What comes after access: hate, harassment and manipulation?
Discussing papers

P. Ramasoota

Exploring Digital Fake News Phenomena in Indonesia *by Riri Kusumarani, HangJung Zo*

- *Very well-executioned research with impressive results, a little alteration on RQ suggested*

- **Research question:**
  - What Types & Channels of Digital Fake News are Commonly Found in Indonesia?
  - How Digital Fake News Spread in Indonesia?

- The first RQ was well tackled with these research activities

    1) categorizing types of fake news that are retrieved from the database based on criteria established in the reviewed literature which the researcher did really well and was a commendable exercise when two sets of criteria were cross-tabulated to give a more thorough dimensional analysis ;

    2) classifying fake news based on topic, fake news type and keywords;

    3) elaborating on selected cases of fake news that represent different categories outlined

    4) PEST analysis to reflect upon the chosen cases and to attempt to give explanations as to why these fake news spread in Indonesia as they had

Exploring Digital Fake News Phenomena in Indonesia *by Riri Kusumarani, HangJung Zo*

- In order to study the pattern of **how** fake news spread digitally in Indonesia, a different research design is needed – analysis of online information flows, content analysis (of postings and keywords) using data collected through mining and social listening scheme together with data analytics in addition to surveys/interviews with users and agents of fake news scheme.

- Other factors that influence the spread should be given consideration -- the attributes of each online platform (user interface, online grammar), role of social influencers (apart from the buzzers), the type and topic of the news, the presence of a networked scheme(?)

- Perhaps 2nd RQ should be revised to reflect the substance of the research.
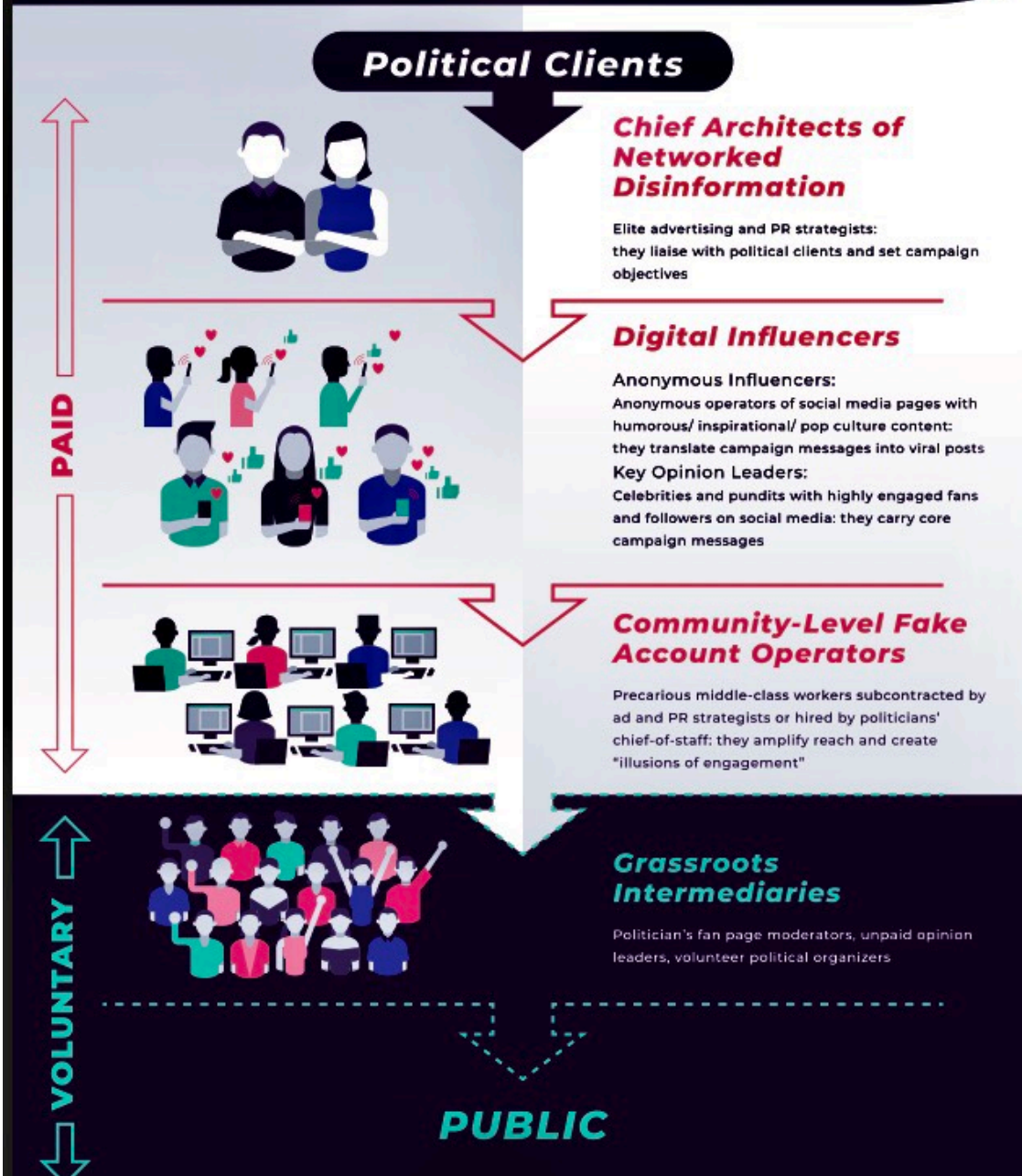
ASEAN case of disinformation campaign

- In the Philippines, social media influencers who command large "troll armies" have been credited with sweeping Rodrigo Duterte into unforeseen election victory in 2016. Under Duterte's presidency, "trolls", or "**Dutertards**" – nickname for his fanatic followers -- are seen to have debased political discourse and silenced dissidents through their extensive sharing of **fake news** and amplification of **hate speech**. Their disinformation campaign is well-designed and features a networked scheme.

BUHAY PA BA KAYO?

Portrait of a president
as a source of disinformation

THE ARCHITECTURE OF NETWORKED DISINFORMATION

Political Clients

PAID

Chief Architects of Networked Disinformation

Elite advertising and PR strategists:
they liaise with political clients and set campaign objectives

Digital Influencers

Anonymous Influencers:
Anonymous operators of social media pages with humorous/ inspirational/ pop culture content: they translate campaign messages into viral posts

Key Opinion Leaders:
Celebrities and pundits with highly engaged fans and followers on social media: they carry core campaign messages

Community-Level Fake Account Operators

Precarious middle-class workers subcontracted by ad and PR strategists or hired by politicians' chief-of-staff: they amplify reach and create "illusions of engagement"

VOLUNTARY

Grassroots Intermediaries

Politician's fan page moderators, unpaid opinion leaders, volunteer political organizers

PUBLIC

Identity, harassment and rights online: Myanmar by *Helani Galpaya*

- Well-designed and executed research, but with more contextualization perhaps
- Research questions:
  - What's the online experience of Myanmar users?
  - Do negative experiences (experienced or perceived) impact their online participation? How?
  - What role do speech and participation online play in a climate of ethnic tension, political persecution?
- Good triangulation with desk research, quantitative online survey and qualitative focus group interviews, with special focus on marginalized groups (women; LGBTQ, Muslim, non-Bamar ethnicity, those with differing political views from the mainstream) Interesting findings with respect to negative online experience and examples of harassment.
- The range of experiences presented in the survey was large and quite exhaustive while the examples from the interviews were striking and validly illustrative of the social situation surrounding Myanmar and not any online environment in general.

Identity, harassment and rights online: Myanmar by **Helani Galpaya**

- Most interesting -- the use of alternate identities (including fake accounts) to facilitate political discourse particularly in the context of a multi-ethnic society with deep-seated racial-religious conflict like Myanmar.

- While the examples of how people conceal and construct more mainstream as well as multiple identities to avoid problems in online interactions are remarkable, what would have been even more interesting is
  - the kind of dialogues these subjects engaged in using their specially constructed identities
  - the types of topics and the types of online public spheres they have been facilitated to participate using these identities
  - the extent of participation these identities have enabled them, esp. in marginalized and vulnerable groups

- The discussion on rights is quite limited, however. Apart from studying the sampled users' awareness of related laws that may curb their online expression, perhaps the research should also tap on more strategic groups – online rights-based groups, academics, advocacy groups.

Understanding and addressing online harassment: Policy initiatives to address online harassment on the Internet and social media in India, Pakistan, Bangladesh and Cambodia *by Tharaka Amarasinghe*

- *Well-carried out comparative national surveys that give broad picture of online sexual harassment experiences in selected countries. Room for improved applicability as policy evidence*

- A broad-based comparative survey of online harassment experience in four different countries – Bangladesh, Cambodia, Pakistan and India. Large sample size and carefully designed research with stratified-sampling plan to achieve representation. Research results reflect similarities as well as differences across national sets of data.

- The earlier version of the findings highlighted certain differences across different countries but the version presented today merged all national findings into one as if it was one homogenous set of data.

- A number of main variables are cross-tabulated with degrees of harassment experienced -- education, gender, urbanity, possession of smart phone, online literacy attributes (verifying information before sharing, and being selective in accepting social media friends)

Understanding and addressing online harassment: Policy initiatives to address online harassment on the Internet and social media in India, Pakistan, Bangladesh and Cambodia *by Tharaka Amarasinghe*

- Well-structured proposed policy recommendations with related sectors and recommended actions clearly outlined

- Limited usefulness as policy evidence for 3 main reasons

  1) despite the broad coverage of the findings, a comprehensive profile of social media users that are most vulnerable to online harassment is still quite . What about marginalized groups like LGBTQ, the disableds, or people with special needs. Why are they not covered as likely user groups to experience online harassment?

  2) some of the findings may not be entirely new – education and media literacy correlated with likelihood for online harassment

  3) With the way that the data is presented today, national differences are not at all delineated, making the comparative point moot.

Controlling hate speech on social media: lesson from Sri Lanka by **Yudhanjaya Wijeratne**

- Interesting research that questions the inadequate frameworks used by related players – government, social media companies – in controlling hate speech, with very provocative policy recommendations, yet marred by one-dimensional understanding of hate speech and

- The focus seems to be on detecting hate speech, accurately and efficiently. Yet, it remains unclear what actually constitutes hate speech, according to the researcher, apart from ethnic slurs and racial epithets?

- Are there degrees of severity of hate speech that need be reckoned with for the purpose of censorship?

Controlling hate speech on social media: lesson from Sri Lanka by **Yudhanjaya Wijeratne**

- Would free speech be affected?
- What are avenues in controlling hate speech, apart from censorship blocking and filtering by human content moderation?

Three essential features of hate speech

- 1) It is directed against a specifically **identifiable** individual or group of individuals based on an arbitrary or normatively irrelevant feature;

- 2) it **stigmatizes** the target group by implicitly or explicitly ascribing it to qualities widely regarded as highly undesirable;

- 3) because of its negative qualities, the target group is viewed as an **undesirable** presence and a **legitimate object of hostility**, that should not be trusted to be a loyal member of society, and presents a threat to its stability and well-being. Thus, **discrimination** against such group or even extermination of such group is **justified**.

Elements of hate speech

Objectives/severity

- **Unintentional discrimination**
- **Deliberate discrimination**
- **Inciteme**

Target group

- **Characteristic that is reflective of group identity**
  - **inherent and cannot be changed such as race, ethnicity, skin color, or place or origin**
  - **Can be changed such**

Communicating hate

- **Use of figures of speech and strategic communication**
- **Issues that spark hostility and hatred**