Friends or Foe: Recommending friends in the misinformation era

Dr Antonela Tommasel

ISISTAN, CONICET-UNICEN, Argentina

antonela.tommasel@isistan.unicen.edu.ar

CONICET

I S I S T A N

Who am I?

- Dr Antonela Tommasel
  - PhD in Computer Sciences at UNICEN
- Work at ISISTAN, CONICET-UNICEN.
- Teacher Assistant at UNICEN.
- Research Interests:
  - Recommender systems
  - Text Mining
  - Social Media
  - Social Computing
  - …

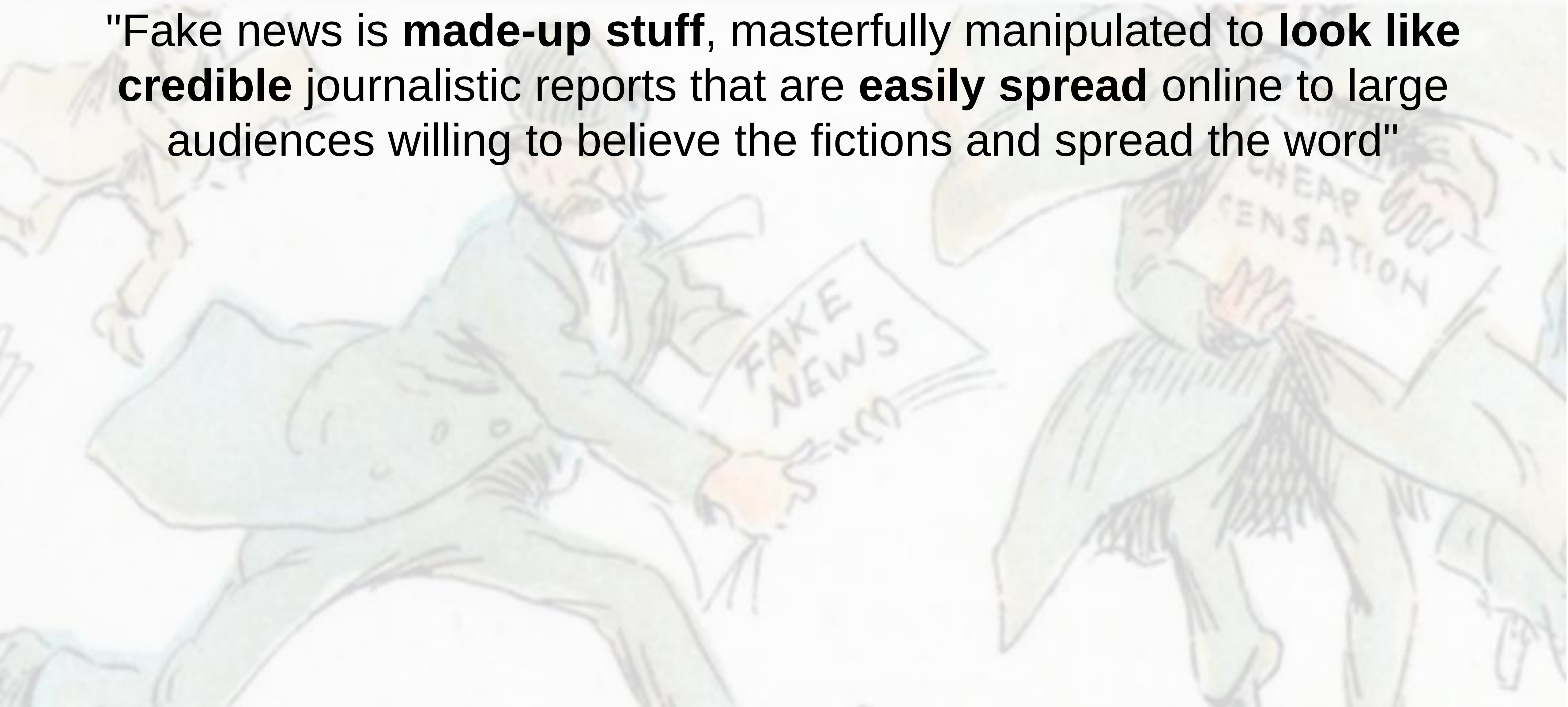
CONICET

I S I S T A N

Table of contents

Faking it!

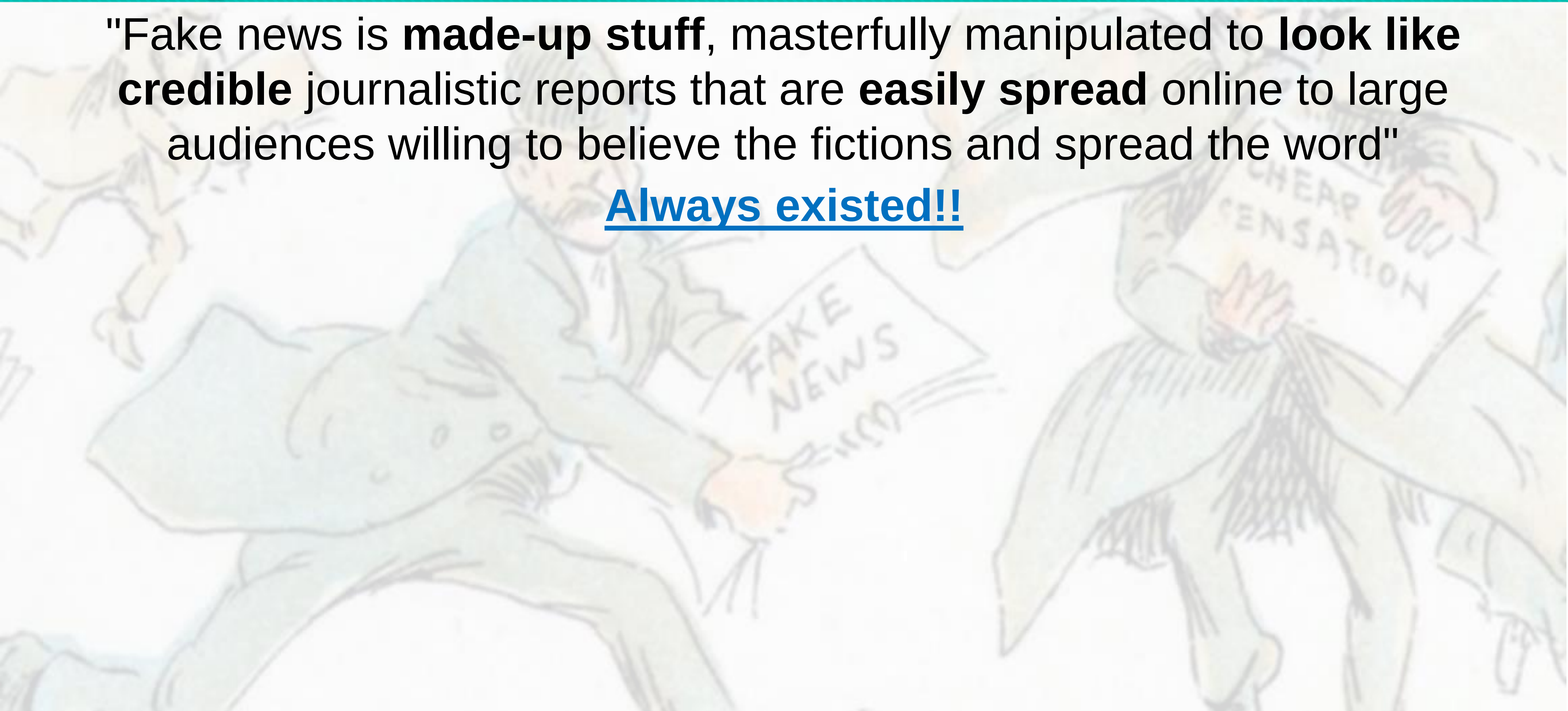
"Fake news is **made-up stuff**, masterfully manipulated to **look like credible** journalistic reports that are **easily spread** online to large audiences willing to believe the fictions and spread the word"



Faking it!

"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!!



Faking it!

"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"

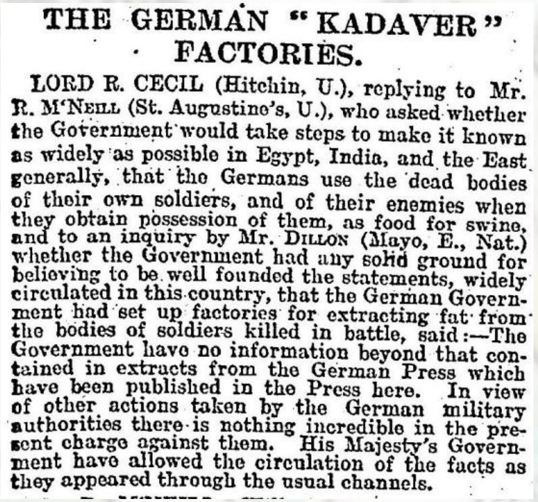
Always existed!!

The King's Health is Failing
(mid 1700s – Jacobite rebellion)

Life on the Moon
(1835)

Jack the Ripper
(1888)

World War One Fake News
(1917)



THE GERMAN "KADAVER" FACTORIES.

LORD R. CECIL (Hitchin, U.), replying to Mr. R. M'NEILL (St. Augustine's, U.), who asked whether the Government would take steps to make it known as widely as possible in Egypt, India, and the East generally, that the Germans use the dead bodies of their own soldiers, and of their enemies when they obtain possession of them, as food for swine, and to an inquiry by Mr. DILLON (Mayo, E., Nat.) whether the Government had any solid ground for believing to be well founded the statements, widely circulated in this country, that the German Government had set up factories for extracting fat from the bodies of soldiers killed in battle, said:—The Government have no information beyond that contained in extracts from the German Press which have been published in the Press here. In view of other actions taken by the German military authorities there is nothing incredible in the present charge against them. His Majesty's Government have allowed the circulation of the facts as they appeared through the usual channels.

Faking it!

"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!!

The King's Health is Failing
(mid 1700s – Jacobite rebellion)

Life on the Moon
(1835)

Jack the Ripper
(1888)

World War One Fake News
(1917)

THE GERMAN "KADAVER" FACTORIES.

LORD R. CECIL (Hitchin, U.), replying to Mr. R. M'NEILL (St. Augustine's, U.), who asked whether the Government would take steps to make it known as widely as possible in Egypt, India, and the East generally, that the Germans use the dead bodies of their own soldiers, and of their enemies when they obtain possession of them, as food for swine, and to an inquiry by Mr. DILLON (Mayo, E., Nat.) whether the Government had any solid ground for believing to be well founded the statements, widely circulated in this country, that the German Government had set up factories for extracting fat from the bodies of soldiers killed in battle, said:—The Government have no information beyond that contained in extracts from the German Press which have been published in the Press here. In view of other actions taken by the German military authorities there is nothing incredible in the present charge against them. His Majesty's Government have allowed the circulation of the facts as they appeared through the usual channels.

Social media aggravates the problem!

Faking it!

- Social media represents the ideal environment for undesirable phenomena!
  - The dissemination of unwanted or unreliable content, and misinformation.

- A threat to the access to reliable and **trustworthy** information and the establishment of **reliable** social relations.

Social media provides a great opportunity to learn about events and news.

Social media produces scepticism amongst users as relevant and accurate information coexist with unreliable and undesired information.

- The growing spread of undesired content motivated the assessment of the reliability of information.
- The vulnerability of individuals and society to the manipulations is still unknown.

Faking it!

increasing availability and popularity of social media

+

low cost of producing fraudulent sites

=

rapid creation and dissemination of misinformation

Faking it!

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

Faking it!

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!!

What has been done?

- The development of methods for automatically detecting undesired content is **essential**.
  - Such detection is not simple.
  - Mainly based on **one** of three aspects:
    - Textual content.
    - The responses received.
    - The identification of the content promoters.

What has been done?

- The development of methods for automatically detecting undesired content is **essential.**
  - Such detection is not simple.
  - Mainly based on **one** of three aspects:
    - Textual content.
    - The responses received.
    - The identification of the content promoters.

Each type of unwanted content may have different textual indicators.

Responses focus on content propagation requiring access to large amounts of data.

- Based on the same characteristics as the detection of unreliable content.
- Mostly, techniques attempt to determine only if an account is a certain type of unwanted user (binary classification).

Challenges

1. Identification requires more than text analysis, hence multiple sources of information must be integrated.

2. Bots and accounts spreading misinformation modify their behaviour patterns in an attempt to go unnoticed.

3. Techniques may be over-trained for a specific type of misinformation or spam campaign, limiting its applicability in broader domains.

4. As undesired content does not appear spontaneously, it is vital to analyse who published it, its intentions and processes.

Faking it!

Challenges

It can be challenging given the incompleteness, noise and volume of it.

1. Identification requires more than text analysis, hence multiple sources of information must be integrated.

2. Bots and accounts spreading misinformation modify their behaviour patterns in an attempt to go unnoticed.

3. Techniques may be over-trained for a specific type of misinformation or spam campaign, limiting its applicability in broader domains.

4. As undesired content does not appear spontaneously, it is vital to analyse who published it, its intentions and processes.

Faking it!

Challenges

It can be challenging given the incompleteness, noise and volume of it.

1. Identification requires more than text analysis, hence multiple sources of information must be integrated.

Detection will be less effective if training data is not periodically updated.

2. Bots and accounts spreading misinformation modify their behaviour patterns in an attempt to go unnoticed.

3. Techniques may be over-trained for a specific type of misinformation or spam campaign, limiting its applicability in broader domains.

4. As undesired content does not appear spontaneously, it is vital to analyse who published it, its intentions and processes.

Faking it!

Challenges

1. Identification requires more than text analysis, hence multiple sources of information must be integrated.

It can be challenging given the incompleteness, noise and volume of it.

2. Bots and accounts spreading misinformation modify their behaviour patterns in an attempt to go unnoticed.

Detection will be less effective if training data is not periodically updated.

3. Techniques may be over-trained for a specific type of misinformation or spam campaign, limiting its applicability in broader domains.

Each type of misinformation presents particular characteristics, which must be taken into account for detection.

4. As undesired content does not appear spontaneously, it is vital to analyse who published it, its intentions and processes.

Faking it!

Challenges

1. Identification requires more than text analysis, hence multiple sources of information must be integrated.

It can be challenging given the incompleteness, noise and volume of it.

2. Bots and accounts spreading misinformation modify their behaviour patterns in an attempt to go unnoticed.

Detection will be less effective if training data is not periodically updated.

3. Techniques may be over-trained for a specific type of misinformation or spam campaign, limiting its applicability in broader domains.

Each type of misinformation presents particular characteristics, which must be taken into account for detection.

4. As undesired content does not appear spontaneously, it is vital to analyse who published it, its intentions and processes.

Affected by the lack of integration of multiple sources of information, the updating of techniques and the disregard of the interrelation between different social platforms.

And what about recommendations?

Recommending similar elements
is **NOT** sufficient to ensure reliability!

And what about recommendations?

Recommending <u>similar</u> elements
is **<u>NOT</u>** sufficient to ensure reliability!

We need **<u>Relevant</u>** **and** **<u>Trustworthy</u>** recommendations!

And what about recommendations?

Recommending <u>similar</u> elements
is **<u>NOT</u>** sufficient to ensure reliability!

We need **<u>Relevant</u> and <u>Trustworthy</u>** recommendations!

New challenges!

User trustworthiness

Measuring trustworthiness has been important in psychology and social sciences.

User trustworthiness

Measuring trustworthiness has been important in psychology and social sciences.

Various factors to consider:

- Personal relationships.
- Past experiences of a user with their friends.
- Actions and opinions made in the past.
- …

In social media focuses on behaviours expressed in the way information is produced and shared.

No attention to the principle of unequal participation. Largest proportion of content is created by the minority of users, whilst the other are lurkers.

User trustworthiness

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

User trustworthiness

Measuring trustworthiness has been important in psychology and social sciences.

- Generally studied in the context of collaborative filtering to:
  - Determine the reliability of users' ratings.
  - Mitigate the cold start problem.

Few studies have incorporated this concept to:

- Social media
- The proliferation of unwanted content recommendations

- Aspects specifically related to unwanted content are not considered.
- Require explicit reputation Indicators.
- Do not consider the dynamism of the social environments.

Table of contents

What do we want?

The goal is to **define a <u>profile</u>** to

What do we want?

The goal is to **define a <u>profile</u>** to

describe and estimate the trustworthiness or reputation **of users**,

What do we want?

The goal is to **define a <u>profile</u>** to

describe and estimate the trustworthiness or reputation **of users**,

to **avoid favouring the propagation of unreliable content** and polluting users

What do we want?

The goal is to **define a profile** to

describe and estimate the trustworthiness or reputation **of users**,

to **avoid favouring the propagation of unreliable content** and polluting users

to be **integrated into a recommender system**,

What do we want?

The goal is to **define a <u>profile</u>** to

describe and estimate the trustworthiness or reputation **of users**,

to **avoid favouring the propagation of unreliable content** and polluting users

to be **integrated into a recommender system**,

aiming at **balancing** both the **relevance** and **reliability** of recommendations.

What do we want?

Research Questions

RQ1. What is the **utility** of the multiple **sources of heterogenous information** available for the detection of unreliable content and malicious accounts.

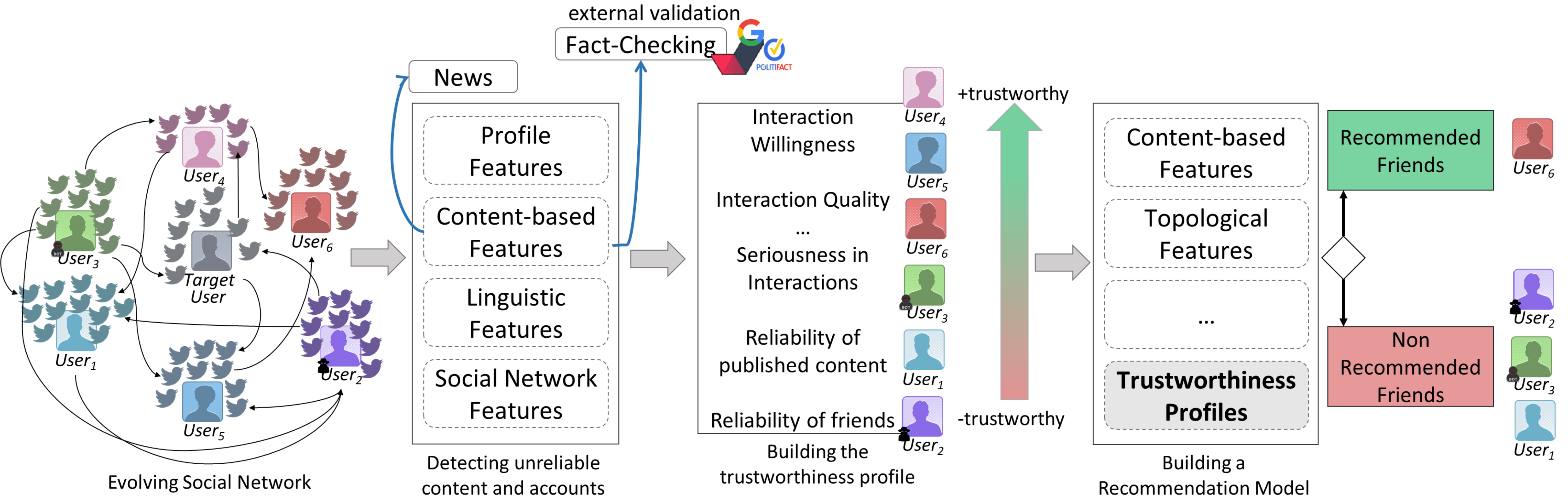
RQ2. How the detection of unreliable content and accounts can be **integrated** for the definition of a **trustworthiness user profile** in relation to their social and publication patterns?

RQ3. How to **adapt** said level of **trustworthiness** to **changes** in the undesired **behaviours** of said users or accounts?

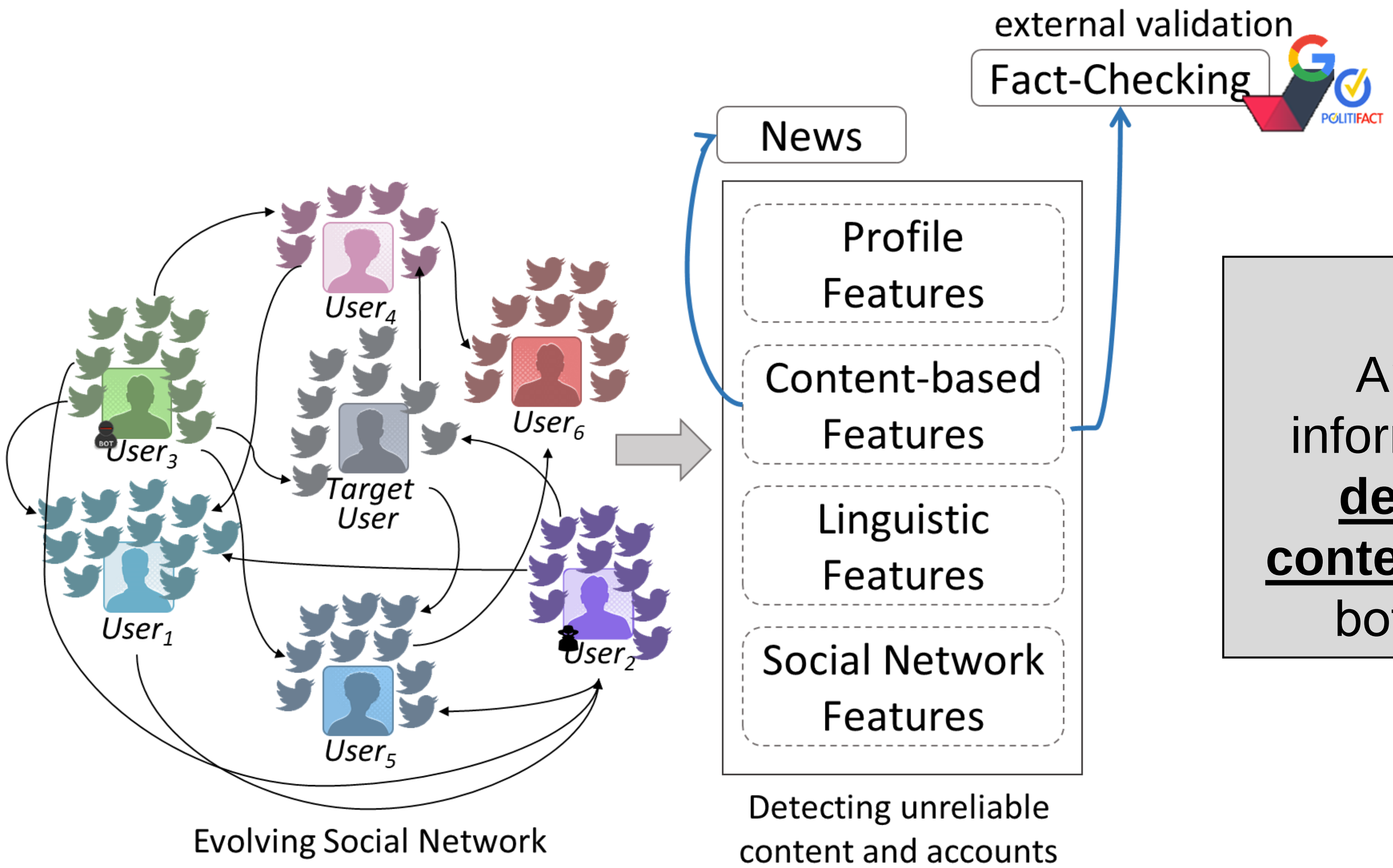
RQ4. How to **integrate** the **trustworthiness profile** in a **recommendation system** that leverages on the characteristics and behaviour of users for personalising the recommendations.

RQ5. To what extent the **quality of recommendations** can be **improved** if the **evolution** of the **interests** and behavioural patterns of users **is also considered**?

Trustworthy recommendations



Trustworthy recommendations



Goal #1

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

Undesired content & users detection

- Even though some computational solutions have been presented, the lack of common ground and public datasets has become one of the major barriers.

- Not only datasets are **rare**, but also they are **mostly limited** to only the actual shared **text**.

Undesired content & users detection

- Even though some computational solutions have been presented, the lack of common ground and public datasets has become one of the major barriers.

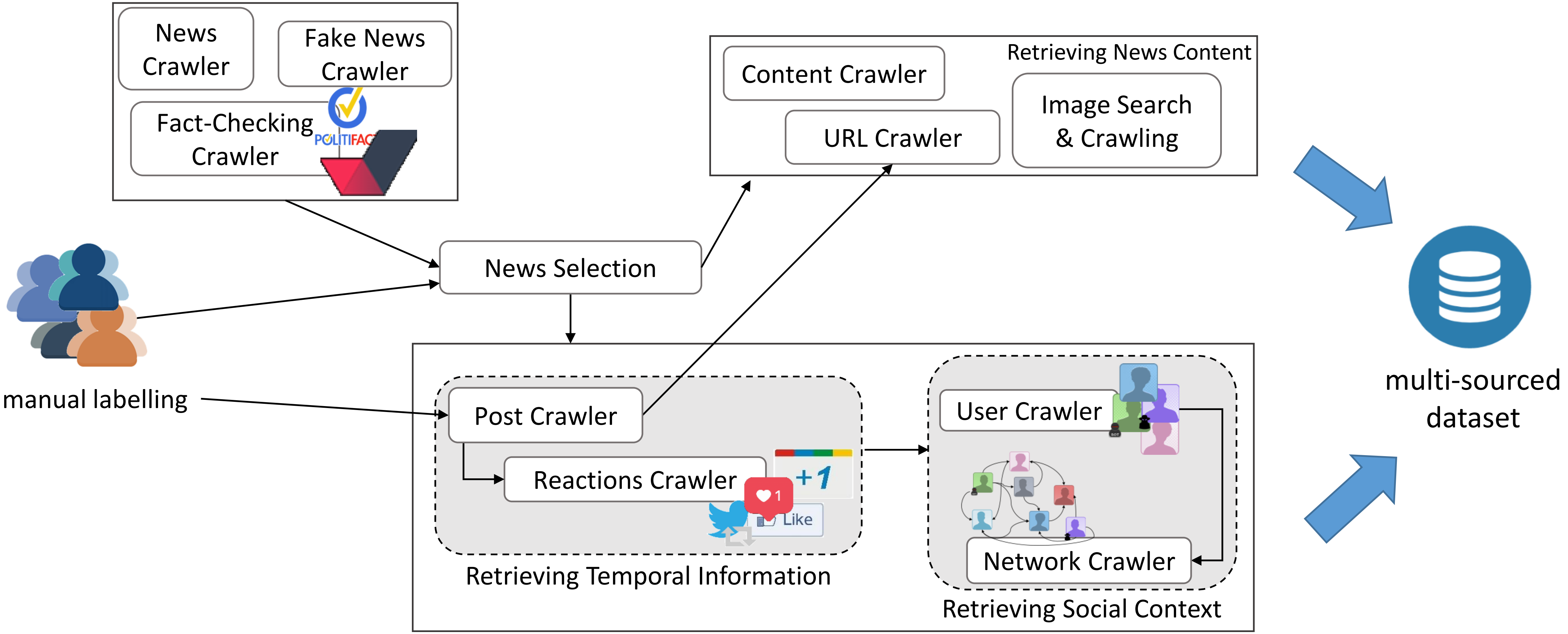
- Not only datasets are **rare**, but also they are **mostly limited** to only the actual shared **text**.

- **Create a publicly available dataset!**
  - Comprising multi-sourced data including: → Distributed as a tool
    - Textual and multimedia content.
    - Social Context.
    - Temporal information.

- Potential uses:
  - Undesired content and user detection.
  - Evolution and engagement cycle.
  - Debunking process

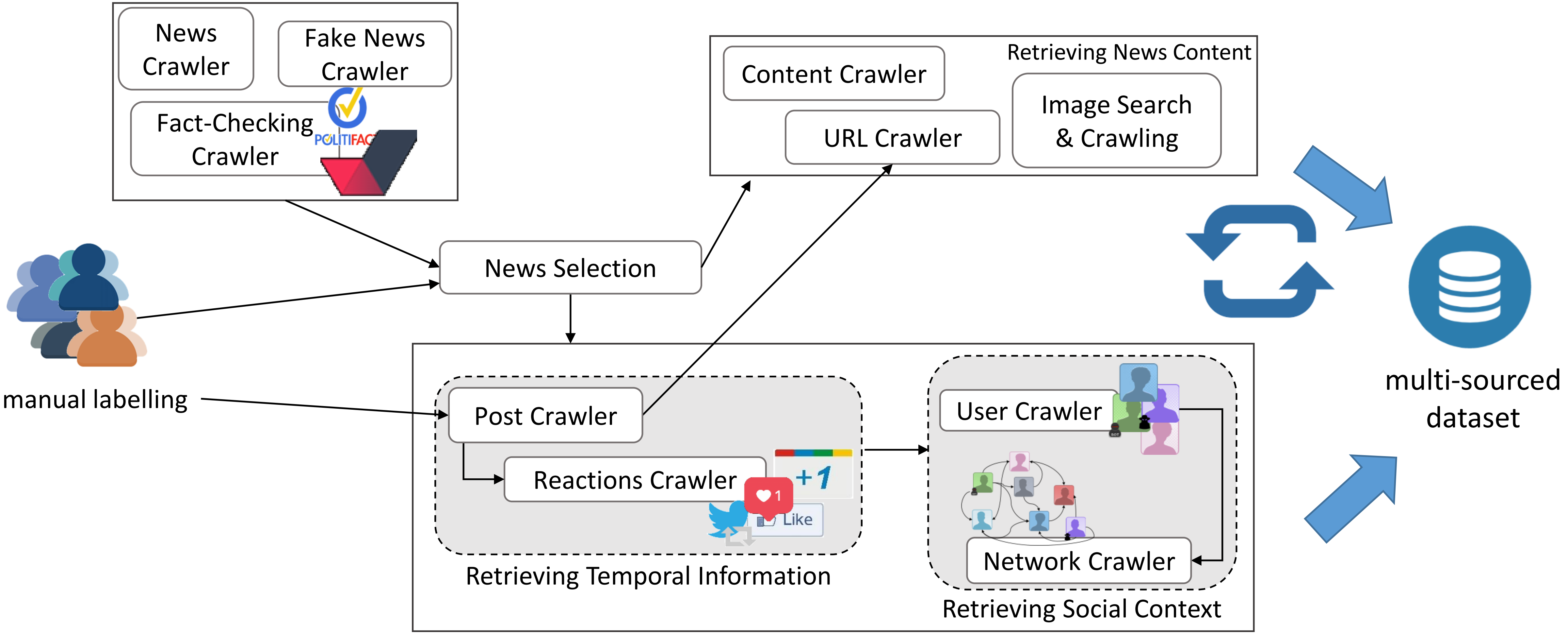
Trustworthy recommendations

Undesired content & users detection

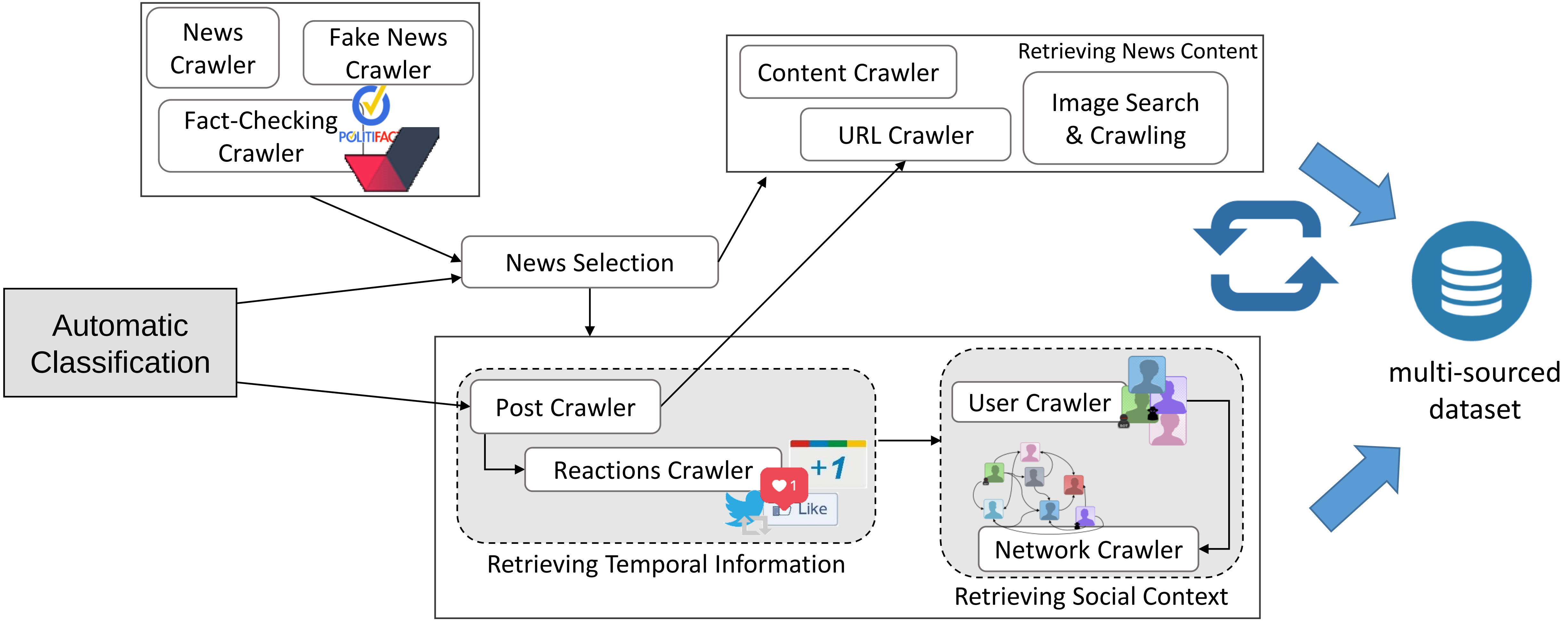


Trustworthy recommendations

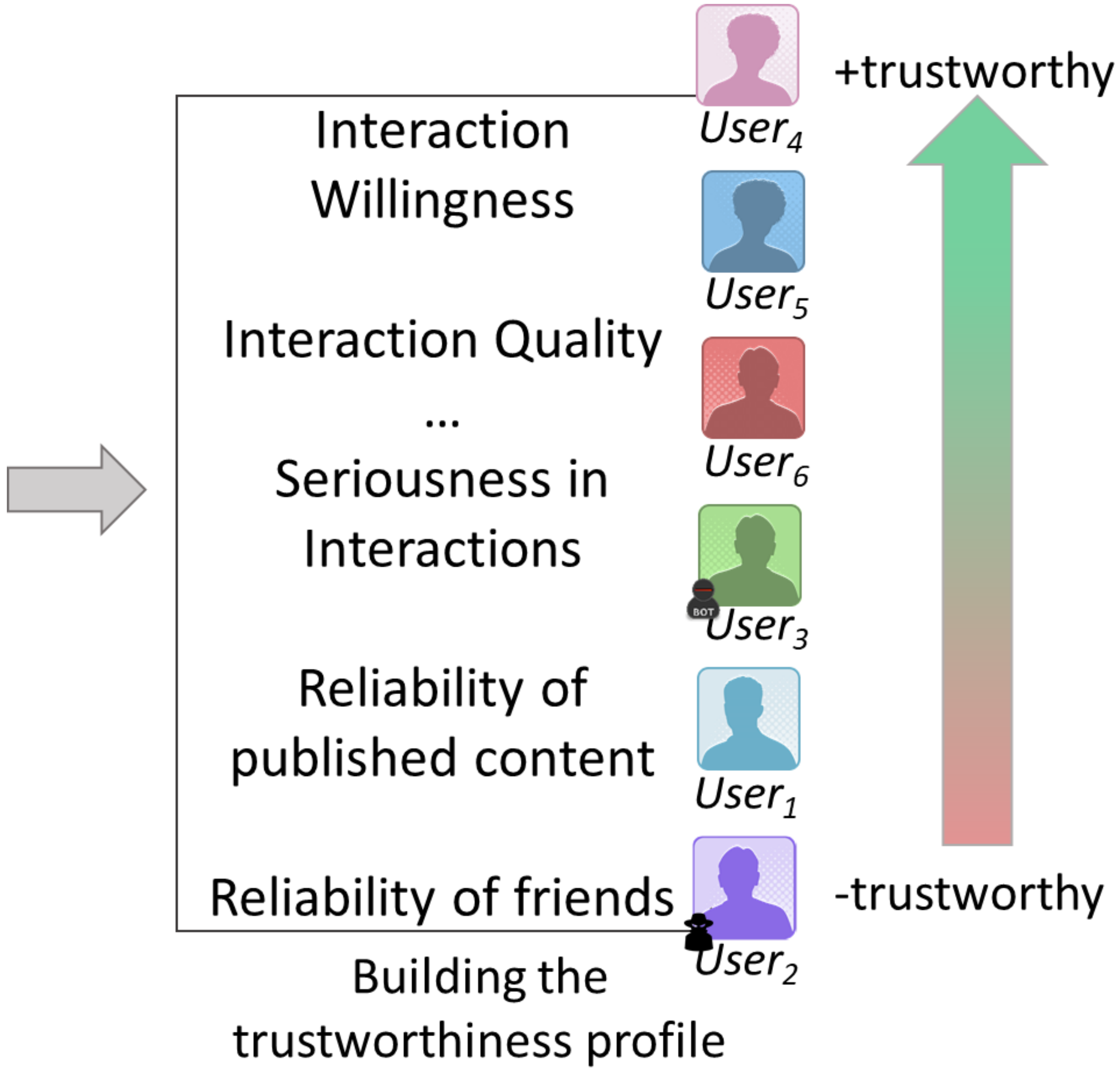
Undesired content & users detection



Undesired content & users detection



Trustworthy recommendations



Goal #2

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

Behaviour conceptual model

- Most studies focus on examining users' motivations and attitudes towards adopting a particular social media network, instead of investigating the processes of information (and misinformation) diffusion.

- We aim at proposing a <u>hierarchical conceptual model</u> to characterise different aspects of the <u>information diffusion process</u>, focusing not only in the <u>information being disseminated</u>, but also on the <u>role of users</u> in such process.

- Shed some light on the psychological and social motivations, and <u>attitudes towards the diffusion and consumption</u> of content in social media.

Behaviour conceptual model

- The layers aim at presenting different and complementary characterisations of the information diffusion process.

Behaviour conceptual model

- The layers aim at presenting different and complementary characterisations of the information diffusion process.
- ***Message Layer*.**
- ***Conversational Thread Layer*.**
- ***User Layer*.**
- ***Group Layer*.**

.

Trustworthy recommendations

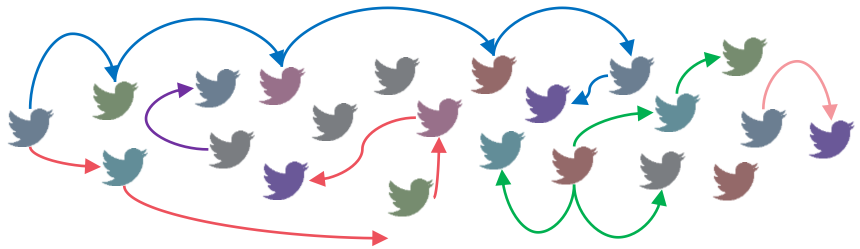
Behaviour conceptual model

- The layers aim at presenting different and complementary characterisations of the information diffusion process.
- ***Message Layer*****.** The top layer of the hierarchy focuses on the characteristics of the individual published messages.
- ***Conversational Thread Layer*****.**
- ***User Layer*****.**
- ***Group Layer*****.**

.

Behaviour conceptual model

- The layers aim at presenting different and complementary characterisations of the information diffusion process.
- ***Message Layer*. The top layer of the hierarchy focuses on the characteristics of the individual published messages.**

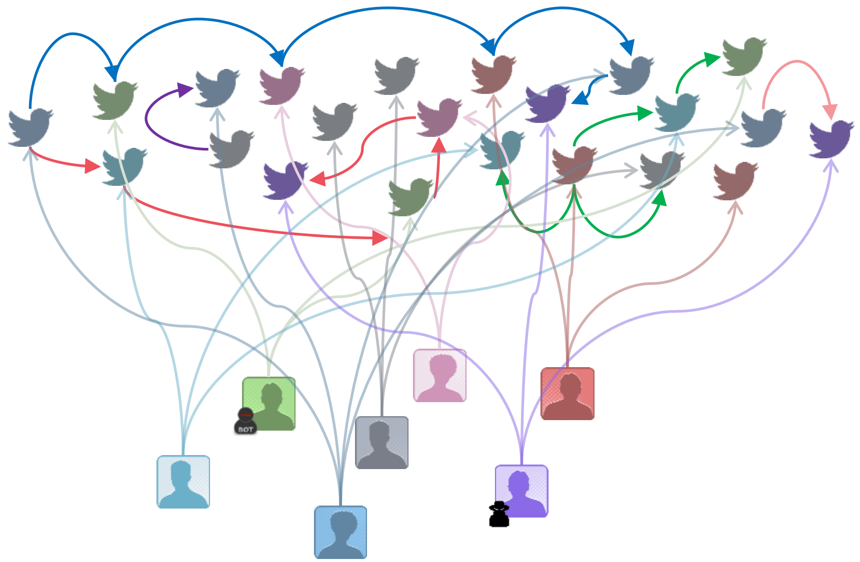


- ***Conversational Thread Layer*.** Messages generally belong to defined conversational threads in which users interact, and engage with the shared content by replying to it.
- ***User Layer***
- ***Group Layer*.**

.

Behaviour conceptual model

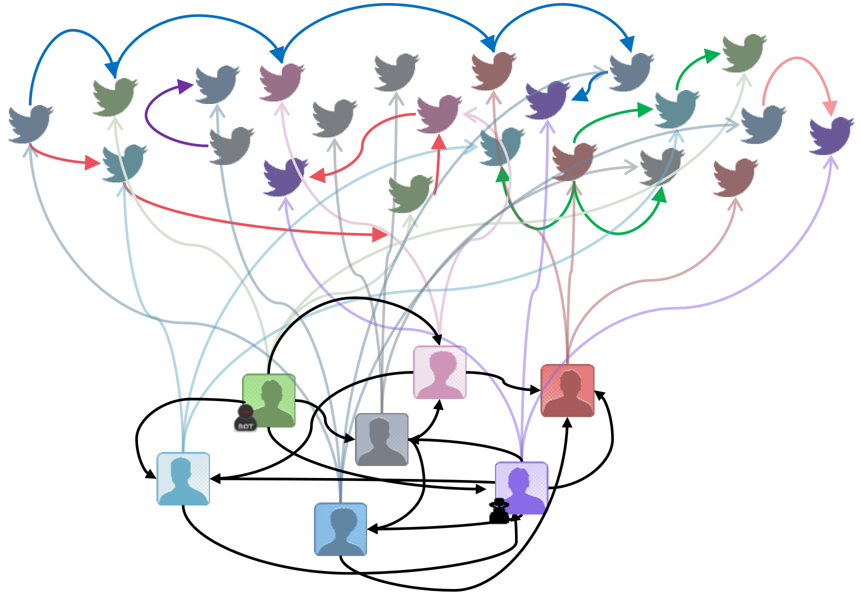
- The layers aim at presenting different and complementary characterisations of the information diffusion process.
- ***Message Layer*. The top layer of the hierarchy focuses on the characteristics of the individual published messages.
- ***Conversational Thread Layer*. Messages generally belong to defined conversational threads in which users interact, and engage with the shared content by replying to it.
- ***User Layer*. In contrast with the conversational thread layer, this layer focuses on the messages and interactions of one individual user.
- ***Group Layer*.



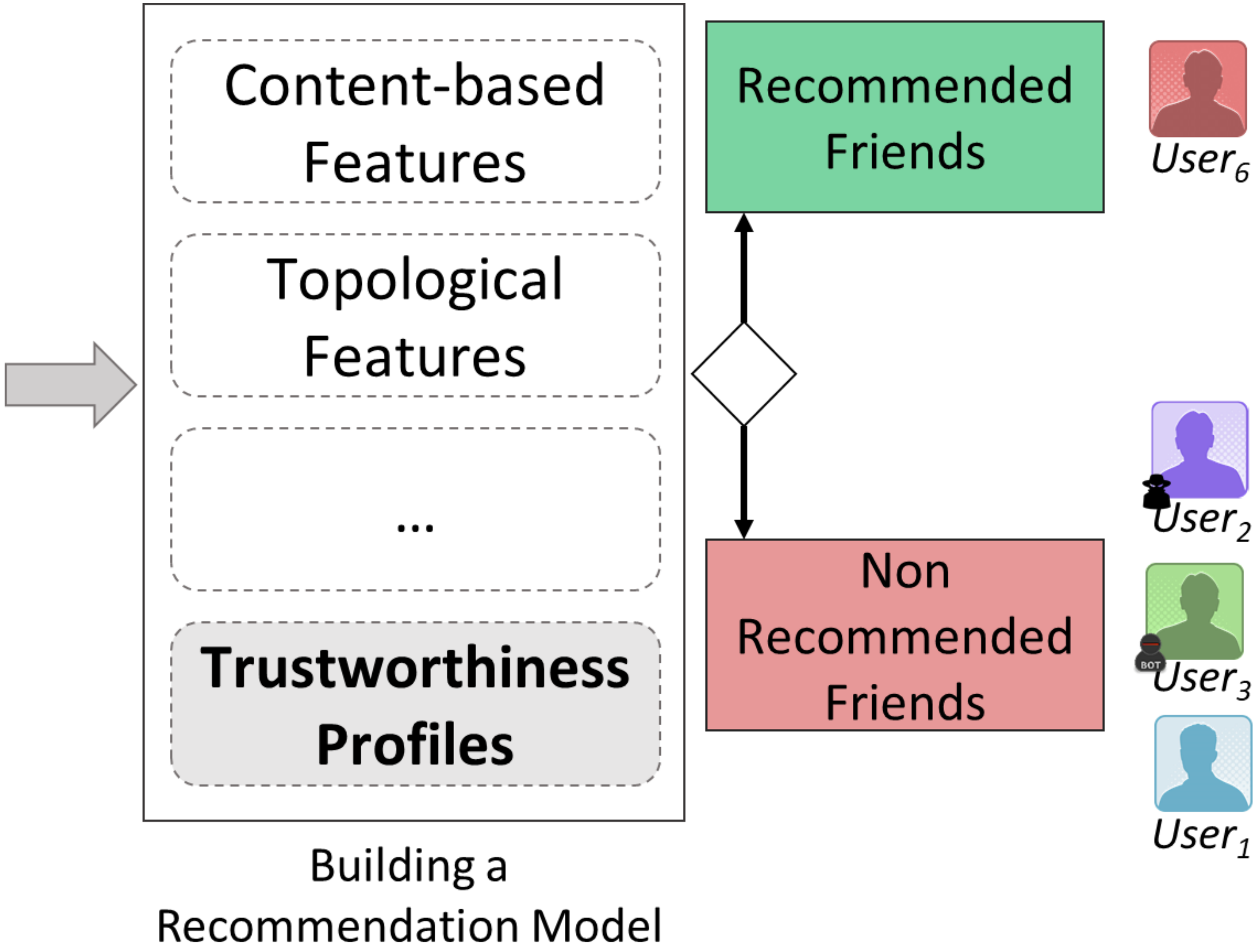
.

Behaviour conceptual model

- The layers aim at presenting different and complementary characterisations of the information diffusion process.
- ***Message Layer*.** The top layer of the hierarchy focuses on the characteristics of the individual published messages.
- ***Conversational Thread Layer*.** Messages generally belong to defined conversational threads in which users interact, and engage with the shared content by replying to it.
- ***User Layer*.** In contrast with the conversational thread layer, this layer focuses on the messages and interactions of one individual user.
- ***Group Layer*.** Users are not isolated nor act individually. This layer focuses on the behaviour of users in relation to others in terms of their activity and interaction .



Trustworthy recommendations



Goal #3

Integrate the **profile** in a recommendation system!

Table of contents

Trustworthy recommendations

What have we already done?

- Explored the dynamics of social networks in terms of homophily.

- Studied the importance of personality and user behaviour in user recommendation.

- Proposed a recommendation technique that adapted over time the recommendation criteria to the characteristics of previously selected friends.

- Defined a metric of user influence.

What have we already done?

- Exploited the linked nature of social media for community detection.

- Applied community detection for discovering groups of friends and provide recommendations tailored to the characteristics of each group.

- Studied writing styles in relation to personality and gender.

- Explored the detection of aggressive content and aggressors in the context of cyberbullying.

Trustworthy recommendations

We are far from finished!

- 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!!!

Trustworthy recommendations

We are far from finished!

- Defining rumour detection.
- Defining spam detection.

Integrate the techniques for multi-class detection!

- Refining aggressive and hate speech detection.

- Extending unreliable content detection
for unreliable user detection.

- Integrating everything in the profile!

- <u>More</u> Evaluations!!!

Trustworthy recommendations

We are far from finished!

- Defining rumour detection.
- Defining spam detection.

Integrate the techniques for multi-class detection!

- Refining aggressive and hate speech detection.

- Extending unreliable content detection for unreliable user detection.

Aiming for a multi-class classification!

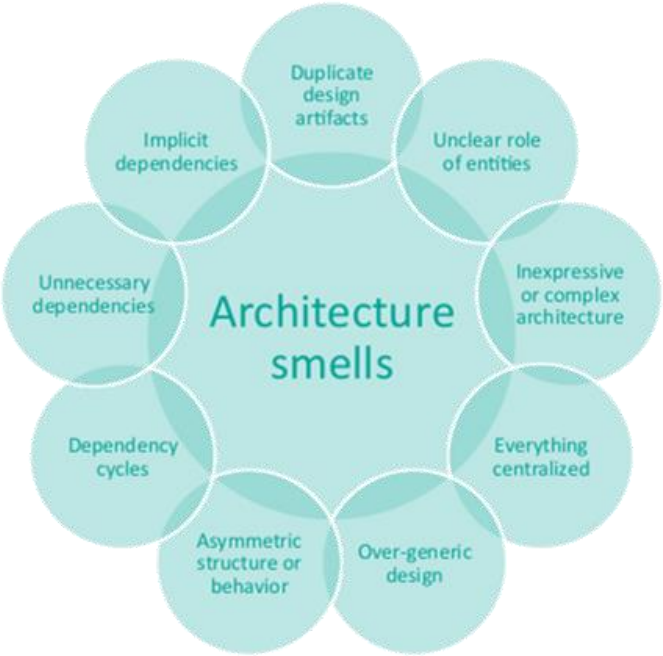
- Integrating everything in the profile!

- More Evaluations!!!

Bonus track

Bonus track

Keeping one-step ahead of architectural smells

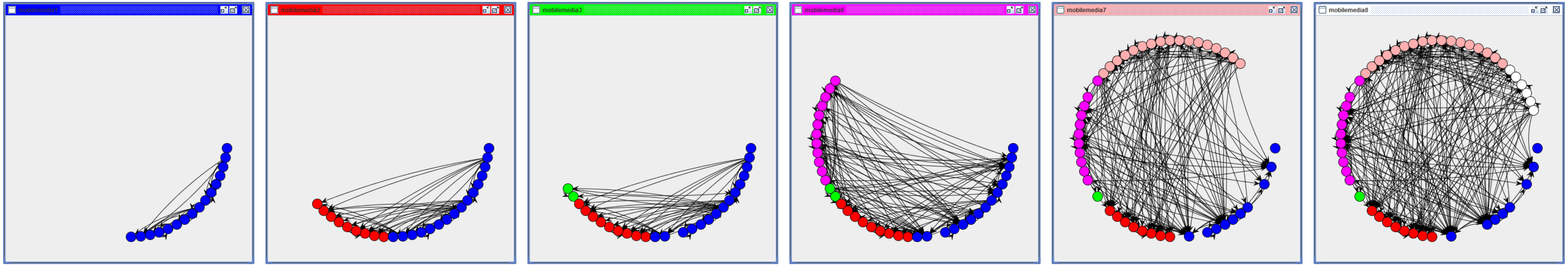


Software evolution & dependencies

- As software systems evolve, the **amount and complexity of the interactions** amongst their components often increases.
  - More coupling.
  - **“*Undesired*” dependencies** amongst certain components (e.g., layer bridging, direct access to databases, cycles).
  - <u>Degradation</u> of intended design.

Software evolution & dependencies

- As software systems evolve, the **amount and complexity of the interactions** amongst their components often increases.
  - More coupling.
  - **“*Undesired*” dependencies** amongst certain components (e.g., layer bridging, direct access to databases, cycles).
  - <u>Degradation</u> of intended design.



Software evolution & dependencies

- Conscious efforts must be made to stop (or alleviate) degradation.
  - Plan for corrective actions (e.g., refactoring).
  - Monitor system health (e.g., via metrics).
  - Conformance checks.

Software evolution & dependencies

- Conscious efforts must be made to stop (or alleviate) degradation.
  - Plan for corrective actions (e.g., refactoring).
  - Monitor system health (e.g., via metrics).
  - Conformance checks.

The early detection of architectural smells is **important** for architects, so that they can **plan ahead** for actions that **preserve** the **quality** of the system.

What can we do about it?

- In a forward-looking scenario, **architects** would want to know:
  - Which modules are <u>likely to be coupled</u> in the near <u>near future</u>.
  - Which smells are <u>more harmful</u> for the system.

This architecture-level analysis requires **<u>to anticipate dependency-related problems</u>** in order **<u>to proactively look for solutions</u>**.

What can we do about it?

- In a forward-looking scenario, **architects** would want to know:
  - Which modules are likely to be coupled in the near future.
  - Which smells are more harmful for the system.

This architecture-level analysis requires **to anticipate dependency-related problems** in order **to proactively look for solutions**.

Predict when a dependency-related problem is likely to manifest!

Social network analysis to the rescue!

- In the last decade, research has been devoted to study:
  - How smells are **introduced**.
  - How smells **evolve**.
  - What their **effect** is on program comprehension.

- However, research on how to **predict the appearance** of architectural smells **has been scarce**.

- A particular graph-based approach is Social Networks Analysis (SNA), which has been used for modelling both nature and human phenomena.

SNA techniques can predict links that yet do not exist between pairs of nodes in a network.

Social network analysis to the rescue!

… and software engineering?

We hypothesise that **software systems** and
their underlying architectures **behave** as **social networks**.

- **Evidence** that the **topological** features of **dependency graphs** can reveal interesting properties of the software system under analysis.

- Nonetheless, ***SNA*** techniques has not yet greatly exploited in the Software Engineering community.

Social network analysis to the rescue!

… and software engineering?

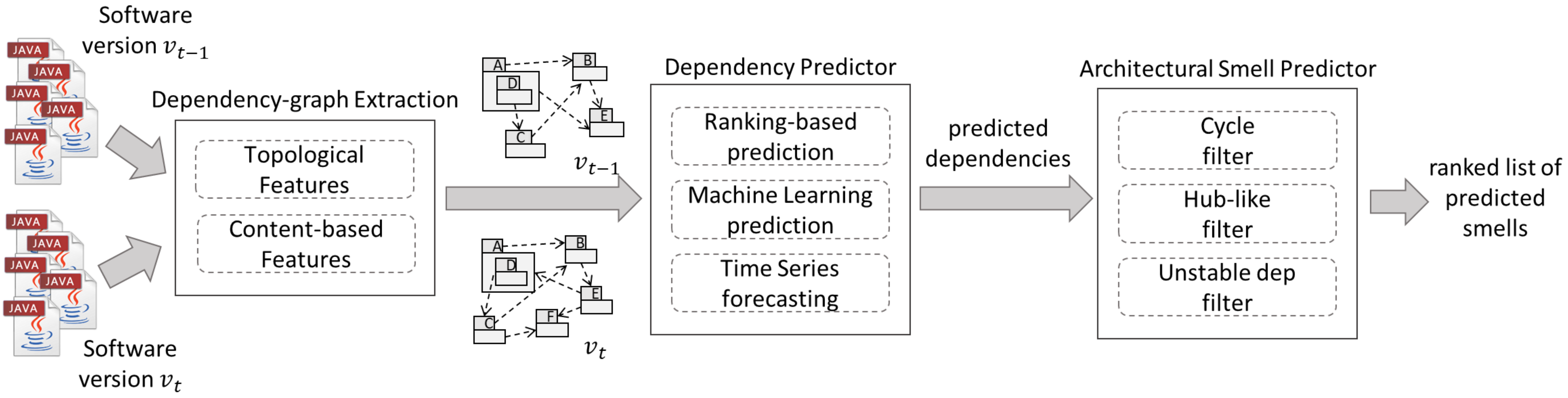
- **RQ1.** How do architectural **smells evolve** over system versions, in terms of increasing or decreasing their dependency configurations?

- **RQ2.** What criteria are useful for **assessing similarity of design elements** with respect to link prediction?

- **RQ3.** Can **past system versions affect, and improve the predictions** of, the design structure of a future version?

- **RQ4.** To what extent **Machine Learning** techniques can **aid** in the **prediction** of architectural smells?

Prediction Overview



What have we done so far?

- Ongoing research has primary focused on the definition of the dependency graph and the evaluation of the dependency predictor.

- Complemented the definition of the dependency graph with a statistical analysis of software versions and the evolution of SNA metrics.

- Analysed how past decisions reflected in the software structure affect the future occurrence of dependencies, and smells thereof.

- Analysed the descriptive power of both topological and content-based features for defining the similarity of components.

- Smell prediction focused on cycles and hubs.

What do we do now?

- Perform a systematic study with more systems and other dependency-based smells.

- The prediction capabilities are sensitive to the prediction model.
  - Analyse and extend the set of features used.
  - Considering software specific-metrics?

- Smells might not be harmful.
  - How can we train a model to discard them?

- Can we predict the appearance of new nodes (e.g. new packages, classes)?

- Can we predict the disappearance of dependencies?

Thanks!

Questions?

Contact me at: antonela.tommasel@isistan.unicen.edu.ar

A few papers…

Recommending friends in the misinformation era

- Tommasel A., Rodriguez J.M., Godoy D. **”An experimental study on feature engineering and learning approaches for aggression detection in social media”**. Inteligencia Artificial. ISSN: 1137-3601. Iberamia. DOI: 10.4114/intartif.vol22iss63pp81-100 http://journal.iberamia.org/index.php/intartif/article/view/264/83

- Antonela Tommasel, Daniela Godoy. “**Multi-view Community Detection with Heterogeneous Information from Social Media Data**” Neurocomputing. ISSN: 0925-2312. Elsevier. https://www.sciencedirect.com/science/article/pii/S0925231218301528

- Tommasel A., Godoy D. *”***A Social-aware Online Short-text Feature Selection Technique for Social Media***”*. Information Fusion. ISSN: 1872-6305. Elsevier. DOI: 10.1016/j.inffus.2017.05.003 Vol. 40. pp 1–17. March 2018. http://www.sciencedirect.com/science/article/pii/S1566253516302354

- Tommasel A., Godoy D. ***“Learning and Adapting User Criteria for Recommending Followees in Social Networks”.*** Journal of the Association for Information Science and Technology. ISSN: 1532-2890. Wiley-Blackwell. DOI: 10.1002/asi.23861 http://onlinelibrary.wiley.com/doi/10.1002/asi.23861/abstract

- Tommasel A., Corbellini A., Godoy D., Schiaffino S. **"Personality-aware followee recommendation algorithms: An empirical analysis"**. Engineering Applications of Artificial Intelligence. ISSN 0952-1976. Elsevier. 2016. DOI: 10.1016/j.engappai.2016.01.016. http://www.sciencedirect.com/science/article/pii/S0952197616000208

A few papers…

Keeping one-step ahead of architectural smells

- Diaz-Pace, J.A., Tommasel, A., and Godoy, D. "**Can Network Analysis Techniques help to Predict Design Dependencies? An Initial Study**". In Proceedings of the IEEE INTERNATIONAL CONFERENCE ON SOFTWARE ARCHITECTURE (ICSA 2018). Seattle USA. April, 2018. https://arxiv.org/abs/1808.02776v1

- Diaz-Pace, J.A., Tommasel, A., and Godoy, D. "**Towards Anticipation of Architectural Smells using Link Prediction Techniques**". In Proceedings of the 18th IEEE International Working Conference on Source Code Analysis and Manipulation (SCAM 2018). Madrid, Spain. September, 2018. http://arxiv.org/abs/1808.06362

- Tommasel A., Diaz-Pace J.A., Godoy D. "**A Dependency-based Prediction Approach for Architectural Smells**". Sent to Journal of Systems and Software. Elsevier. (Tool included!)

Friends or Foe: Recommending friends in the misinformation era

Dr Antonela Tommasel

ISISTAN, CONICET-UNICEN, Argentina

antonela.tommasel@isistan.unicen.edu.ar

CONICET

I S I S T A N