Image based Misinformation on WhatsApp

Kiran Garimella, Dean Eckles
WhatsApp

- Peer to peer messaging platform
- 1.5 Billion monthly active users
- Multimedia heavy – roughly 50% is images and video
- End-to-end encrypted
Image based misinformation

• Much more powerful!

• Almost no large scale studies.

• Closed platforms are important!

• Sense of intimacy and trust
Dataset

• WhatsApp

• Over 6,000 public political groups from India

• ~2 million images

• Over 6 months starting December 2018
Fact checked images

- Fact checking websites
  - ~450 images
- Manual annotation
  - 3 journalists
  - ~400 images
When were these images shared?

Number of times fake images were shared

Mostly during the India-Pakistan conflict!
Primary Categories

• Images taken out of context (~30%)
Out of context images

“Cheap fakes"
Primary Categories

- Images taken out of context (~30%)
- Photoshopping (~20%)
Photoshopped images
Primary Categories

• Images taken out of context (~30%)
• Photoshopping (~20%)
• False statistics and quotes (~10%)
Fake Quotes/Stats

“...leaders failing to perform under pressure, but PM Narendra Modi is perhaps the only leader I've come across who's unaffected under pressure.”

BILL GATES
BUSINESS MAN

---

BBC DECLARES CONGRESS AS ONE OF THE MOST CORRUPT POLITICAL PARTIES IN THE WORLD

Top 10 corrupt political party in the world 2017

1. Pakistan Muslim League Nawaz, PML(N)- Pakistan
2. National Resistance Movement - Uganda
3. Progressive Action Party - Cuba
4. INDIAN NATIONAL CONGRESS - INDIA
5. Vietnam's Communist party - Vietnam
6. Kuomintang - China
7. Nationalist Fascist Party - Italy
8. Nazi Party - Germany
9. Communist Party of China- China
10. Communist Party of the Soviet Union - Russia

Source: BBC Survey
Primary Categories

- Images taken out of context (~30%)
- Photoshopping (~20%)
- False statistics and quotes (~10%)
- Other
Misinformation topics

• Primarily political
  • Memes/Nationalism/Religion
• Rumors/Urban legends
• Feel good stuff
• Health (roughly 15%!)
Why is this important?

• Automated fact checking
• Out of context images ✓
• Photoshopped ✓
• Fake quotes/stats ×
Why are they sharing these?

• Caring about others

• Low cost of forwarding (much easier than fact checking)

• Lack of awareness

• Personal belief in these networks

• Malicious intent? (hard to identify)

• Prejudice and ideology rather than out of ignorance or digital illiteracy.
What could be done?

• Only central authorities can make a difference!

• Need to have a common repository
Summary

• Categorized image misinformation

• Public groups

• The same categorization generalizes across countries and platforms
Thank you!

@gvrkiran
garimell@mit.edu