ISISTAN
Instituto Superior de Ingeniería de Software Tandil

CONICET

Friend or Foe: Studying user trustworthiness for friend recommendation in the era of misinformation

Dr Antonela Tommasel

UNIVERSIDAD NACIONAL DEL CENTRO DE LA PROVINCIA DE BUENOS AIRES
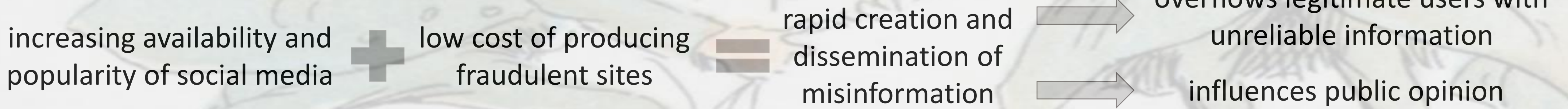
Fake News

"Fake news is made-up stuff, masterfully manipulated to look like credible journalistic reports that are easily spread online to large audiences willing to believe the fictions and spread the word"

Always existed!!

Social media aggravates the problem

increasing availability and popularity of social media **+** low cost of producing fraudulent sites **=** rapid creation and dissemination of misinformation ⟶ overflows legitimate users with unreliable information

⟶ influences public opinion

The value and quality of the social Web diminishes!!

And what about recommendation?

Recommending <u>similar</u> elements ⟶ **NOT** sufficient to ensure reliability! ⟶ We need **Relevant** AND **Trustworthy** recommendations!

⬆ Measuring trustworthiness has been important in psychology and social sciences.
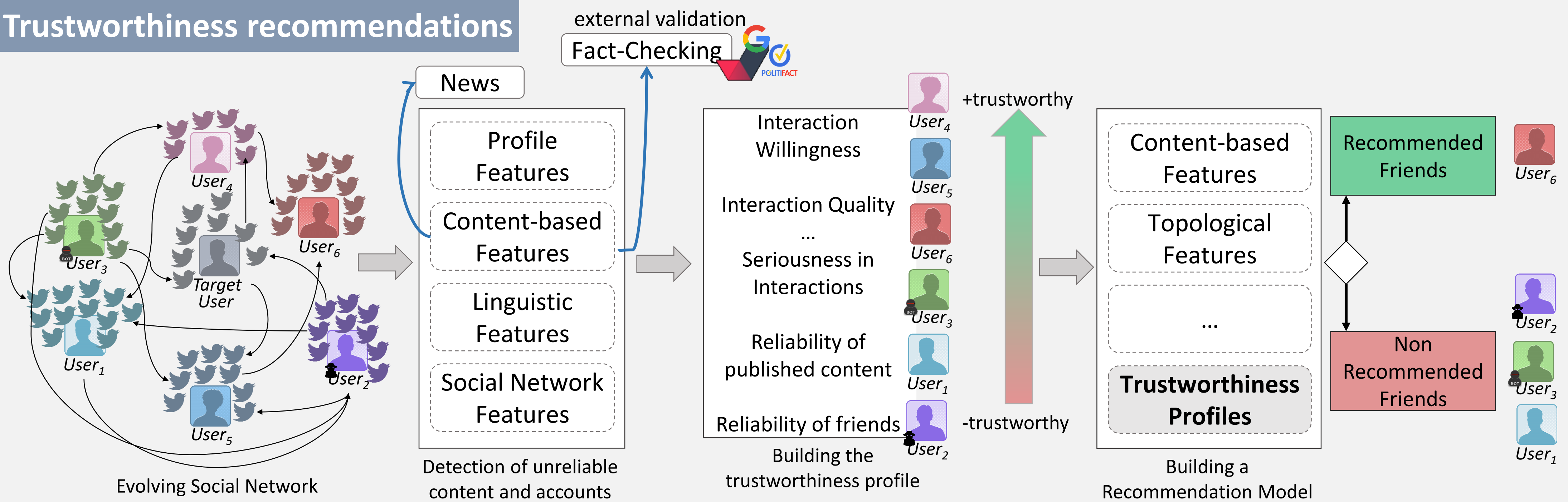
⬇ Few studies have incorporated this concept to:
• Social media
• The proliferation of unwanted content
• Recommendations

New challenges!
• How to determine the trustworthiness of friends, content, locations in social media …
• How such trustworthiness can be integrated into recommender systems.

Trustworthiness recommendations

external validation
Fact-Checking

Goal #1
Analyse multiple information sources to **detect unreliable content**, false accounts, bots, spammers…

Goal #2
Define a **user trustworthiness** profile based on <u>user behaviour</u> and the <u>detection of unreliable content</u> and <u>accounts</u>.

Goal #3
Integrate the **profile** in <u>recommendation systems</u>!

What we have already done

• Explored the dynamics of social networks.
• Studied the importance of personality and user behaviour in user recommendation.
• Proposed a recommendation technique that adapted the recommendation criteria to the previously selected followees.
• Defined a metric of user influence.
• Exploited the linked nature of social media for community detection.
• Applied community detection for personalised recommendation.
• Explored the detection of aggressive content and aggressors in the context of cyberbullying.
• Studied writing styles in relation to personality and gender.

There is still work to do!

We are currently working on <u>Goals #1-2</u>

• Defining rumour detection.
• Defining spam detection.
• Refining aggressive and hate speech detection.
• Extending unreliable content detection for unreliable user detection.
• Integrating everything in the profile!
• More Evaluations!!!