

Hi Doppelgänger: Detecting Manipulation in News Comments

Jan Pennekamp, Martin Henze, Oliver Hohlfeld, Andriy Panchenko

<https://comsys.rwth-aachen.de/>
<https://www.b-tu.de/en/it-security>

San Francisco / CyberSafety, 13th May 2019

- **Plethora of opportunities to express opinions**
 - ▶ Easy access and mostly without real authentication
- **Media increasingly reports about opinion manipulation**
 - ▶ (Paid) trolls, fake news, ...
 - ▶ Topics: Ukraine crisis, Brexit, US election, ...

Which information can we still trust?

Facebook purges tens of thousands of fake accounts

Twitter admits former Russian bots had disclosed its and reported ion of figures

The readers' editor on... pro-Russia trolling below the line on Ukraine stories

Chris Elliott

Global crackdown on 'ina on Friday targeting a sing

Guardian moderators, who deal with 40,000 comments a day, believe there is an orchestrated pro-Kremlin campaign

By Gerrit D January 27,

- **An attacker targeting public opinion**
 - ▶ Controlling multiple pseudonyms on a single news website
 - ▶ Utilizing the bandwagon effect
 - Frequently commenting similar beliefs to generate a false impression
- **Actively spoofing side channel information**
 - ▶ Lack of identifiers for matching accounts

Stylometric details are hard to spoof

- 1. Crawl datasets from news websites**
- 2. Artificially split profiles into multiple identities**
- 3. Extract stylometric features for each identity**
- 4. Apply, evaluate, and adjust Doppelgänger Finder**
- 5. Collect impressions on scenarios without artificially split identities**

Doppelgänger Finder is a state-of-the-art implementation by Afroz et al.



Dataset Generation

- Compile large, realistic, up-to-date datasets of user comments on news (*The Guardian* , *SPIEGEL ONLINE* , *ZEIT ONLINE* *Z*)



Adaptation of Existing Stylometric Approaches

- Extend state-of-the-art feature set to cope with modern challenges (significantly shorter texts, Internet slang, ...)



Doppelgänger Evaluation

- Identify a baseline scenario for user comments (artificial scenario)
- Present a metric to determine the required parameters
- Apply our approach to a real-world evaluation (without artificial identities)

- **The Guardian**

- ▶ Two accounts opposing the Brexit and extremists
- ▶ False positive? Only one account with frequent line breaks

- **SPIEGEL ONLINE**

- ▶ Potentially linked a main account (from 2008, >1.700 comments) and a second account (from 2015, only 80 comments)
 - Both actively commenting on politics

- **ZEIT ONLINE**

- ▶ Linked two accounts publishing messages in English

- **Plenty of “suspicious” texts indexed by search engines**

- **Collection of large-scale datasets**
 - ▶ to detect doppelgänger on
- **Adaptation of Doppelgänger Finder for shorter texts**
- **Applicable in scenarios w/ & w/o artificial doppelgänger**
- **Get first impressions during real-world experiment**
 - ▶ Discovery of (potential) doppelgänger on random news articles
 - ▶ Future work (!), important research topic

Thank you for your attention!