Friends or Foe: Recommending friends in the misinformation era

Dr Antonela Tommasel
ISISTAN, CONICET-UNICEN, Argentina
antonela.tommasel@isistan.unicen.edu.ar
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
  - …
1. Introduction & motivation

2. Work proposal

3. Previous work & research agenda
"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"
"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!!
"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)

**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!!**
"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)

**Faking it!**

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.

  ![Diagram]

  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
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!!
Faking it!

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).
Faking it!

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.
Friends or Foe: Recommending friends in the misinformation era

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.
Friends or Foe: Recommending friends in the misinformation era

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.

Detection will be less effective if training data is not periodically updated. It can be challenging given the incompleteness, noise and volume of it.
Faking it!

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.

- Detection will be less effective if training data is not periodically updated.
- It can be challenging given the incompleteness, noise and volume of it.
- Each type of misinformation presents particular characteristics, which must be taken into account for detection.
Faking it!

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.

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

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

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

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 similar elements is NOT sufficient to ensure reliability!

We need Relevant and Trustworthy recommendations!
And what about recommendations?

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

We need **Relevant** and **Trustworthy** recommendations!

**New challenges!**
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.
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.

Few studies have incorporated this concept to:
• Social media
• The proliferation of unwanted content recommendations

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

• 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

1. Introduction & motivation

2. Work proposal

3. Previous work & research agenda
The goal is to define a profile to
What do we want?

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

describe and estimate the trustworthiness or reputation of 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.
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 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, aiming at balancing both the relevance and reliability of recommendations.
Research Questions

RQ1. What is the utility of the multiple sources of heterogeneous 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

Evolution Social Network

Detecting unreliable content and accounts

Profile Features
Content-based Features
Linguistic Features
Social Network Features

Interaction Willingness
Interaction Quality
Seriousness in Interactions
Reliability of published content
Reliability of friends
Building the trustworthiness profile

Content-based Features
Topological Features
Trustworthiness Profiles

Building a Recommendation Model

External validation: Fact-Checking

Recommended Friends
Non Recommended Friends
Trustworthy recommendations

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

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.
Trustworthy recommendations

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:
    • 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

- News Crawler
- Fake News Crawler
- Fact-Checking Crawler

Content Crawler
- URL Crawler
- Image Search & Crawling

News Selection

Post Crawler
- Reactions Crawler

Retrieving Temporal Information

User Crawler
- Network Crawler

Retrieving Social Context

manual labelling

multi-sourced dataset

A. Tommasel
Friends or Foe: Recommending friends in the misinformation era
ISISTAN, CONICET-UNICEN
Trustworthy recommendations

Undesired content & users detection

- News Crawler
- Fake News Crawler
- Fact-Checking Crawler
- Content Crawler
- URL Crawler
- Image Search & Crawling
- News Selection
- Post Crawler
- Reactions Crawler
- User Crawler
- Network Crawler

manual labelling

Retrieving Temporal Information

Retrieving Social Context

multi-sourced dataset

A. Tommasel
Friends or Foe: Recommending friends in the misinformation era
Friends or Foe: Recommending friends in the misinformation era

Trustworthy recommendations

Undesired content & users detection

- News Crawler
- Fake News Crawler
- Fact-Checking Crawler
- Content Crawler
- URL Crawler
- Image Search & Crawling
- News Selection
- Post Crawler
- Reactions Crawler
- User Crawler
- Network Crawler
- Retriving Temporal Information
- Retrieving News Content
- Retrieving Social Context

multi-sourced dataset
Trustworthy recommendations

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

Diagram:
- Interaction Willingness
- Interaction Quality
  - Seriousness in Interactions
- Reliability of published content
- Reliability of friends
- Building the trustworthiness profile

Nodes:
- User_1: -trustworthy
- User_2: -trustworthy
- User_3: +trustworthy
- User_4: +trustworthy
- User_5: +trustworthy
- User_6: +trustworthy
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 hierarchical conceptual model to characterise different aspects of the information diffusion process, focusing not only in the information being disseminated, but also on the role of users in such process.

Shed some light on the psychological and social motivations, and attitudes towards the diffusion and consumption of content in social media.
Trustworthy recommendations

Behaviour conceptual model

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

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.**
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.** 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.**
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.** 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.**
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.** 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 patterns.
Trustworthy recommendations

Goal #3
Integrate the profile in a recommendation system!
# Table of contents

1. Introduction & motivation

2. Work proposal

3. Previous work & research agenda
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.
Trustworthy recommendations

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.
- 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.

• Refining aggressive and hate speech detection.

• Extending unreliable content detection for unreliable user detection.

• Integrating everything in the profile!

• More Evaluations!!!

Integrate the techniques for multi-class detection!

Aiming for a multi-class classification!
Bonus track
Keeping one-step ahead of architectural smells
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).
- **Degradation** 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).
  - **Degradation** 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 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.
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!
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.
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

Dependency-graph Extraction

Software version $v_{t-1}$

Topological Features
Content-based Features

Software version $v_t$

Dependency Predictor

Ranking-based prediction
Machine Learning prediction
Time Series forecasting

predicted dependencies

Architectural Smell Predictor

Cycle filter
Hub-like filter
Unstable dep filter

ranked list of predicted smells

A. Tommasel
Friends or Foe: Recommending friends in the misinformation era
Lessons Learned

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.
Lessons Learned

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  

- Antonela Tommasel, Daniela Godoy. “Multi-view Community Detection with Heterogeneous Information from Social Media Data” Neurocomputing. ISSN: 0925-2312. Elsevier.  


A few papers...

*Keeping one-step ahead of architectural smells*


Friends or Foe: Recommending friends in the misinformation era

Dr Antonela Tommasel
ISISTAN, CONICET-UNICEN, Argentina
antonela.tommasel@isistan.unicen.edu.ar