TITLE:

Evolutions of "Fake News" from the South: Tracking Disinformation Innovations and Interventions between the 2016 and 2019 Philippines Elections

ABSTRACT:

A Facebook executive famously called the Philippines as "patient zero" in the so-called "global misinformation epidemic", referring to how techniques of media manipulation were observed first in the 2016 Philippine election months before the Brexit vote in the UK and Trump's election in the US. Indeed, some political pundits and journalists have expressed moral panics that Rodrigo Duterte's 2016 victory and continued popularity are outcomes of dark technological alchemy–from paid trolls to state-sponsored propaganda to even shadowy foreign influence, if we are to believe that data analytics firm Cambridge Analytica had a hand in Duterte's 2016 campaign. In the recent May 2019 Philippine midterm election where Duterte's allies shut out opposition figures from the senate races, world governments, big tech, and foreign journalists kept high alert for the export potential of emerging disinformation innovations and the threat these represent for elections in global context. With its reputation as the "social media capital of the world" and having a large precariat digital workforce seeking entrepreneurial opportunities in the digital underground, the Philippines is fertile ground for disinformation experimentation with potentially global ramifications.

This paper is based on a research intervention and public engagement effort to monitor disinformation innovations and interventions in the 2019 midterm elections. With partner researchers in the Philippines and overseas, we tracked disinformation as both emerging *technological innovations* and systematically constructed *narratives* that attempted to manipulate political discussion and deliberation. Our approach is ethnographically inspired, following on previous research that combined interviews and digital ethnography with the digital strategists, fake account operators, and influencers that comprised the 2016 "disinformation architecture" (Ong & Cabanes 2018). This enabled us to "follow the trail" through digital ethnography of the Facebook pages, YouTube channels, Twitter hashtags, and Instagram accounts that aimed for "attention-hacking" (Marwick & Lewis 2017) in the 2019 elections. With this longer view and concern for both fake news as technology and narrative, we observed how the disinformation playbook has evolved and morphed into incorporating micro-targeted techniques as well as mobilizing both old and new narratives. With a team of trained student volunteers, we published articles that fact checked the claims of candidates that took part in town hall debates. We also "fact checked" the public sphere, where we tracked fake accounts that attempted to plant questions, divert attention and hijack the conversation.

This paper thus reports on our main findings for the 2019 midterm elections, where we observed four key developments. First, campaigners experimented with new strategies of micro-targeting that seeded specific political ads and messages aimed at discrete groups of potential voters. Second, digital underground operations of mobilizing click armies to hack public attention and circulate insidious narratives became more prevalent, if strategically obscured and harder to spot. Third, while both administration and opposition candidates shared emotionally manipulative and disinformative propaganda, their underlying narratives ranging from hypernationalism to historical revisionism are

different. Lastly, disinformation interventions from big tech and government have so far only gone after easy targets. We expound on these findings in the succeeding paragraphs, and conclude with recommendations for regulatory approaches moving forward.

The rise of micro-targeted content as well as micro- and nano-influencers comes on the heels of frequent criticisms and even content regulation for macro-influencers, which led the disinformation drive in the 2016 elections. What micro- influencers lack in broader reach, they gain in manoeuvrability and 'contrived authenticity.' Contrived authenticity is the term media anthropologists use to describe internet celebrities whose carefully calculated posts seek to give an impression of raw aesthetic, spontaneity and therefore relatability. This makes it easier for them to infiltrate organic communities and evade public monitoring. We classify these micro- and nano-influencers into three categories: political parody accounts, pop culture accounts, and thirsttrap instagrammers. Unlike celebrities or the aforementioned political bloggers with millions of social media followers and mainstream media visibility, these personalities cultivate more intimate and interactive relationships with their fans. When they post an election- related joke, tweet a hashtag, or share a candidate's video, their message comes across as spontaneous and sincere. Their 'authentic' exuberance for a political cause becomes an aspirational model for their followers' own political performance.

Another form of micromedia manipulation we observed is the diversification of alternative news across social media platforms. Hyper-partisan news channels continue to pose as imposters of news sites while manufacturing more contrived authenticity. Meanwhile, thematic and local news pages try to appear nonpartisan by curating news with a particular focus, but occasionally slip in clients' political propaganda, such as their track record in office or their campaign promises.

These first two innovations make regulation difficult because of the increase in disinformative actors. First, the sheer volume of the actors makes it harder to monitor and regulate all of them. Second, the change in following size also makes it difficult to observe all disinformative actors since micro-targeting now means content is dispersed across different interest groups online, requiring tracing more than just one flow of information. The shift from macro-influences to micro-influencers means increased difficulty in following the disinformation trail online. These accounts operate in a coordinated manner, making it possible to see a fraction of the disinformation activity online and regulating them, but the entire network is still missed. Even with platform regulation leading to hundreds of page takedowns by Facebook, more accounts manage to evade sanctions since they now operate at smaller and less perceptible scales. Furthermore, the thrust of Facebook's regulation is more towards page behavior than content. This means while general patterns of questionable behaviour are checked, social media platform's community standards for posting are seemingly insufficient in preventing disinformation.

The last innovation we want to highlight is the more prolific use of Facebook closed groups in spreading election- related disinformation. Facebook closed groups cater to diverse communities. Disinformation tactics here can either be explicitly political and hyper-partisan or discreetly seeding paid political propaganda. Closed groups are not subject to takedowns due to their privacy protections. Instead, there are group administrators or moderators who police the group and prevent potential spies from opposing political camps from gaining access. The moderators ban individual 'bad actors' within the closed groups when they are found violating the group's rules of communication exchange. Identifying 'bad actors' is especially difficult when closed groups often operate as an echo chamber, or filter bubble, of zealous fans who affirm each other's beliefs. Moderators also actively promote 'authenticity' in these groups by encouraging visual displays of community and sociality. Moderators encourage members to post selfies of themselves during holidays to neutralise potential critiques of 'fake' interactions.

Our paper offers reflection on the opportunities and risks facing disinformation interventions in the global South through the particular experience of the Philippines. Disinformation tactics are fast-evolving, creative, and increasingly undetectable. Even in the best of circumstances, regulation on content moderation, among others, can only do so much to anticipate innovations of digital underground operations designed precisely to circumvent existing rules. The challenge, therefore, is broader and more systemic. Avenues for reform need to take more seriously the enabling environment for disinformation to thrive, from the vulnerabilities of digital workers that drive them to join underground operations to the ethics of the advertising industry that allows unscrupulous practices unchecked.

We suggest that there is no one-size-fits-all solution to the complex problem of digital disinformation. Understanding local contexts of disinformation production and the ways that architects of disinformation evade responsibility and entice other workers to join them in the digital underground, allows us to suggest bespoke interventions that, for example, spotlight the culpability of the advertising and PR industries in the normalization and expansion of an underground industry of political trolling that only hides in plain sight.

We also highlight how national government regulation of social media also poses social and political risks. The Philippines' proximity to other Southeast Asian countries with fake news laws that can be weaponized to curb free speech and silence dissent heightens the risk of chilling legislation to snowball to the Philippine context, accelerating the current regime's creeping authoritarianism. We argue for a shift in regulatory interventions around social media from content regulation to process regulation. While content regulation often bristles against liberal principles of free speech particularly when universalising concepts such as 'hate speech' are applied to diverse cultural contexts, process regulation upholds instead the virtues of transparency, accountability, and fairness in the conduct of political campaigns as well as the decision-making around social media regulation. We argue that regulatory reform should be guided by bold ethical principles but it should always be handled with a legal soft touch so as not to compromise liberal principles of free speech and tolerance.

A process-oriented approach can apply to both political campaigning, and fact checks and platform bans online. For political campaigning, we need to shine a light on the ways in which contemporary campaigns are funded, managed, and executed. This requires shifting regulatory impulses from banning or censoring to openness through disclosure. The first step to take is to continue a public conversation about the scale of the issue, and how deep incentives go within local advertising industries.

The second step is to review possibilities for a broader legal framework that might encourage transparency and accountability in political consultancies. In this light a legal framework for a Political Campaign Transparency Act might provide opportunities to create better checks- and-balances in political consultancy work arrangements, campaign finance disclosures, and campaign donations of 'outsourced' digital strategy.

The third step is to review the Commission on Election (COMELEC)'s existing frameworks for campaign finance and social media regulation. While guidelines have improved to create transparency and accountability in social media campaigning in 2019, the current framework also has several vulnerabilities, particularly in its extensive focus on the reporting and monitoring of politicians' official social media accounts, and requirement of attaching receipts of transactions. As our study has shown, digital campaigns involve both official and underground operations. Facebook ads, influencer collaborations, and many political consultancies do not have formal documentation and fails the

requirement. This loophole enables politicians to skirt responsibility to report on informal work arrangements.

Finally, we encourage COMELEC to form intersectoral alliances with the academe, civil society, and industry in the monitoring of traditional and digital campaigns. The issue of precarious digital workers being lured to underground operations is a matter of labour rights, and so collaboration with labour groups could be explored to provide support to overworked and exploited digital workers who wish to report the perils of their jobs. Industry players could also begin developing collaborative strategies to develop a code of conduct or a broad framework for political advertising.

On the other hand, a process-oriented policy model establishes the necessary conditions for social media content moderation in the context of elections. Even prior to the stricter enforcement of community standards and ad policies, a process-oriented approach to election-related content moderation means developing transparent and context-sensitive systems of institutional cooperation between social media firms and local election commissions, fact checkers, and civil society watchdogs. This means creating inclusive oversight committees that can inform social media firms of local standards and concerns around 'harmful' and 'inauthentic' content without compromising free speech or falling into political partisanship. A process-oriented model means establishing transparency and feedback loops between social media firms and their local collaborators.

REFERENCES

Abidin, Crystal. 2017. "Vote for My Selfie: Politician Selfies as Charismatic Engagement." In Selfie Citizenship, edited by Adi Kuntsman, 75–87. New York: Springer.

Cabañes, Jason, CW Anderson, and Jonathan Corpus Ong. 2019. "Fake News and Scandal." In The Routledge Companion to Media and Scandal, edited by Howard Timber and Silvio Waisbord. London: Routledge.

Casilli, Antonio. 2016. "Digital Labor Studies Go Global: Toward a Digital Decolonial Turn." International Journal of Communication 11: 3934-3954.

Chakravartty, Paula and Srirupa Roy. 2017. "Mediatized Populisms: Inter-Asian Lineages." International Journal of Communication 11 (2017): 4073–92.

De Guzman, Warren. 2019. "EXCLUSIVE: Twinmark Media earned millions of dollars before Facebook takedown." ABS- CBN News, February 28. Accessed July 1, 2019. https://news. abs-cbn.com/news/02/26/19/exclusive-twinmark-media-earned- millions-of-dollars-before-facebook-takedown.

Guess, Andrew, Jonathan Nagler, and Joshua Tucker. 2019. "Less than you think: Prevalence and predictors of fake news dissemination on Facebook." Science Advances 5 (9). Published electronically January 9, 2019. DOI: 10.1126/sciadv. aau4586.

Harbath, Katie. 2018. "360/OS: Facebook's Katie Harbath on protecting election integrity." YouTube video, 13:55. Posted by "Rappler," June 23, https://www.youtube.com/ watch?v=dJ1wcpsOtS4.

Hau, Caroline. 2018. "Why I Distrust Solita Monsod's 'Why Filipinos Distrust China.'" Esquire, November 27. Accessed July 21, 2019. https://www.esquiremag.ph/politics/opinion/caroline- hau-winnie-monsod-a2262-20181128-lfrm.

Lazer, David M. J., Matthew A. Baum, Yochai Benkler, Adam J. Berinsky, Kelly M. Greenhill, Filippo Menczer, Miriam J. Metzger, Brendan Nyhan, Gordon Pennycook, David Rothschild, Michael Schudson, Steven A. Sloman, Cass R. Sunstein, Emily A. Thorson, Duncan J. Watts, and Jonathan L. Zittrain. 2018. "The Science of Fake News." Science 359 (6380): 1094—96.

Lewis, Rebecca. 2019. Alternative Influence: Broadcasting the Reactionary Right on YouTube. New York: Data & Society Research Institute.

Marwick, Alice and Rebecca Lewis. 2017. Media Manipulation and Disinformation Online. New York: Data & Society Research Institute.

Ong, Jonathan Corpus and Jason Vincent Cabañes. 2018. Architects of Networked Disinformation: Behind the Scenes of Troll Accounts and Fake News Production in the Philippines. Accessed July 1, 2019. https://newtontechfordev.com/wp- content/uploads/2018/02/ARCHITECTS-OF-NETWORKED-DISINFORMATION-FULL-REPORT.pdf.

Pazzibugan, Dona. 2017. "Pro-Duterte blogger quits DFA contract." Inquirer.net, October 11. Accessed July 1, 2019. https://newsinfo.inquirer.net/937020/pro-duterte-blogger-quits- dfa-contract.

Philippine Center for Investigative Journalism. 2019. "PCIJ Audit: Candidates on Social Media." February 14. Accessed July 1, 2019. https://pcij.org/stories/pcij-audit- candidates-on-social-media/.

Pohjonen, Matti and Sahana Udupa. 2017. "Extreme Speech Online: An Anthropological Critique of Hate Speech Debates." International Journal of Communication 11 (2017): 1173–91.

Punongbayan, JC. 2019. "Why the influx of Chinese in the Philippines?" Rappler, June 17. Accessed July 16, 2019. https:// www.rappler.com/thought-leaders/233238-reasons-influx- chinese-philippines.

Ranada, Pia. 2018. "PCOO Assistant Secretary Mocha Uson resigns." Rappler, October 3. Accessed June 28, 2019. https:// www.rappler.com/nation/213380-mocha-uson-resigns-from- pcoo-october-2018.

Rappler. 2017. "Mocha Uson named PCOO assistant secretary." May 9. Accessed July 1, 2019. https://www.rappler. com/nation/169371-mocha-uson-presidential-communications-operations-office-assistant-secretary. **Ressa, Maria**. 2016. "Propaganda War: Weaponizing the Internet." Rappler, October 3. Accessed July 1, 2019. https:// www.rappler.com/nation/148007-propaganda-war-weaponizing- internet.

Sharma, Amogh Dhar. 2019. "How far can political parties in India be made accountable for their digital propaganda?" Scroll.in, May 10. Accessed July 1, 2019. https://scroll.in/ article/921340/how-far-can-political-parties-in-india-be-made- accountable-for-their-digital-propaganda.

SocialBlade. N.d. "Top 250 Youtubers in the Philippines Sorted by SB Rank." Accessed April 5, 2019. https://socialblade. com/youtube/top/country/ph.

Stuart-Santiago, Katrina. 2018. "State of Blogging, Microblogging, Media #Crisis." Katrina Stuart Santiago Blog, January 9. Accessed July 1, 2019. https://katrinasantiago.com/ state-of-blogging-microblogging-media-crisis/.

Tapsell, Ross. 2019. "The Smartphone as the 'Weapon of the Weak': Assessing the Role of Communication Technologies in Malaysia's Regime Change." Journal of Current Southeast Asian Affairs 37 (3): 9–29.

Wardle, Claire and Hossein Derakhshan. 2017. "Information Disorder: Toward an Interdisciplinary Framework for Research and Policy Making." Accessed July 13. https:// rm.coe.int/information-disorder-toward-an-interdisciplinary- framework-for-researc/168076277c.

Wissman, Barrett. 2018. "Micro-Influencers: The Marketing Force of The Future?" Forbes, March 2. Accessed July 1. https:// www.forbes.com/sites/barrettwissman/2018/03/02/micro- influencers-the-marketing-force-of-the-future/#6b90cd756707.

Wong, Tessa. 2019. "Singapore fake news law polices chats and online platforms." BBC, May 9. Accessed July 1, 2019. https:// www.bbc.com/news/world-asia-48196985.