14 May 2019).

Noemi Derzsy contributed to this article in her personal capacity; the views expressed are her own and do not necessarily represent the views of AT&T.

This article was published online on 21 August 2019.