

## 小端网络观察：一个“怒”=五个讚？！Facebook曾用表情符号放大仇恨？

这几年假新闻，仇恨及争议内容在脸书上泛滥，伤害或许已经造成。



2018年4月26日，伦敦英国议会委员会听证会场地外，有戴著Facebook 创办人 Mark Elliot Zuckerberg 的面具及愤怒表情面具的示威者。摄：Simon Dawson/Bloomberg via Getty Images

端传媒实习记者 郑又祯 | 2021-10-28

“小端网络观察”主要处理网络热议事件，简析事件原因、始末、经过及相关的网友反应，主要发表于端传媒脸书平台，为端传媒社媒组特色栏目。

你在Facebook（脸书）上喜欢用表情符号吗？华盛顿邮报昨（26）日刊出报导《Five points for anger, one for a 'like': How Facebook's formula fostered rage and misinformation》，揭露脸书从2017年开始以演算法加强推送让使用者更多的使用“表情符号”，而受众越使用怒(angry)来回复贴文，越会被演算法推送更多情绪化甚至令人愤怒的内容，而华邮的报导更指出，你按一个“怒”的影响力相当于一个“讚(like)”的五倍。这项争议措施在脸书实行多年，直到2020年才得到较大幅度修正。

这也是继日前脸书前员工法兰西斯·豪根（Frances Haugen）在美国国会听证会中以大量文件指控脸书为高流量用户放松事实查核、高层明知Instagram对青少年身心有害却不采取任何改善行动、在公众利益与集团获利之间，选择极大化集团获利等多项问题之后，再有报导揭露脸书争议。

到底脸书如何让你的页面充满愤怒内容的呢？ **在脸书“按个怒”到底意味著什么？**

你点赞，点爱心或者点怒，你的每次贴文互动都会决定刷新脸书时动态消息将显示的内容。根据脸书内部文件揭示，脸书有超过万种方式衡量贴文显示顺序的标准，其中一些更细微到用户难以察觉，例如一则贴文会产生多少长篇的留言、影片是直播的还是事先录制的、留言是纯文字还是用卡通人物贴图，甚至还会计算一则贴文会占用电脑多少记忆体，和使用者的网路讯号强弱等等。

脸书在五年前推出除了赞以外的五种表情符号反应，包含“爱心”、“哈哈”、“哇”、“哭泣”、“怒”等。华邮引述文件内容指出，从2017年开始，脸书演算法为因应平台用户在发布贴文和留言互动上的下滑，将使用者按下表情符号的反馈，权重设定为单纯按赞的五倍，原因在于这类激出更多情绪反应的贴文，让使用者的参与度更高——而这点，也正是脸书的利基所在。

值得注意的是，华邮报导更指出，当表情符号作为一种贴文“加分”手段，碰上了旨在限制有害内容的“扣分”指标时，两者碰撞可能使问题变得复杂。脸书流出的内部文件显示，一篇贴文平均可获得几百分，但该机制并未限制单篇文章的得分上限。这意味著，当脸书演算法认为一篇贴文“不好”，此机制确实可透过演算法将其分数减半以减少曝光，但当一则贴文通过表情符号“加分”机制累积到数十亿分，即使内容有害导致扣分使分数折半，该贴文的分数可能还是高到会被推到用户动态消息的首位。

文件亦显示，脸书在推广新功能时，会将以新功能发布的贴文，透过演算法加强曝光，以鼓励用户多加使用。如此一来，便可以确保新功能为大家所知，并被广为使用，而在用户端，那些希望得到更高曝光度的发文者，例如公众帐户与政治人物，也会发现特定类型的贴文传播效果更好。

脸书也曾为促进用户使用直播功能，演算法一度将其权重拉高到600倍，这使得低品质的影片得以快速得到关注并被广为散播。而脸书的直播影片在政治性活动中也有相当重要的地位，包含去年在美国黑人乔治·佛洛伊德（George Floyd）遭杀害之后的一系列种族正义抗争，与今年1月6日的美国国会山庄暴动——在该次事件之后，脸书紧急启动先前一度取消的安全措施，让直播影片的权重由600降低至60。

## 争议机制在脸书内部争辩已久

一名匿名脸书员工曾在内部文件中指出，透过表情符号加强特定贴文推送，恐怕会让争议内容得到更高曝光度，这项预警后来也成为现实。脸书在2019年发现，获得越多“怒”的贴文，当中更有可能包含大量劣质、错误甚至有害信息。

脸书执行长马克·扎克伯格（Mark Zuckerberg）曾鼓励用户对公开贴文中的留言使用“怒”，以表达他们并不喜欢这些内容——但按当时的演算法，你越按“怒”，你的脸书反而更常显示这些内容。

脸书在2018年，第一次下调“怒”的权重，将各个表情符号的权重从此前的5个赞调整为4个赞，但因表情符号被使用的频率不一，这样的调整并未完全解决问题。到2019年7月，有员工提议将部分表情符号的权重降为一个赞，或不列入计算，因为“怒”反应与“哇”、“哈哈”等反馈，更常被使用在有毒内容和错假信息上。在当年稍晚的另一份提议中，则提及要将“大心”和“哭泣”等值于四个赞，因为他们“比较安全”，有员工甚至提议直接将整个表情符号的机制移除，但这些加权和功能在内部激辩之后，仍未做出太大的改变。

直到2020年，有研究指出脸书用户并不喜欢他们的贴文收到“怒”，不论是来自朋友或陌生人。脸书将所有表情符号的加权降低至相当于1.5个赞。同年九月，脸书终于将“怒”的权重将至零，将其移出加权计算的因素之外，同时将“大心”和“哭泣”等值于两个赞。

然而这几年假新闻，仇恨及争议内容在脸书上泛滥，伤害或许已经造成。

## “脸书”正在放大仇恨？

本周一（10/25）豪根前往英国国会，在因应《网路安全法》草案（Draft Online Safety Bill）成立的联合委员会作证时，再次强调脸书“毫无疑问地放大了仇恨”，且“不愿意为了安全，对自己的营收做出任何一点让步”。

除了华盛顿邮报的报导外，美联社（Associated Press）、纽约时报（The New York Times）、大西洋月刊（The Atlantic）等多家美国媒体，亦有使用豪根于美国国会作证当时公开的大量“脸书文件”（The Facebook Papers）进行报导。

其中美联社的报导提到，苹果在两年前，曾因Facebook和Instagram平台有涉及买卖中东女性佣人的疑虑，威胁从App Store移除两软体。另外一篇美联社报导亦指出，由于阿拉伯语在“标准阿拉伯语”之外，各地方言仍存在用字，书写的不同，这些因素都让不同地区，作为脸书第三大语言群体的阿拉伯语使用者，不一定能辨识其他地区使用的阿语。因此，不论是由演算法自动审查，或是由懂阿语的内容审查员进行人工排查，都未能精确侦测、排除真正有问题的内容，而这使得脸书在2011年阿拉伯之春时快速传播了抗争信息，但近几年却成为多个国家的仇恨内容散播的源头。

在缅甸，“穆斯林”、“伊斯兰教”等用词，经常与种族、宗教仇恨辞令与相关错假信息共同出现，而脸书的演算法缺陷，使得针对信仰伊斯兰教的罗兴亚人的仇恨言论未能被有效阻止，反而被广为推送；巴勒斯坦记者和异议人士谈论巴勒斯坦处境时，也常因为被系统判定为“违反社群守则”，而令自己的脸书帐户遭到永久删除。