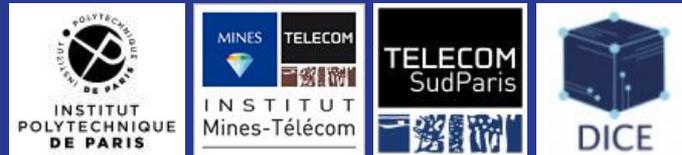


Impersonation in Online Social Networks

Koosha Zarei

*Under the supervision of Noel Crespi and Reza Farahbakhsh
Institut Polytechnique de Paris, Telecom SudParis*



About me!

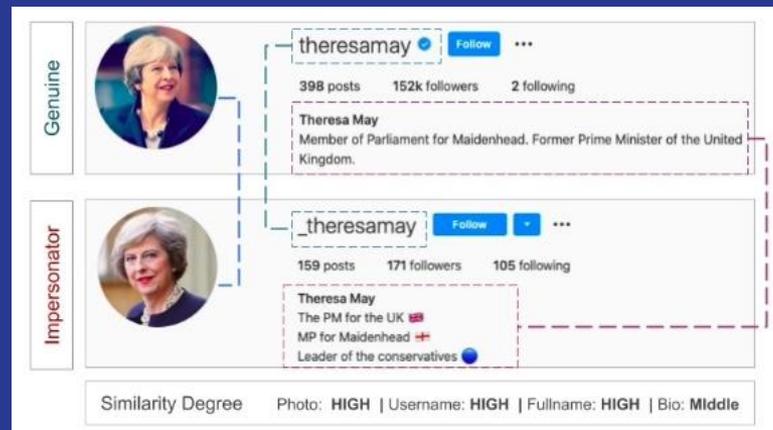


- He is a Research Engineer & PhD Candidate at Institut Polytechnique de Paris, Telecom SudParis, France.
- He is studying **Computer Science** and currently, he is a member of "*Data Intelligence and Communication Engineering*" Lab.
- He is working on **Bot Detection**, **Fake Content**, and **User Behaviour Analysis** on Online Social Networks such Instagram & Twitter.
- His research interests include:
 - Online Social Networks,
 - Data Analysis,
 - Deep Learning,
 - NLP,
 - Fake News,
 - Big Data.

Introduction & Motivation

Impersonation

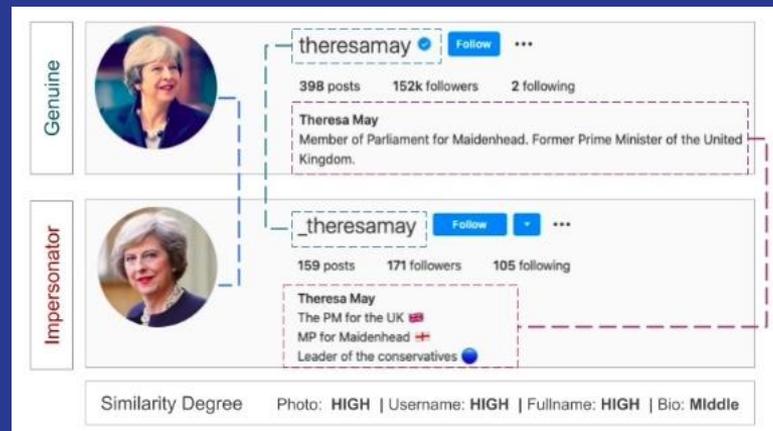
- An impersonator is someone who **pretends** or **copies** the behaviour or actions of another.
- Operate with different
 - Aims
 - Characteristics
 - Behaviours



Introduction & Motivation

Impersonation

- An impersonator is someone who **pretends** or **copies** the behaviour or actions of another.
- Operate with different
 - Aims
 - Characteristics
 - Behaviours
- Play an important role in the **production** and **propagation** of the content on OSNs



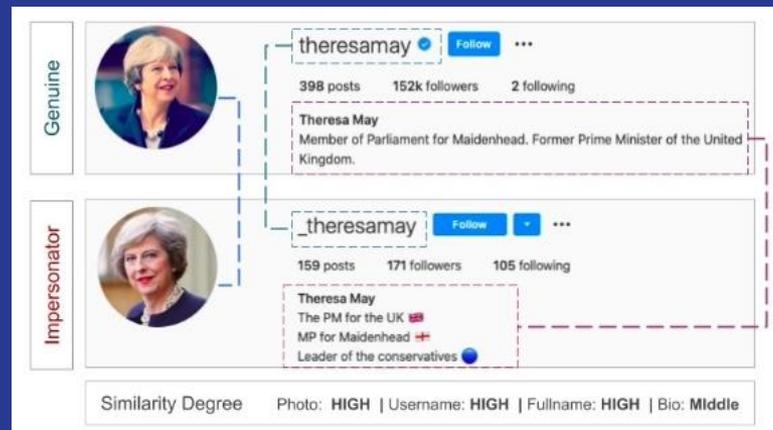
Introduction & Motivation

Impersonation

- An impersonator is someone who **pretends** or **copies** the behaviour or actions of another.
- Operate with different
 - Aims
 - Characteristics
 - Behaviours
- Play an important role in the **production** and **propagation** of the content on OSNs

Lawsuits

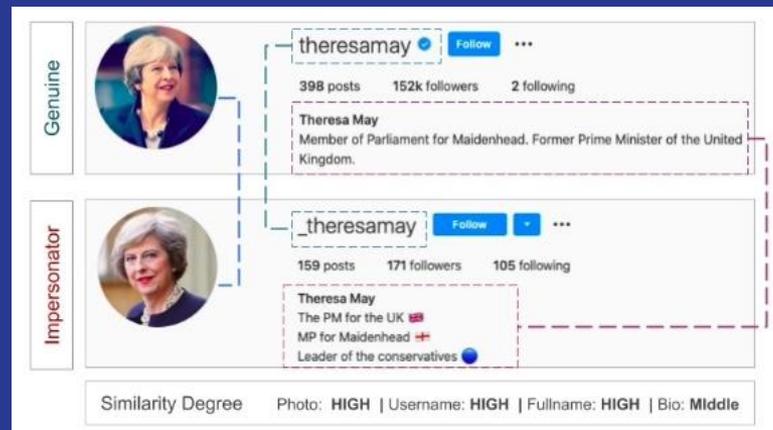
Several lawsuits have taken place in the United State, UK, and EU where criminal impersonation is a **crime!**



Introduction & Motivation

Research Questions

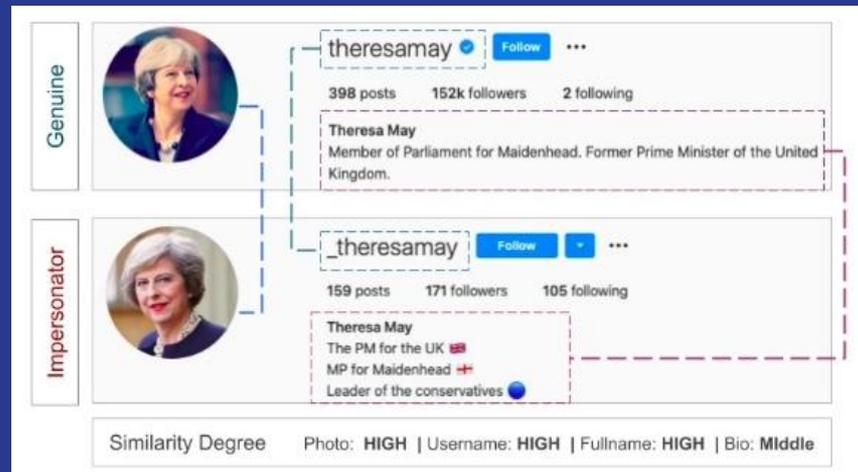
1. Who are the impersonators and what are their behaviours?
2. How we can identify impersonator?
3. Among them, how many distinct hidden groups exist?
4. What are their characteristics?
5. Can we identify impersonator-generated content?



Introduction & Motivation

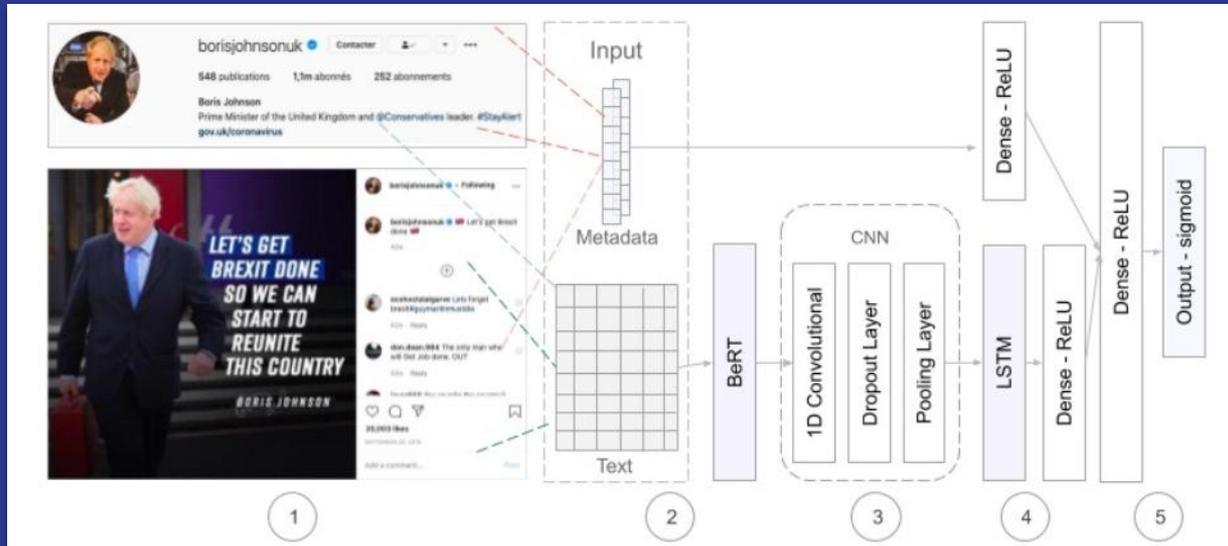
Main Contributions

- Develop an exclusive **crawler** to collect necessary data from Instagram aligned with GDPR rules.
- A novel **dataset** containing the content and activities of impersonators in various communities.
- Focus on the **detection, Clustering, and Identifying** of impersonators by developing a **Deep Neural Network** architecture.
- Leverage advanced **NLP** techniques to analyse the **behaviour** of identified fake identities.



A Deep Neural Approach

- Impersonator-generated vs. Genuine-generated content.
- Predict content type by a **DNN model** ('bot-generated', 'Fan-generated', 'genuine')
- **Bot**, as a fake identity, produces **untrustworthy content** and **fake engagements**.
- CNN, LSTM, BERT and Dense Layers.
- impersonators target some **specific topics**.



A Deep Neural Approach

- Impersonator-generated vs. Genuine-generated content.
- Predict content type by a **DNN model** ('bot-generated', 'Fan-generated', 'genuine')
- **Bot**, as a fake identity, produces **untrustworthy content** and **fake engagements**.
- CNN, LSTM, BERT and Dense Layers.
- impersonators target some **specific topics**.

Performance of the proposed architecture

Model	Accuracy	Precision	Recall	F1
Random Forest Classifier	0.76	0.78	0.77	0.76
Proposed DNN (post)	0.78	0.79	0.76	0.78
Proposed DNN (post + profile)	0.83	0.82	0.83	0.82
Proposed DNN (post + profile) + BERT	0.86	0.85	0.86	0.85

Feature Set used in Deep Neural Network.

Post Features		Publisher Features	
Feature	Type	Feature	Type
caption text	text	similarity username	numeric
caption topics (LDA)	text	similarity fullname	numeric
post hashtag	text	similarity bio	numeric
tagged users in post	text	profile biography	text
like count	numeric	similarity photo	numeric
comment count	numeric	follower/followee/post	numeric
tagged users count	numeric	full name	text
mention users count	numeric	biography	text
hashtag count	numeric	username	text
overall sentiment of caption	numeric	following followers ratio [29]	numeric
overall sentiment of hashtag	numeric	followers posts ratio	numeric
media type (image or video)	numeric	bio emoji count	numeric
emoji count	numeric	bio hashtag count	numeric
url/website exist	numeric		numeric
date	numeric		

Thank you!

Koosha.zarei @ telecom-sudparis.eu



ANNEX - Future Direction

- An **Hybrid Bot Detection** in OSNs to detect and identify fake identities and fake content.
- Using **BERT** and **Transfer Learning** to increase the overall accuracy.
- Considering more data types such as comments, stories, Reels and the connection between impersonators.

ANNEX - Publications

Conference

1. Koosha Zarei, D. Ibosiola, R. Farahbakhsh, Z. Gilani, K. Garimella, N. Crespi, G. Tyson. Characterising and Detecting Sponsored Influencer Posts on Instagram. In 2020 ACM/IEEE ASONAM, 2020.
2. Koosha. Zarei, R. Farahbakhsh, N. Crespi and G. Tyson. Impersonation on Social Media: A Deep Neural Approach to Identify Ingenuine Content. In 2020 ACM/IEEE ASONAM, 2020.
3. Koosha. Zarei, R. Farahbakhsh, and N. Crespi. How impersonators exploit instagram to generate fake engagement? In ICC 2020, pages 1–6, 2020.
4. Koosha. Zarei, R. Farahbakhsh, and N. Crespi. Typification of impersonated accounts on instagram. In 2019 IEEE 38th IPCCC, pages 1–6, 2019.
5. Koosha. Zarei, R. Farahbakhsh, and N. Crespi. Deep dive on politician impersonating accounts in social media. In 2019 ISCC, pages 1–6, 2019.

Dataset

1. Koosha Zarei, R. Farahbakhsh, N. Crespi, G. Tyson. A First Instagram Dataset on COVID-19. arXiv, 2020.

Ongoing

1. Koosha Zarei, R. Farahbakhsh, N. Crespi, G. Tyson. A Dataset and Exploratory Study of Coronavirus on Instagram. Journal Paper.
2. Koosha Zarei, R. Farahbakhsh, N. Crespi, G. Tyson. Characterizing Impersonators on Online Social Networks. Journal Paper.
3. Koosha Zarei, R. Farahbakhsh, N. Crespi, G. Tyson. Deep Hybrid Bot Identification: Exploring Generated Content During Coronavirus. Conference Paper.