CONICET

ISISTAN

Recommender systems and misinformation
What can we do about them?

ANTONELA TOMMASEL

About me

- Dr Antonela Tommasel
  - PhD in Computer Sciences at UNICEN (Argentina).
  - Bachelor in Software Engineering.
- Assistant Researcher at ISISTAN, CONICET-UNICEN.
- Assistant Professor at UNICEN.
- Research Interests:
  - Recommender systems
  - Text mining
  - Social media
  - Social computing
  - eCitizenship
  - Hate speech

CONICET

ISISTAN

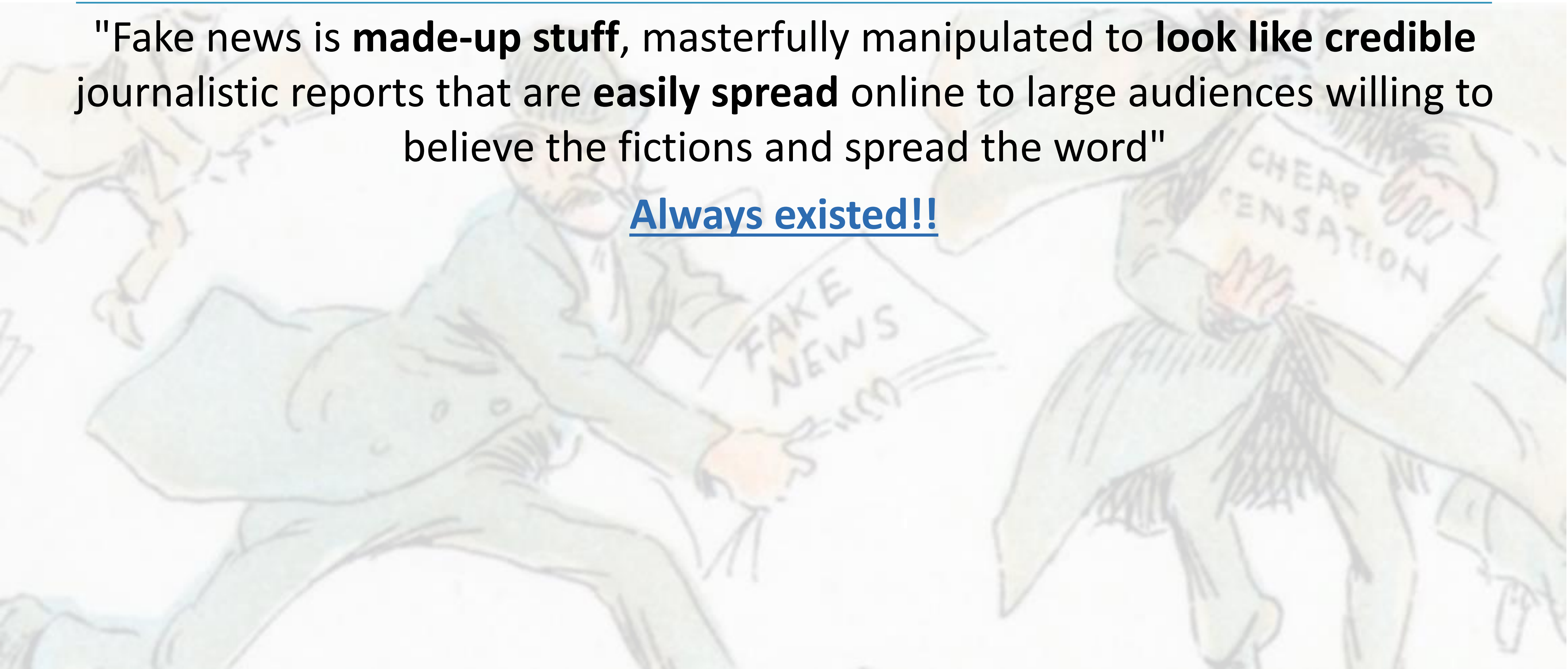
Misinformation is among us!

"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"

Misinformation is among us!

"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!!



Misinformation is among us!

"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"

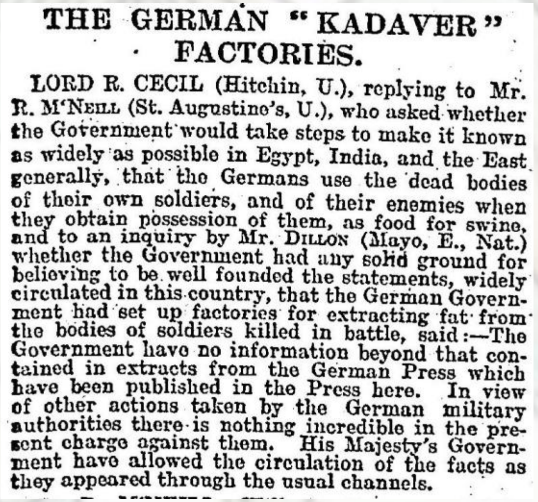
Always existed!!

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)

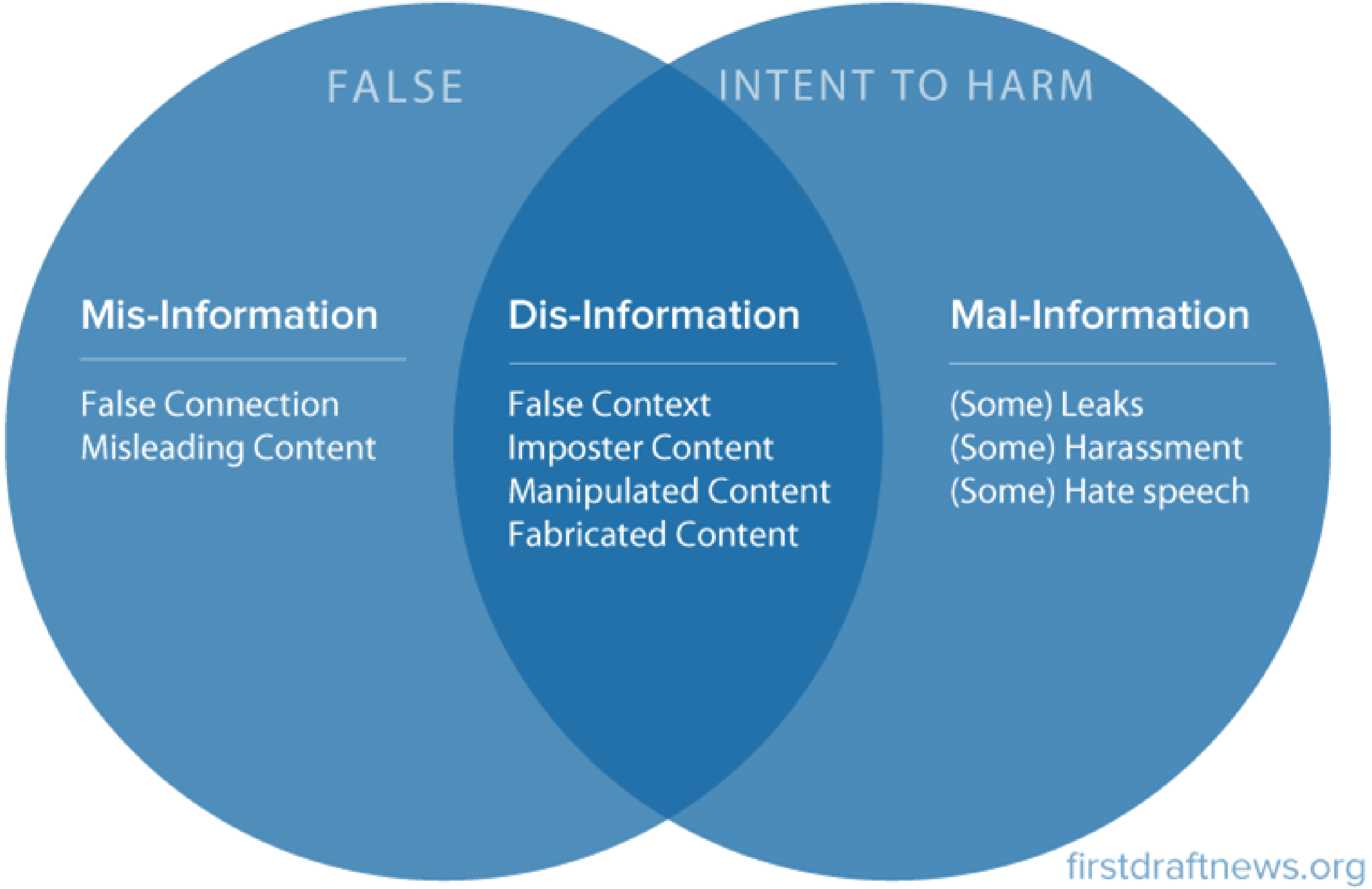


THE GERMAN "KADAVER" FACTORIES.

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.

Misinformation is among us!

…but it is not only misinformation!



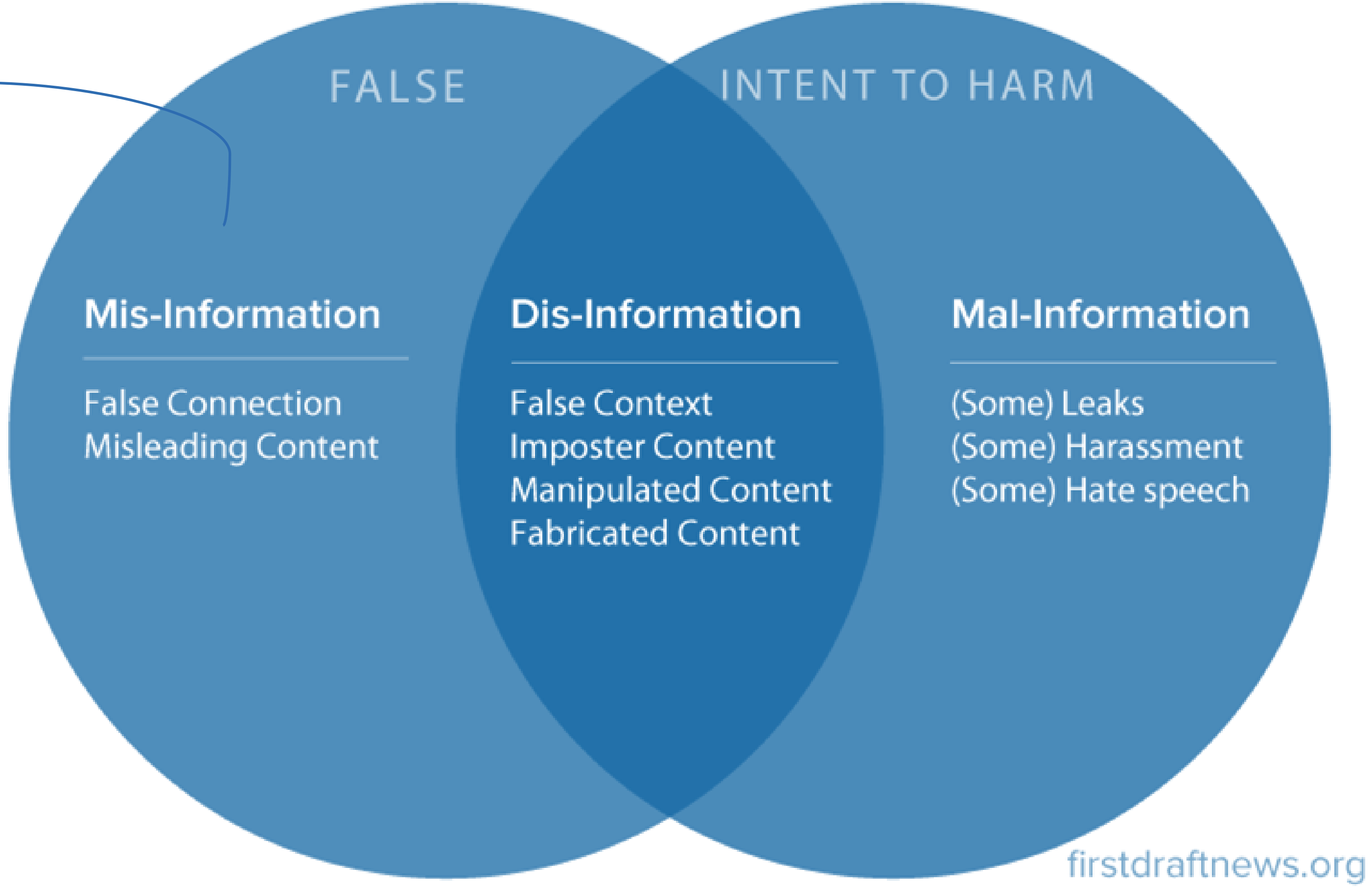
https://firstdraftnews.org/latest/fake-news-complicated/

Misinformation is among us!

…but it is not only misinformation!

Rumours. For example, related to the attack of Charlie Hebdo, advices for preventing COVID…

In most cases, no harm is meant, people is caught up in the moment trying to be helpful and do not inspect what they are sharing.

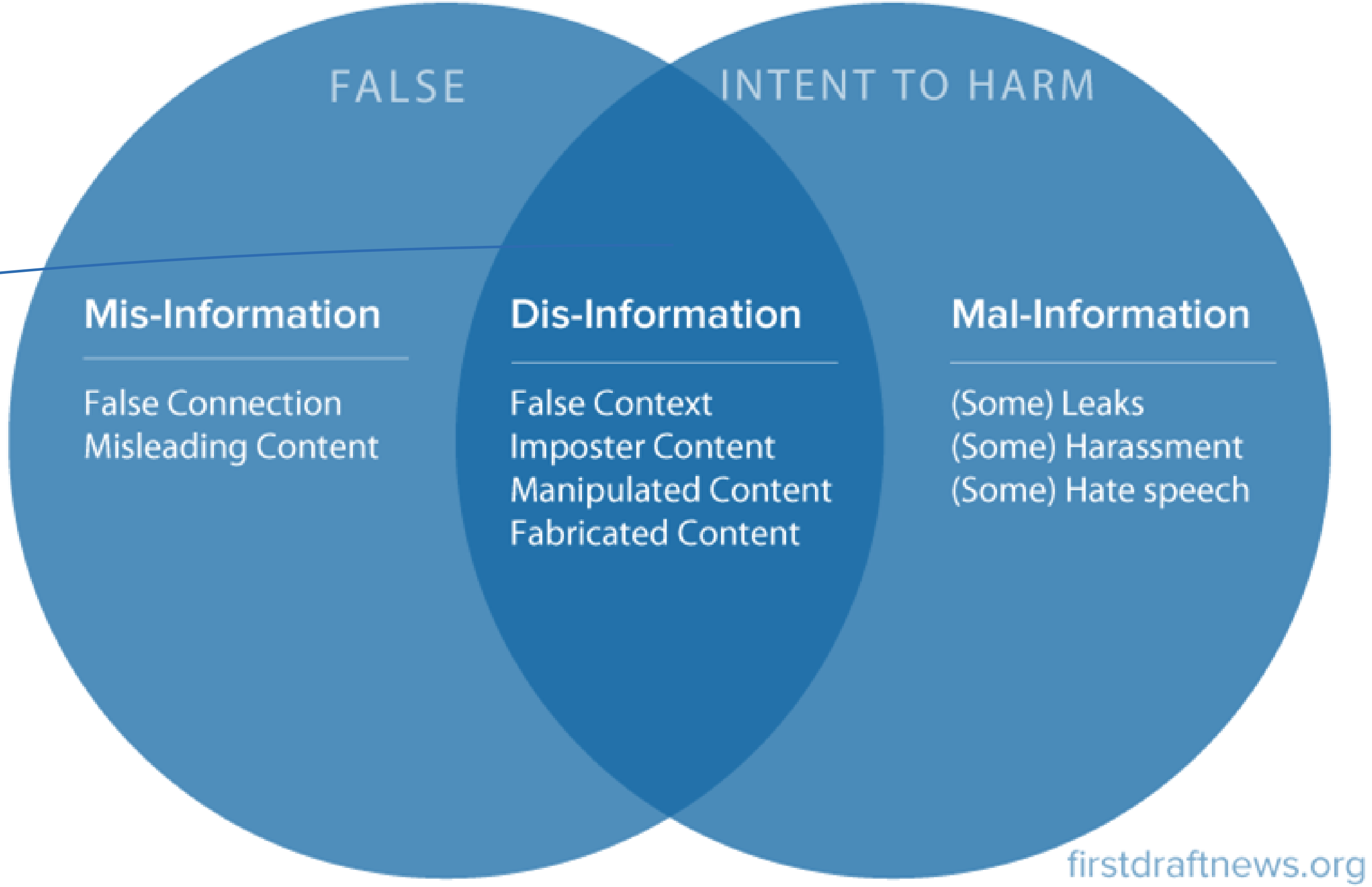


https://firstdraftnews.org/latest/fake-news-complicated/

Misinformation is among us!

…but it is not only misinformation!

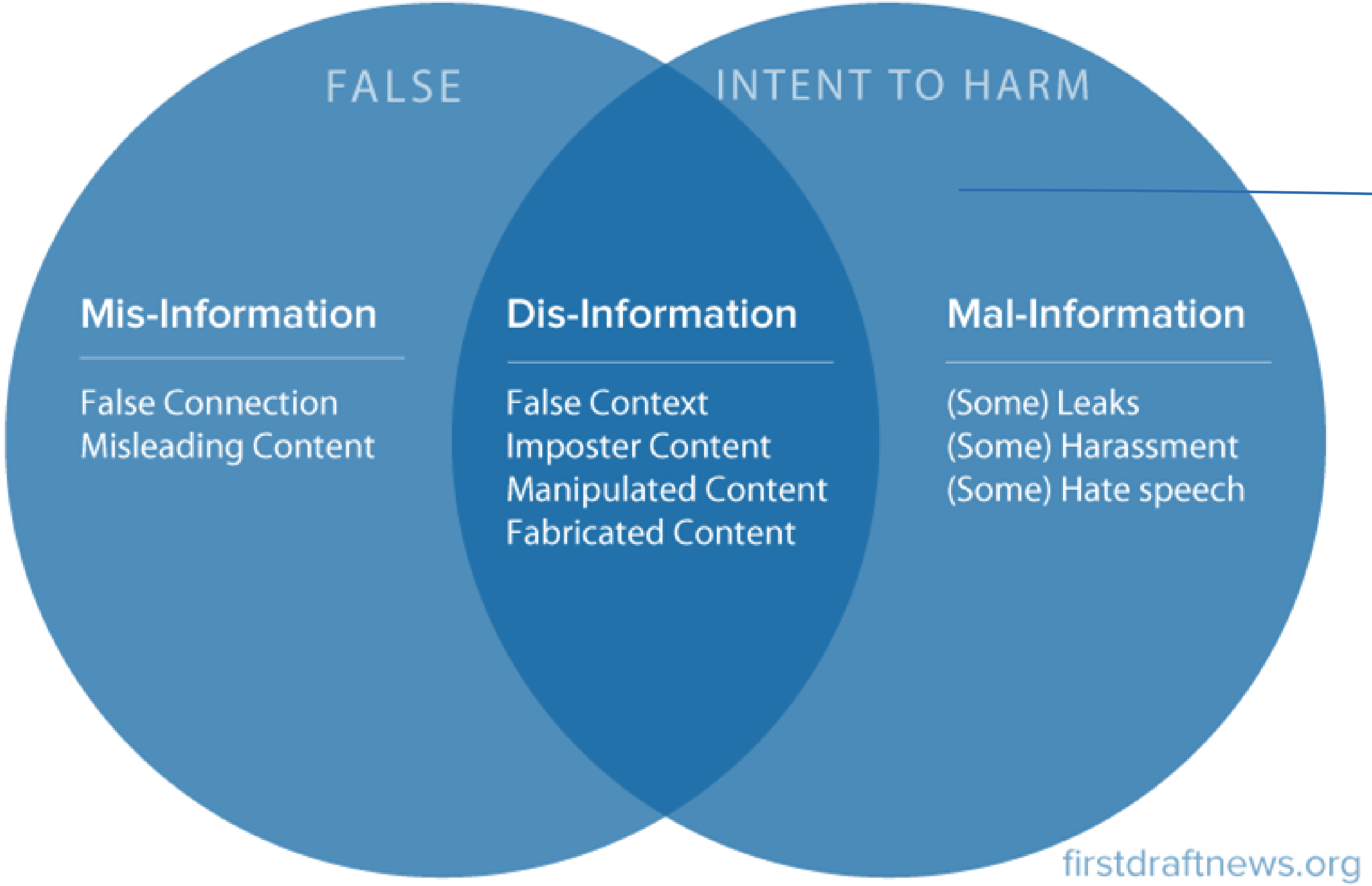
Hoaxes, rumours that could have legal consequences, accusations.



https://firstdraftnews.org/latest/fake-news-complicated/

Misinformation is among us!

…but it is not only misinformation!

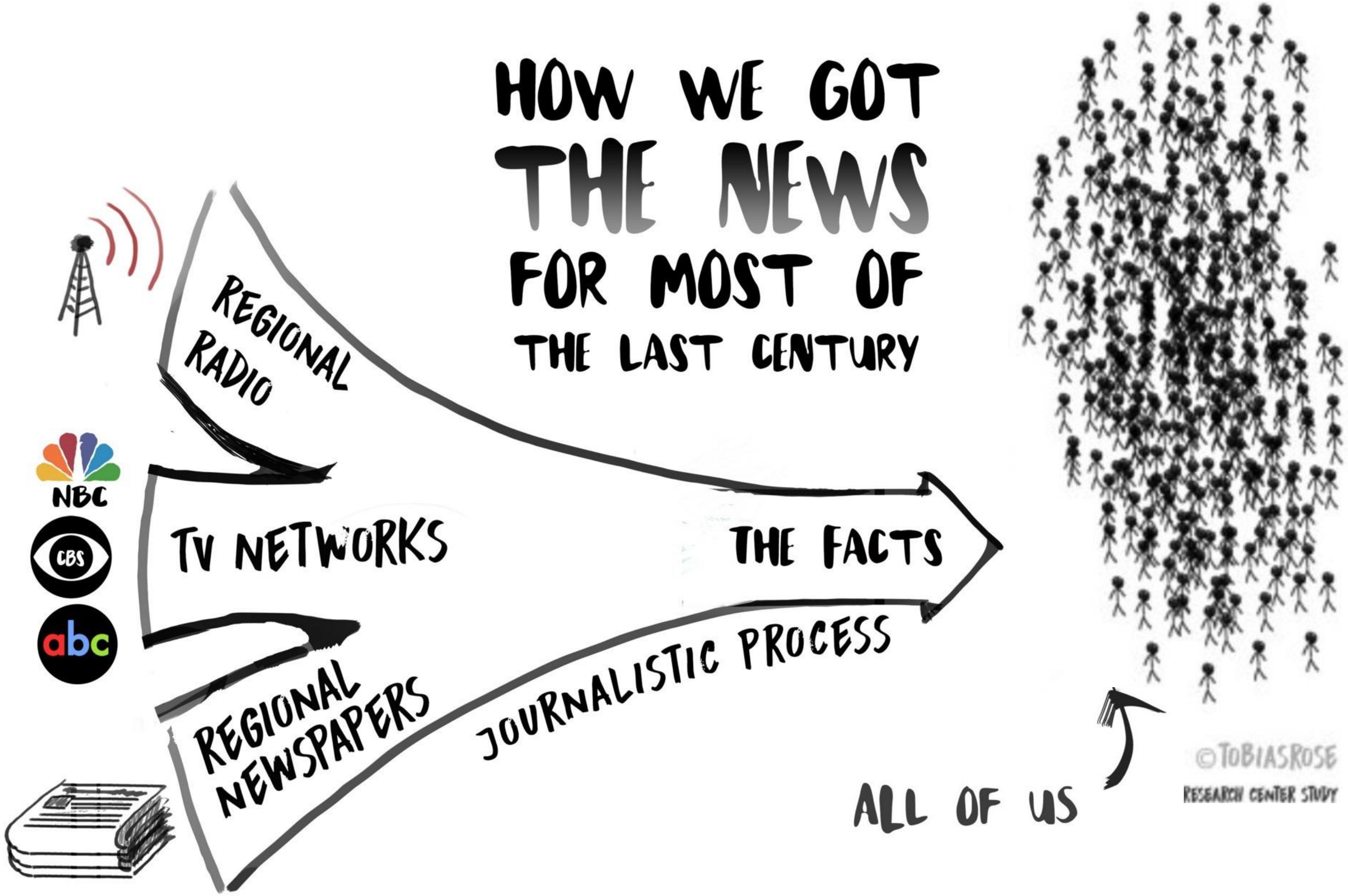


Leaking information with the intent to harm, but not realizing that the leaked information is false.

https://firstdraftnews.org/latest/fake-news-complicated/

Misinformation is among us!

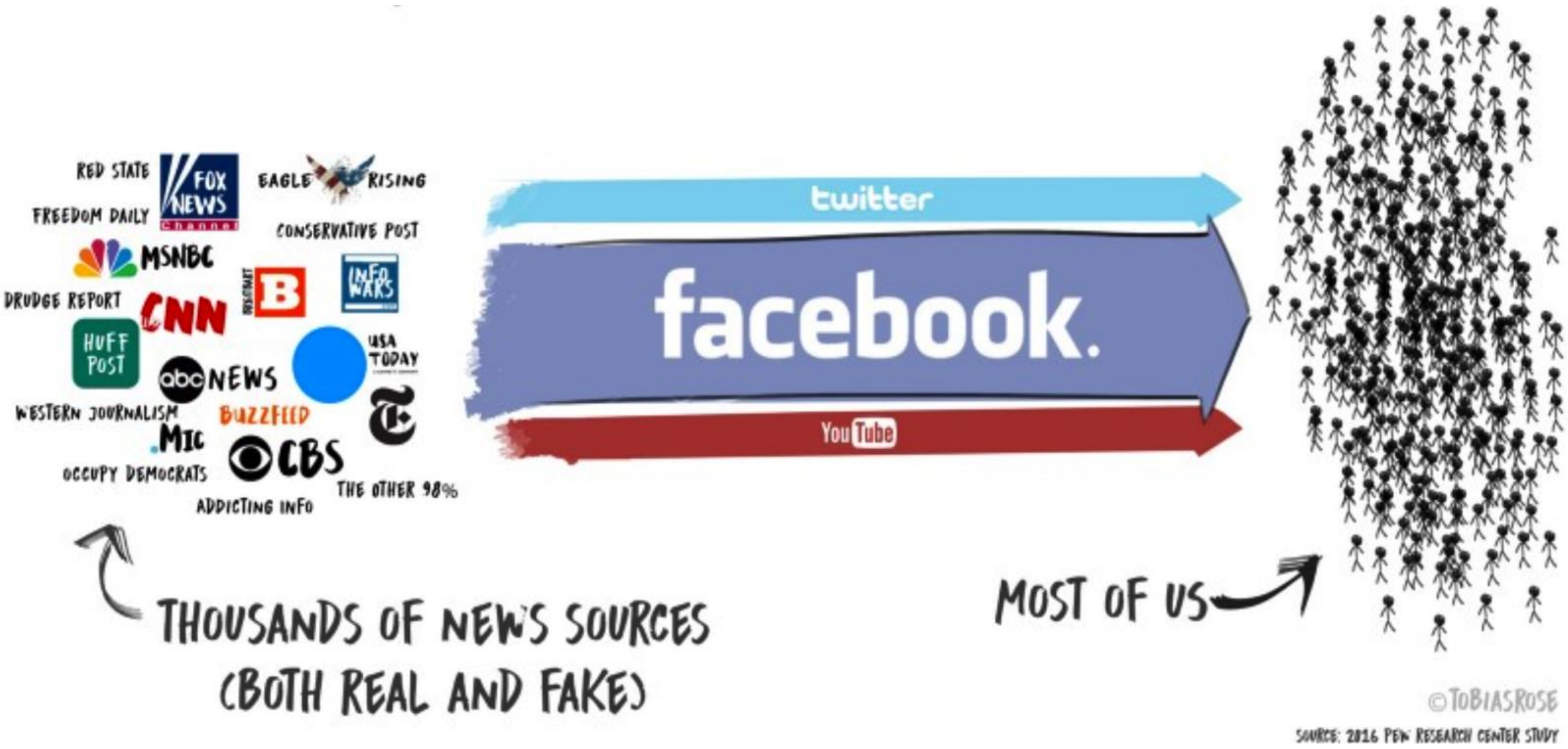
Social media aggravates the problem!



https://medium.com/@tobiasrose/empathy-to-democracy-b7f04ab57eee#.100kciuhj

Misinformation is among us!

Social media aggravates the problem!



https://medium.com/@tobiasrose/empathy-to-democracy-b7f04ab57eee#.100kciuhj

Misinformation & Social Media!

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

Misinformation & Social Media!

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

- **Fake or unreliable content can severely affect society**, posing significant threats to democracies and economy.
  - With the COVID-19 pandemic, health misinformation arose as a threat to public health.

- **Can affect how people perceive content.**
  - Repeated exposure can alter the likelihood of accepting fake content as truth, especially when the fake content aligns with internal beliefs.
  - The line between what is fake or not becomes more uncertain hindering the differentiation between fake and authentic content.
  - <u>The trustworthiness of the entire news ecosystem might be at risk.</u>

Misinformation is among us!

Social media aggravates the problem!

fvckyofeelings
@yvngtwat

WOW, can't believe sharks are swimming the streets/yard of NYC/New Jersey. Scary😱😱 keep yourselves safe guys #Sandy

♡ 85 5:18 AM - Oct 30, 2012

590 people are talking about this

Anonymous
@YourAnonNews

The Statue of Liberty right now - pic.twitter.com/WaXBbZUc | #sandy

♡ 243 3:05 PM - Oct 29, 2012

1,779 people are talking about this

Alberto Fernández difundió una "fake news" sobre la pandemia del coronavirus: qué recomendó

"La Organización Mundial de la Salud, entre las cosas que recomienda, **es que uno tome muchas bebidas calientes, porque precisamente el calor mata el virus**", dijo Fernández.

Alberto Fernández
@alferdez

Según un informe publicado por el@@washingtonpost y realizado por el Instituto Tecnológico de Massachusetts (MIT) Evo Morales ganó los comicios electorales del año pasado por más de 10 puntos de diferencia, sin que mediara fraude alguno
clarin.com/mundo/informe-...

3:06 pm · 29 Feb 2020 · Twitter for iPhone

10.6K Retweets **23.3K** Likes

Anibal Fernández
@FernandezAnibal

Llegó el Corralito
El BCRA anuncia que no se venden mas dólares

OFICINA DE PRENSA Y DIFUSIÓN

Ministerio de Hacienda
Presidencia de la Nación

13/08/2018

ANUNCIO DEL MINISTERIO DE HACIENDA

El Ministerio de Hacienda informa que en consideración a la posición de liquidez en pesos que ha acumulado, ha instruido al Banco Central de la República Argentina a discontinuar las ventas de dólares diarias hasta que las necesidades de pesos lo requieran nuevamente.

Hipólito Yrigoyen 250 5to piso Oficina 501, CABA
Tel: 4349-5112 / 5105

♡ 1.136 13:51 - 13 ago. 2018

1.308 personas están hablando de esto

What are we doing about it?

How misinformation spreads?

How to avoid misinformation derived effects?

How to identify users propagating misinformation?

What are we doing about it?

How misinformation spreads?

How to avoid misinformation derived effects?

How to identify users propagating misinformation?

- How can recommender systems affect misinformation spreading patterns?
- How can recommender systems boost the influence of misinformation spreaders?

What are we doing about it?

How misinformation spreads?

How to avoid misinformation derived effects?

How to identify users propagating misinformation?

- How can recommender systems help avoid the formation of echo chambers/filter bubbles?

What are we doing about it?

How misinformation spreads?

How to avoid misinformation derived effects?

How to identify users propagating misinformation?

- How can we characterize misinformation spreaders?
- Are social interaction patterns relevant?

What are we doing about it?

How misinformation spreads?

How to avoid misinformation derived effects?

How to identify users propagating misinformation?

Antonela Tommasel, Filippo Menczer. "**Do Recommender Systems Make Social Media More Susceptible to Misinformation Spreaders?**". In proceedings of the 16th ACM Conference on Recommender Systems. Association for Computing Machinery. Seattle, USA. DOI: 10.1145/3523227.3551473

Recommender systems and misinformation propagation

- **Recommender systems** play an important role as **mediators of information propagation**.

- They have been deemed as **one of the major culprits of misinformation spreading**.
  - **Disruptive consequences** in our society.

Recommender systems and misinformation propagation

- **Recommender systems** play an important role as **mediators of information propagation**.

- They have been deemed as **one of the major culprits of misinformation spreading**.
  - **Disruptive consequences** in our society.

- **People-recommender systems** or **link prediction** techniques are of special interest.
  - Directly contribute to the **evolution of the social network structure**, affecting the information and the opinions users are exposed to.

Recommender systems and misinformation propagation

- **Recommender systems** play an important role as **mediators of information propagation**.

- They have been deemed as **one of the major culprits of misinformation spreading**.
  - **Disruptive consequences** in our society.

- **People-recommender systems** or **link prediction** techniques are of special interest.
  - Directly contribute to the **evolution of the social network structure**, affecting the information and the opinions users are exposed to.

How can we assess the effect of link prediction techniques on misinformation propagation and polarization?

Recommender systems and misinformation propagation

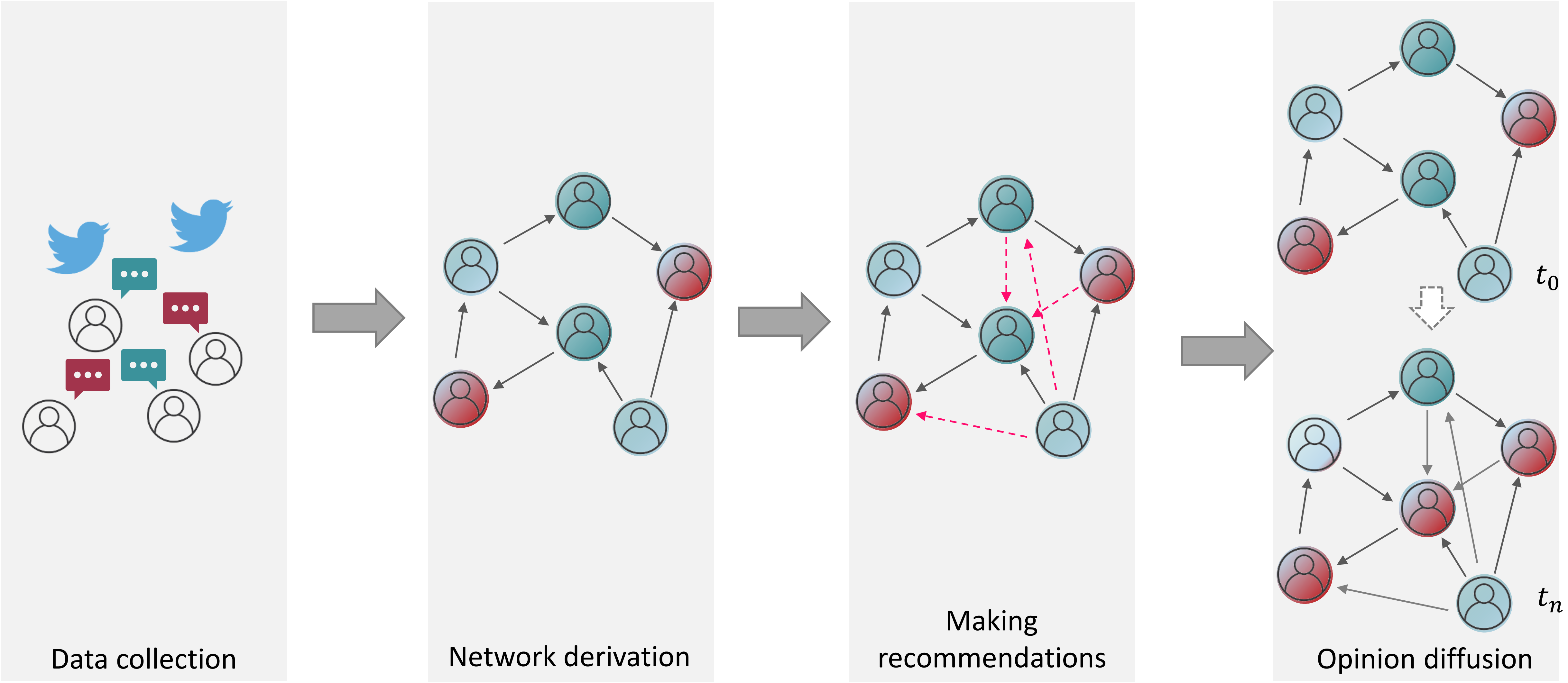
What do we want to study?

How can we assess the effect of link prediction techniques on misinformation propagation and polarization?

We combine a **link prediction technique** with an **opinion dynamics model** to simulate the behavior of individuals changing their opinions as a consequence of their interactions with their neighborhood, within a social network that is continuously evolving.

Recommender systems and misinformation propagation

Pipeline



Recommender systems and misinformation propagation

Pipeline: Data collection

Data collection

- We used the **FibVid** data collection.
  - The collection comprised **772** COVID-related news claims and **112k** relevant tweets belonging to **24k** users, which were shared during 2020.

- Tweets were retrieved using the [Faking it!](#) tool.

Recommender systems and misinformation propagation

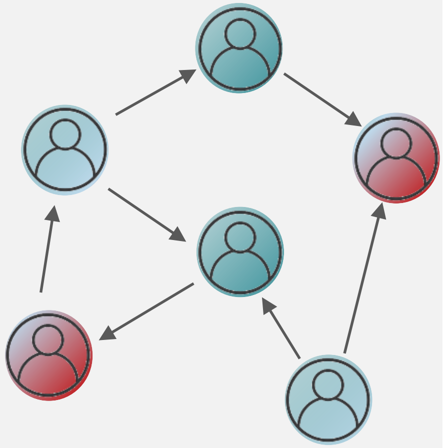
Pipeline: Data collection

Data collection

- We used the **FibVid** data collection.
  - The collection comprised **772** COVID-related news claims and **112k** relevant tweets belonging to **24k** users, which were shared during 2020.
- Tweets were retrieved using the **Faking it!** tool.
- Tweets have a authentic/fake label based on Politifact and Snopes.
  - 26% authentic content, 74% fake content.
  - Labels were used to determine whether users were fake news spreaders.
  - Users were deemed as spreaders if the **proportion of shared fake content was higher than a certain threshold** (0.5).

Recommender systems and misinformation propagation

Pipeline: Network derivation

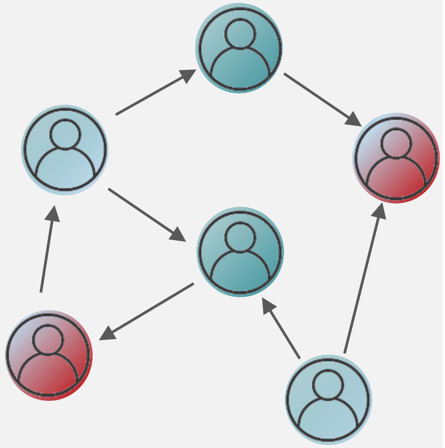


Network derivation

- We build a **user network**.
- Users are related based on tweet interactions.
  - Reply, mention, retweet.
  - Follow the information flow direction.
- Network is directed and weighted.
  - Weighted according to the number of interactions.
- Users are tagged according to whether they share fake or real content.
  - We consider the **percentage of fake tweets shared**.
  - We use a threshold to define whether a user is a misinformation spreader (0.5).

Recommender systems and misinformation propagation

Pipeline: Network derivation

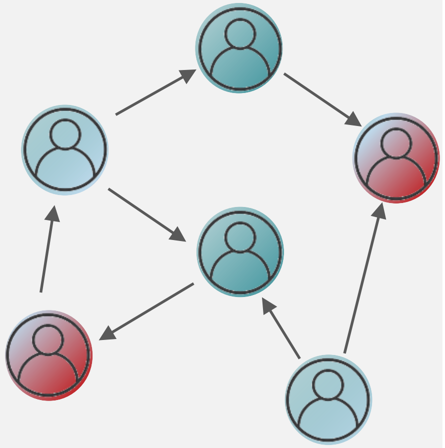


Network derivation

- Temporal partitioning of data.
  - 70% of interactions to training set, 30% to test set.
- In the training network, **41%** of users were deemed as **spreaders**, **49% non-spreaders**, and the remaining **10% were neutral**.
- When considering the entire network, the spreader/non-spreader distribution changed. As users shared more tweets, the proportion of users that could be considered spreaders increased (i.e., the tendency of users to share misinformation increased).
  - **80% of spreaders**, **10% of non-spreaders**, and **10% of neutral users**.
  - Only **2% reverted** their misinformation spreading behaviour.

Recommender systems and misinformation propagation

Pipeline: Network derivation

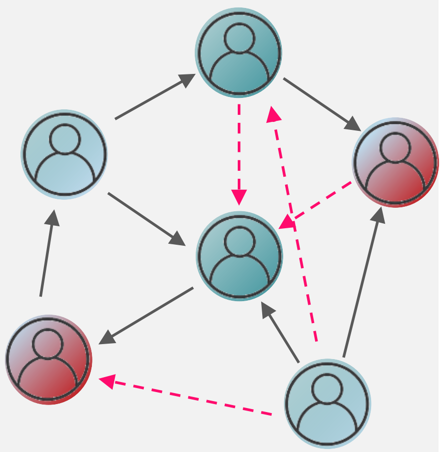


Network derivation

- We quantified the **polarization of the selected users** relying on the relation between **users' score and the score of the users with whom they interacted** to determine the homophily levels of each group.
- On average, **27% of spreaders' interactions were with other spreaders**, with 17% of interactions with users with higher or equal scores.
- **Non-spreaders interacted with users on a broader range of scores**, accounting for only 16% of interactions with spreaders.
- Both spreaders and non-spreaders tended to interact more with non-spreaders, with non-spreaders showing a higher homophilic behaviour than spreaders.

Recommender systems and misinformation propagation

Pipeline: Making recommendations



Making recommendations

Popularity

Random

Topological

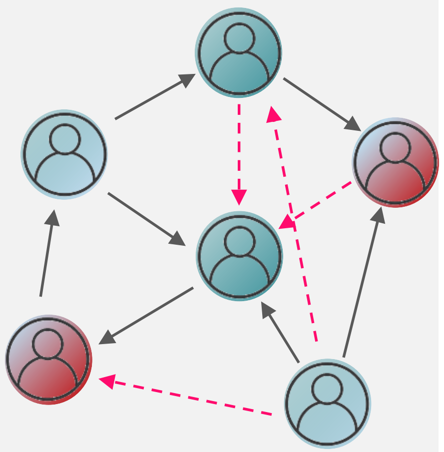
Content

Friend of Friends

Implit MF

Recommender systems and misinformation propagation

Pipeline: Making recommendations



Making recommendations

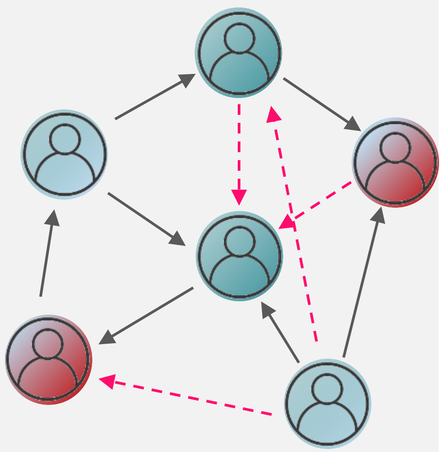
Popularity | Random | Topological | Content | Friend of Friends | Implit MF

- Evaluation did not aim to endorse recommendations but check whether they would be close to the actual user interactions.
  - If recommendations do not follow interactions, then the resulting network structure might not be representative of user behaviour.

Relevance | Diversity novelty | Misinformation exposure

Recommender systems and misinformation propagation

Pipeline: Making recommendations



Making recommendations

Popularity | Random | Topological | Content | Friend of Friends | Implit MF

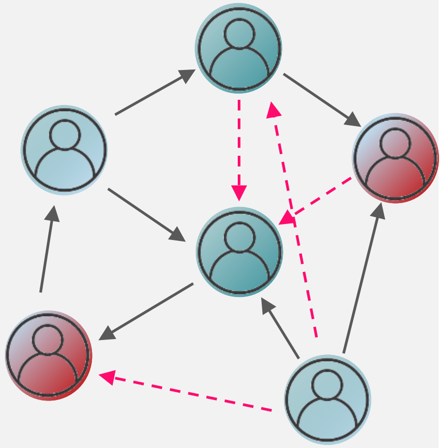
- Evaluation did not aim to endorse recommendations but check whether they would be close to the actual user interactions.
    - If recommendations do not follow interactions, then the resulting network structure might not be representative of user behaviour.

Relevance

- Precision@k
- nDCG@k

Recommender systems and misinformation propagation

Pipeline: Making recommendations



Making recommendations

Popularity

Random

Topological

Content

Friend of Friends

Implit MF

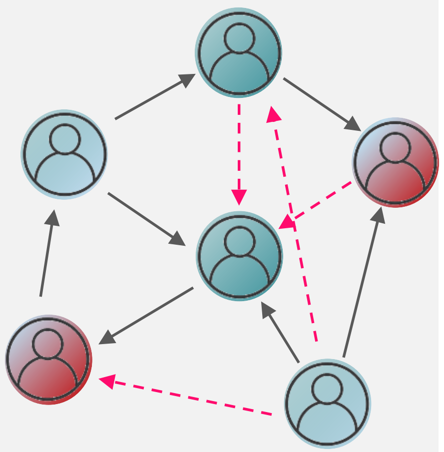
- Evaluation did not aim to endorse recommendations but check whether they would be close to the actual user interactions.
  - If recommendations do not follow interactions, then the resulting network structure might not be representative of user behaviour.

Diversity novelty

- Variations of intra-list dissimilarities:
  - **Diversity**
  - **Novelty**
- Euclidean distance over <u>structural and content-based representations</u>.

Recommender systems and misinformation propagation

Pipeline: Making recommendations



Making recommendations

Popularity | Random | Topological | Content | Friend of Friends | Implit MF

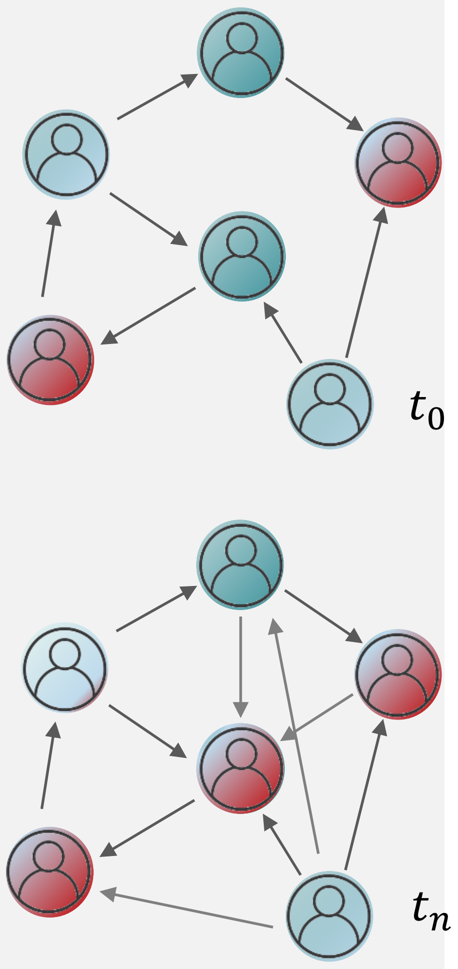
- Evaluation did not aim to endorse recommendations but check whether they would be close to the actual user interactions.
    - If recommendations do not follow interactions, then the resulting network structure might not be representative of user behaviour.

Misinformation exposure

- % of spreaders in recommendations.

Recommender systems and misinformation propagation

Pipeline: Opinion diffusion



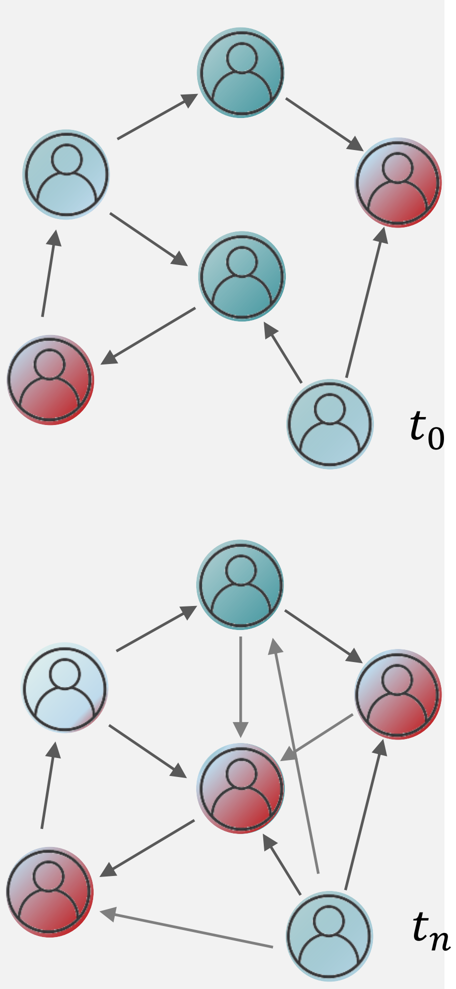
Opinion diffusion

- We **add recommendations to the base graph and simulate.**

- **Q-voter** model.
  - A node is updated is Q randomly selected neighbours shared the same opinion.
  - Simple model, yet commonly used in dynamics analysis.
  - <u>Asynchronous</u>. In each iteration a node is randomly selected to be updated.

Recommender systems and misinformation propagation

Pipeline: Opinion diffusion

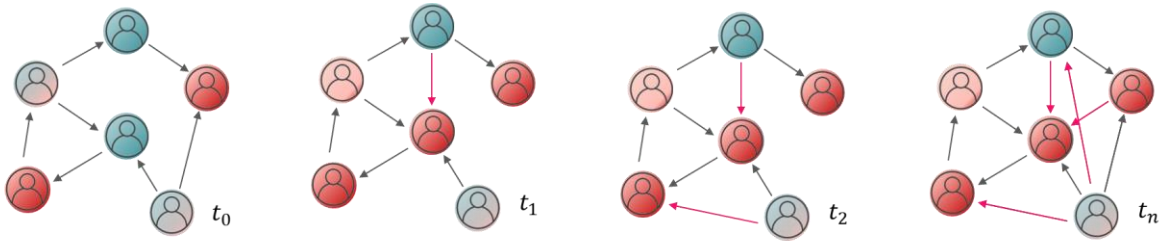


Opinion diffusion

- We **add recommendations to the base graph and simulate.**

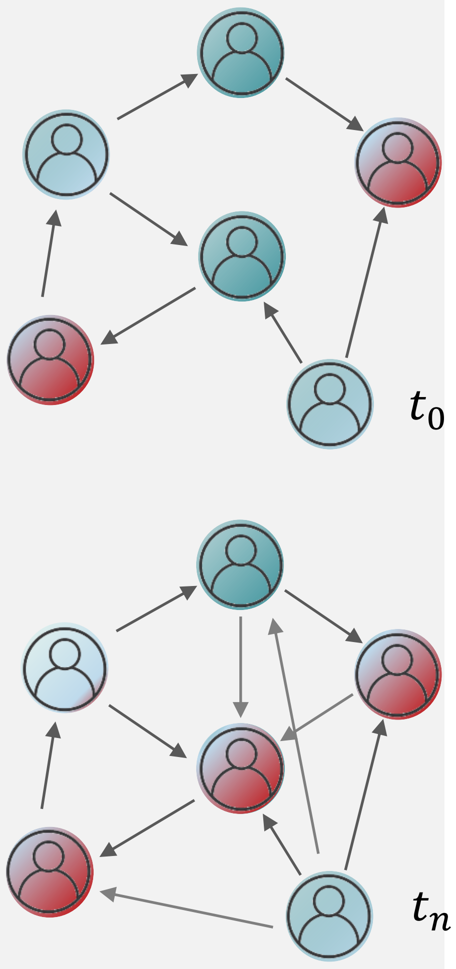
- **Q-voter** model.
  - A node is updated is Q randomly selected neighbours shared the same opinion.
  - Simple model, yet commonly used in dynamics analysis.
  - Asynchronous. In each iteration a node is randomly selected to be updated.

- **Dynamic**.
  - Recommendations are added in steps to the graph during the simulation.
    - One recommendation is added to a random node in each step.



Recommender systems and misinformation propagation

Pipeline: Opinion diffusion



Opinion diffusion

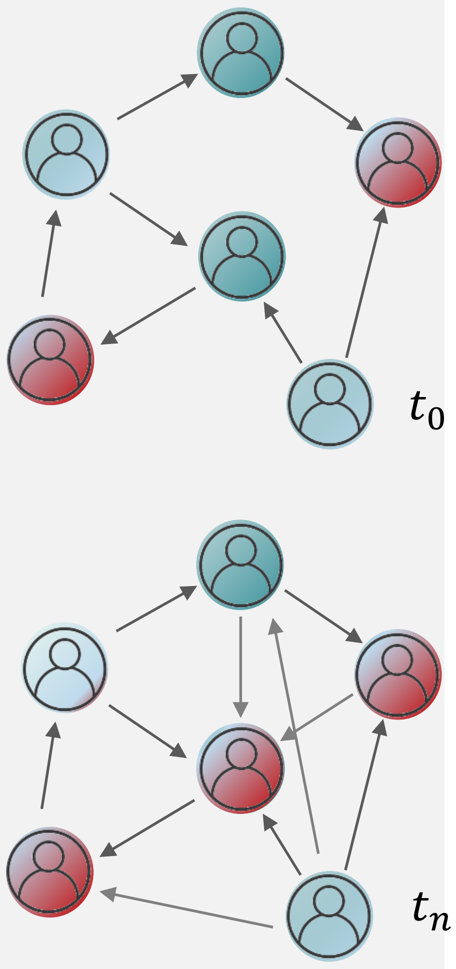
- We **add recommendations to the base graph and simulate.**

- **Q-voter** model.
  - A node is updated is Q randomly selected neighbours shared the same opinion.
  - Simple model, yet commonly used in dynamics analysis.
  - Asynchronous. In each iteration a node is randomly selected to be updated.

- **Edge removal.**
  - For each added edge, we remove one.
    - Keep density.
    - Allows disregarding any effect related to graph densification.
    - Removing eldest edge from the user to which an edge is added.

Recommender systems and misinformation propagation

Pipeline: Opinion diffusion



Opinion diffusion

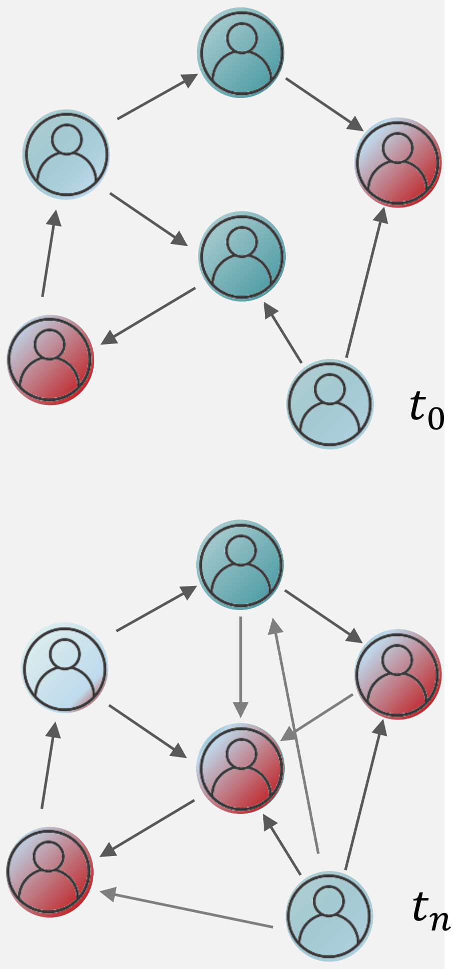
- We **add recommendations to the base graph and simulate.**

- As **evaluation**, compare characteristics of the induced networks and the original one.
  - How metrics varied between the actual network (with real spreaders) and the "recommended network" (with simulated spreaders).

% final spreaders	Clustering coefficient	User interaction polarization	Random Walk Controversy

Recommender systems and misinformation propagation

Pipeline: Opinion diffusion



Opinion diffusion

- We **add recommendations to the base graph and simulate.**

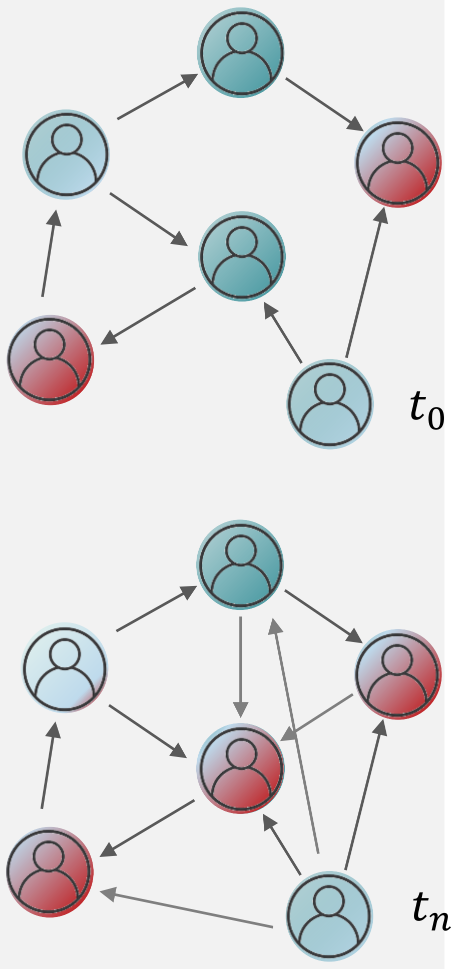
- As **evaluation**, compare characteristics of the induced networks and the original one.
  - How metrics varied between the actual network (with real spreaders) and the "recommended network" (with simulated spreaders).

% final spreaders

- How many users are spreaders at the end of the simulation?
- What was the spreader conversion rate?

Recommender systems and misinformation propagation

Pipeline: Opinion diffusion



Opinion diffusion

- We **add recommendations to the base graph and simulate.**

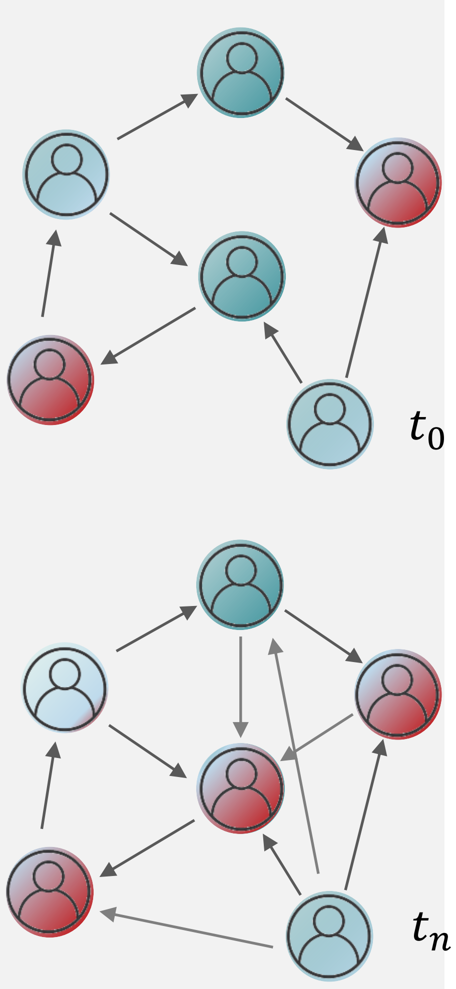
- As **evaluation**, compare characteristics of the induced networks and the original one.
  - How metrics varied between the actual network (with real spreaders) and the "recommended network" (with simulated spreaders).

Clustering coefficient

- Are spreaders more central to the network?
- Do they appear more often in the paths between other nodes?

Recommender systems and misinformation propagation

Pipeline: Opinion diffusion



Opinion diffusion

- We **add recommendations to the base graph and simulate.**

- As **evaluation**, compare characteristics of the induced networks and the original one.
  - How metrics varied between the actual network (with real spreaders) and the “recommended network” (with simulated spreaders).

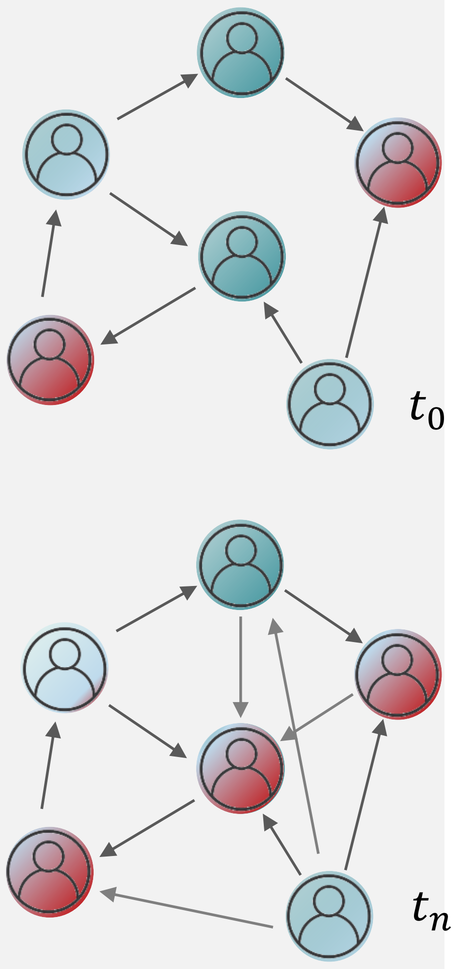
User interaction polarization

- How polarized are the neighbourhoods of users?
- How are spreaders distributed among communities?

- Pearson correlation between users’ spreadness state and their neighbours state.

Recommender systems and misinformation propagation

Pipeline: Opinion diffusion



Opinion diffusion

- We **add recommendations to the base graph and simulate.**
- As **evaluation**, compare characteristics of the induced networks and the original one.
  - How metrics varied between the actual network (with real spreaders) and the "recommended network" (with simulated spreaders).

Random Walk Controversy

- How likely are users to encounter content from the other group?
  - Indication of polarization.

Experimental evaluation

RQ1. How do recommenders contribute to **misinformation spreaders recommendations**?

RQ2. How do recommenders contribute to **amplifying the influence** of misinformation spreaders?

Experimental evaluation

RQ1. Contribution to **misinformation spreader recommendation**

	Precision	Recall	Content-based dissimilarities		Structural dissimilarities		Misinformation Exposure
			Ind Diversity	Ind Novelty	Ind Diversity	Ind Novelty	
Base graph	-	-	0.077	0.294	0.118	0.461	0.18
Random	0.1	0.401	0.373	0.343	0.599	0.547	0.382
Popularity	0.111	0.626	0.248	0.301	0.373	0.431	0.164
Friend-of-Friends	0.203	0.581	0.218	0.258	0.366	0.427	0.196
Topology – Resource Allocation	0.231	0.553	0.248	0.301	0.373	0.431	0.164
Content-based	0.1	0.5	0.381	0.346	0.601	0.548	0.371
Implicit MF	0.109	0.558	0.324	0.347	0.483	0.467	0.161

Experimental evaluation

RQ1. Contribution to **misinformation spreader recommendation**

	Precision	Recall
Base graph	-	-
Random	0.1	0.401
Popularity	0.111	0.626
Friend-of-Friends	0.203	0.581
Topology – Resource Allocation	0.231	0.553
Content-based	0.1	0.5
Implicit MF	0.109	0.558

- In general, **recommenders achieved low precision and moderate nDCG.**
- The **highest relevance** results were obtained with the **topology-based** recommenders.
- **Popularity** achieved the **highest nDCG**.

Experimental evaluation

RQ1. Contribution to **misinformation spreader recommendation**

	Content-based dissimilarities		Structural dissimilarities	
	Ind Diversity	Ind Novelty	Ind Diversity	Ind Novelty
Base graph	0.077	0.294	0.118	0.461
Random	0.373	0.343	0.599	0.547
Popularity	0.248	0.301	0.373	0.431
Friend-of-Friends	0.218	0.258	0.366	0.427
Topology – Resource Allocation	0.248	0.301	0.373	0.431
Content-based	0.381	0.346	0.601	0.548
Implicit MF	0.324	0.347	0.483	0.467

- Generally, recommendations’ **structural diversity/novelty were higher than content-based diversity/novelty**.
    - These recommendations could **affect network rewiring during simulations** as new edges might connect far away users.

- Recommenders achieving high relevance also achieved low diversity/novelty scores (e.g., popularity and topological).
- Recommenders achieving low relevance also achieved the highest diversity/novelty (e.g., content and random).

Experimental evaluation

RQ1. Contribution to **misinformation spreader recommendation**

	Misinformation Exposure
Base graph	0.18
Random	0.382
Popularity	0.164
Friend-of-Friends	0.196
Topology – Resource Allocation	0.164
Content-based	0.371
Implicit MF	0.161

- Recommendations included **less than 40% of misinformation spreaders**.
- Scores showed that the **most popular recommended users were mostly non-spreaders**.
- In general, **the ratio of recommended spreaders was higher for spreaders than for non-spreaders**.
  - Spreaders might be inserted in echo chambers that are strengthened as an effect of recommendations.

Experimental evaluation

RQ1. Contribution to **misinformation spreader recommendation**

	Precision	Recall	Content-based dissimilarities		Structural dissimilarities		Misinformation Exposure
			Ind Diversity	Ind Novelty	Ind Diversity	Ind Novelty	
Base graph	-	-	0.077	0.294	0.118	0.461	0.18
Random	0.1	0.401	0.373	0.343	0.599	0.547	0.382
Popularity	0.111	0.626	0.248	0.301	0.373	0.431	0.164
Friend-of-Friends	0.203	0.581	0.218	0.258	0.366	0.427	0.196
Topology – Resource Allocation	0.231	0.553	0.248	0.301	0.373	0.431	0.164
Content-based	0.1	0.5	0.381	0.346	0.601	0.548	0.371
Implicit MF	0.109	0.558	0.324	0.347	0.483	0.467	0.161

Recommenders fostering relevance	Recommenders fostering diversity/novelty
Topology Popularity	Content-based Random

- The **best relevance performing** recommenders were **the ones recommending the fewest spreaders**.
- Recommenders with **increasing diversity/novelty** tended to **recommend the highest ratio of spreaders**.

Experimental evaluation

RQ2. Contribution to **amplification of spreaders' influence**

	% spreaders	Clustering Coefficient	User interaction polarization	Random Walk Controversy Score
Base graph	0.345	0.033	0.26	0.193
Random	0.042	0.003	0.19	0.147
Popularity	0.366	0.71	0.16	0.34
Friend-of-Friends	0.367	0.137	0.24	0.363
Topology – Resource Allocation	0.368	0.069	0.203	0.445
Content-based	0.2	0.01	0.173	0.143
Implicit MF	0.369	0.03	0.214	0.226

Experimental evaluation

RQ2. Contribution to **amplification of spreaders' influence**

	% spreaders
Base graph	0.345
Random	0.042
Popularity	0.366
Friend-of-Friends	0.367
Topology – Resource Allocation	0.368
Content-based	0.2
Implicit MF	0.369

- **None** of the recommenders **greatly increased the proportion of spreaders in the network**.
  - Content-based and random decreased it.

- Content and random achieved both the **largest number of recommended spreaders** and the **largest number of users becoming spreaders** for at least one iteration, converting 52% of users.

Experimental evaluation

RQ2. Contribution to **amplification of spreaders' influence**

	Clustering Coefficient
Base graph	0.033
Random	0.003
Popularity	0.71
Friend-of-Friends	0.137
Topology – Resource Allocation	0.069
Content-based	0.01
Implicit MF	0.03

- Significant **increments** were observed for **topology and popularity**.
- Significant **decrements** were observed for **content and random**.

- The interactions added by content and random do not contribute to increasing user centrality.
  - The clustering coefficient for non-spreaders was higher than for spreaders.
  - **Non-spreaders had a more central role in the network**, while spreaders might not be well connected, although scattered across the whole network.

Experimental evaluation

RQ2. Contribution to **amplification of spreaders' influence**

	User interaction polarization
Base graph	0.26
Random	0.19
Popularity	0.16
Friend-of-Friends	0.24
Topology – Resource Allocation	0.203
Content-based	0.173
Implicit MF	0.214

- **Most recommenders contributed to increase the proportion of users interacting with spreaders**.
  - Popularity accounting for the 95% of users with a spreader in their neigbourhood.
  - Random greatly decreased the proportion for non-spreaders, and increased it for spreaders, inducing segregation in the network.

Experimental evaluation

RQ2. Contribution to **amplification of spreaders' influence**

	Random Walk Controversy Score
Base graph	0.193
Random	0.147
Popularity	0.34
Friend-of-Friends	0.363
Topology – Resource Allocation	0.445
Content-based	0.143
Implicit MF	0.226

- **FoF increased group segregation, while the other recommenders caused users to mix.**
- When analyzing the individual groups, **interactions between non-spreaders seemed to consolidate**.
  - All recommenders increased the likelihood of spreaders interacting, and perhaps influencing, non-spreaders.
  - Conversely, popularity increased the likelihood of non-spreaders interacting with spreaders.

Experimental evaluation

RQ2. Contribution to **amplification of spreaders' influence**

	% spreaders	Clustering Coefficient	User interaction polarization	Random Walk Controversy Score
Base graph	0.345	0.033	0.26	0.193
Random	0.042	0.003	0.19	0.147
Popularity	0.366	0.71	0.16	0.34
Friend-of-Friends	0.367	0.137	0.24	0.363
Topology – Resource Allocation	0.368	0.069	0.203	0.445
Content-based	0.2	0.01	0.173	0.143
Implicit MF	0.369	0.03	0.214	0.226

Recommending a large number of spreaders does not directly lead to a high conversion rate.
Instead, recommended spreaders also need to be well connected with their neighbours to affect spreading.

Recommenders diversifying interactions and fostering connections with users in other network regions seemed to have a stronger effect on spreaders presence and interaction dynamics.

Summary (I)

We presented a **preliminary exploration** to better understand how **user recommenders affect network dynamics** in terms of **misinformation spreader distribution and influence**.

Simulations could help evaluate potential scenarios to test new or modified recommenders and assess their effects before deployment.

Our study brings to attention the **potential implications of recommenders** in **network evolution and dynamics**.

What are we doing about it?

How misinformation spreads?

How to avoid misinformation derived effects?

How to identify users propagating misinformation?

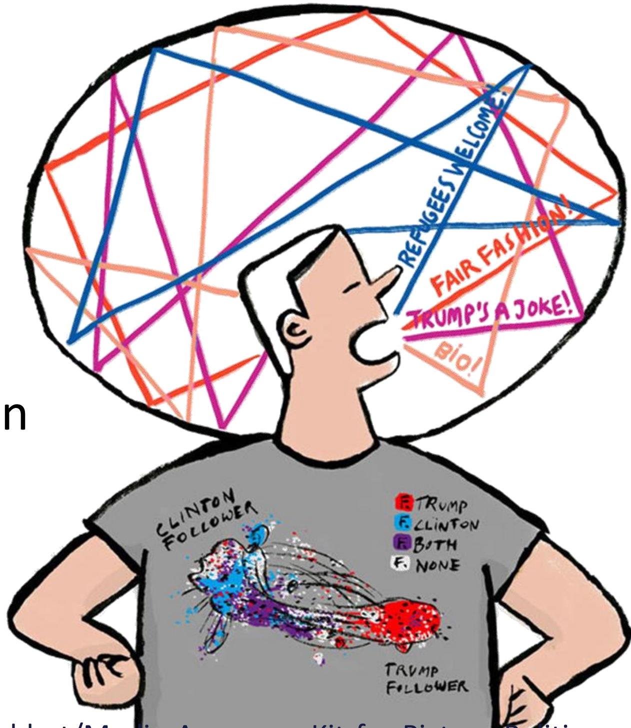
Antonela Tommasel, Juan Manuel Rodriguez, Daniela Godoy, "**I Want to Break Free! Recommending Friends from Outside the Echo Chamber**". In proceedings of the 15th ACM Conference on Recommender Systems. Association for Computing Machinery, Amsterdam, Netherlands. DOI: 10.1145/3460231.3474270

Echo chambers & recommendations

- **Recommender systems** play an important role as **mediators of information propagation**.
  - They are affected by the different forms of online harms, hindering their ability to achieve accurate predictions, thus becoming unintended means for spreading and amplifying harms .

- Echo chambers are related to situations in which **individuals only consume content or interact** with other users expressing their **same points of view**.
  - Selective exposure, biased assimilation, and group polarization.

- Echo chambers concern not only political discourses but also in conspiracy theories, in which they could lead to a **stronger radicalization, seclusion from society** and **destructive actions**.



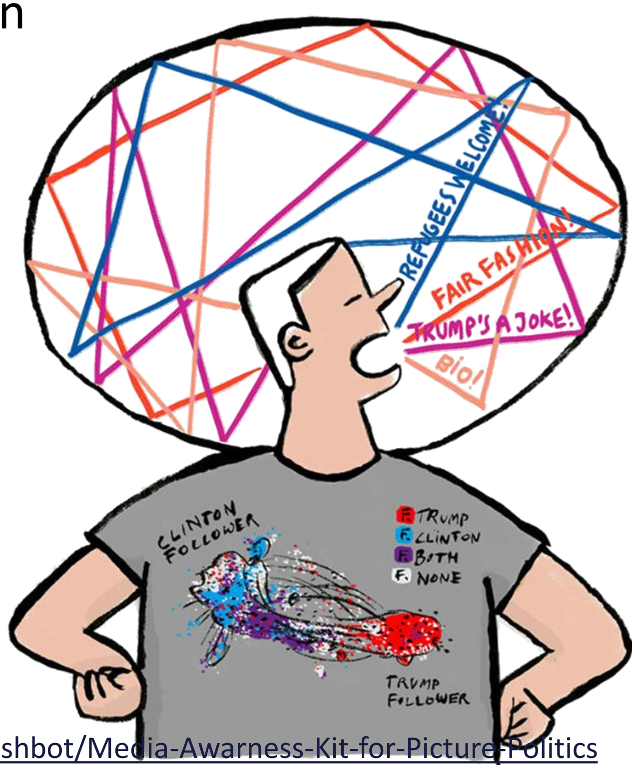
https://alexandraklobouk.com/filter/pushbot/Media-Awarness-Kit-for-Picture-Politics

Echo chambers & recommendations

- Echo chambers are related to situations in which **individuals only consume content or interact** with other users expressing their **same points of view**.
  - Selective exposure, biased assimilation, and group polarization.

- Echo chambers concern not only political discourses but also in conspiracy theories, in which they could lead to a **stronger radicalization, seclusion from society** and **destructive actions**.

Harnessing recommender systems with misinformation- and harm-aware mechanisms becomes **essential to mitigate** the negative effects of the **propagation of online harms** and **increase** the user-perceived **quality** of recommender systems.



https://alexandraklobouk.com/filter/pushbot/Media-Awarness-Kit-for-Picture-Politics

Echo-chamber aware recommendations

The problem

We tackle the **friend recommendation problem** by fostering **recommendation diversification** in an **echo chamber awareness** setting.

We rely on **implicitly modeling the echo chamber** membership of users to present them with **relevant friend recommendations from outside** the influence of their community.

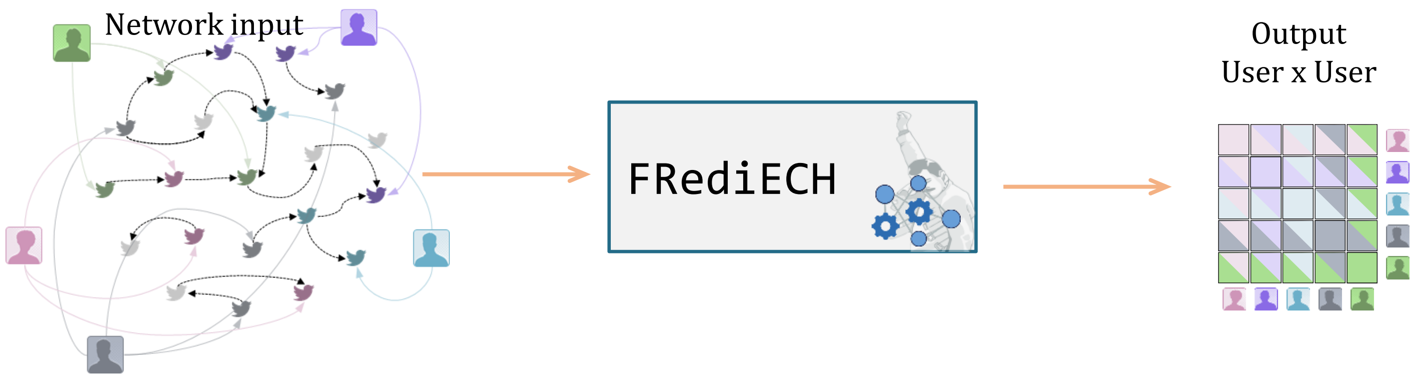
Echo-chamber aware recommendations

The problem

We tackle the **friend recommendation problem** by fostering **recommendation diversification** in an **echo chamber awareness** setting.

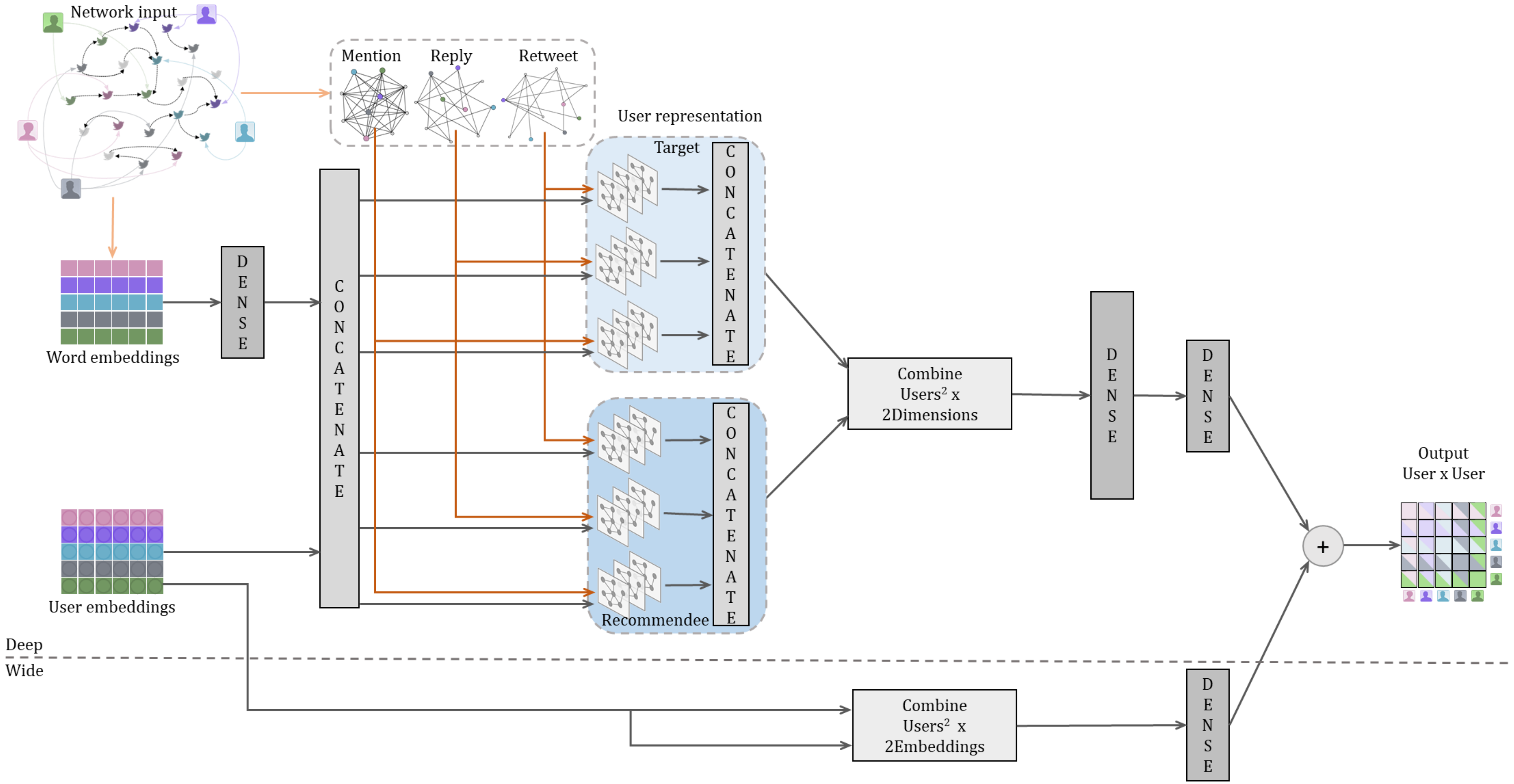
We rely on **implicitly modeling the echo chamber** membership of users to present them with **relevant friend recommendations from outside** the influence of their community.

`Friend REcommenDer for breakIng Echo CHambers`



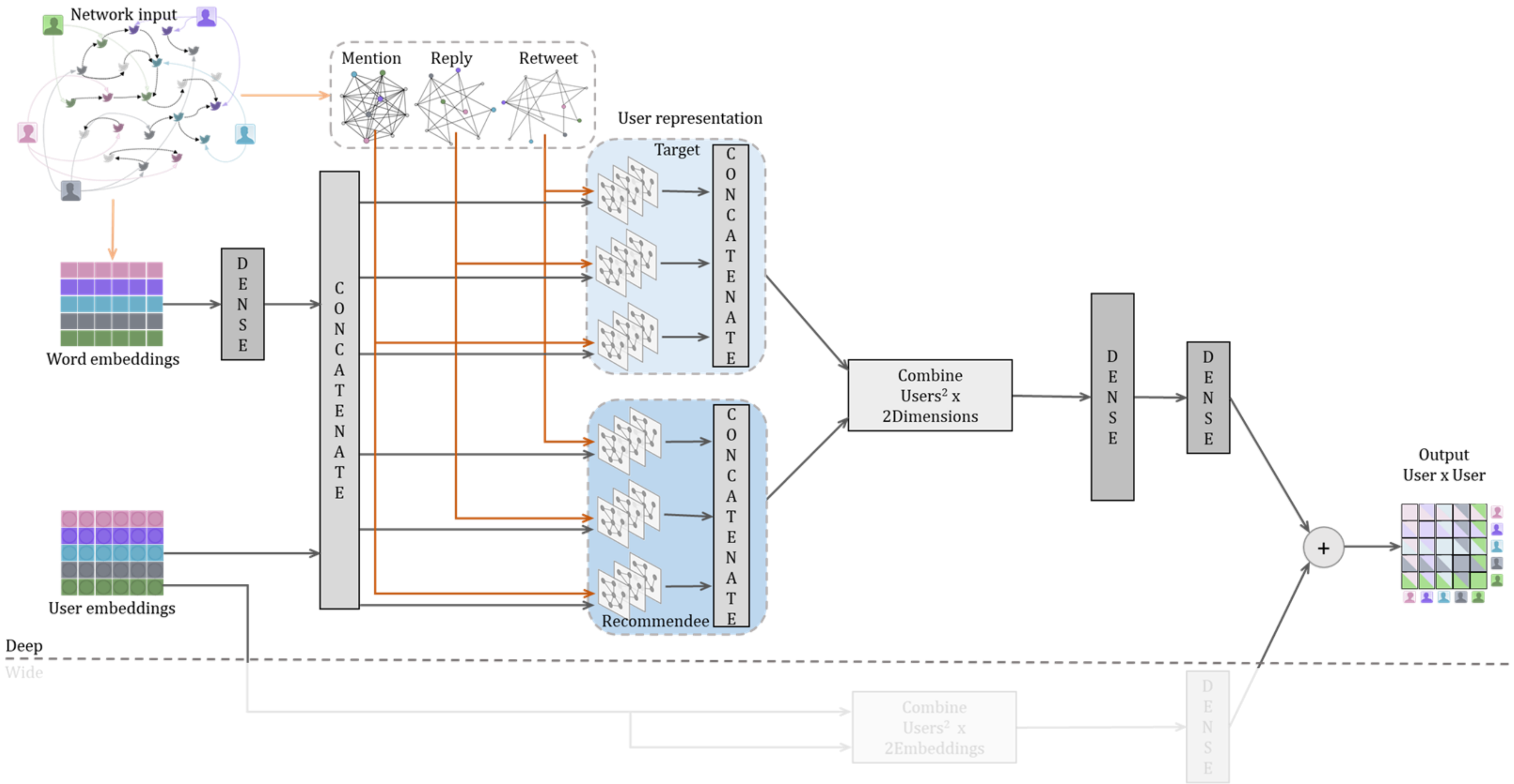
FRediECH

Friend REcommenDer for breakIng Echo CHambers



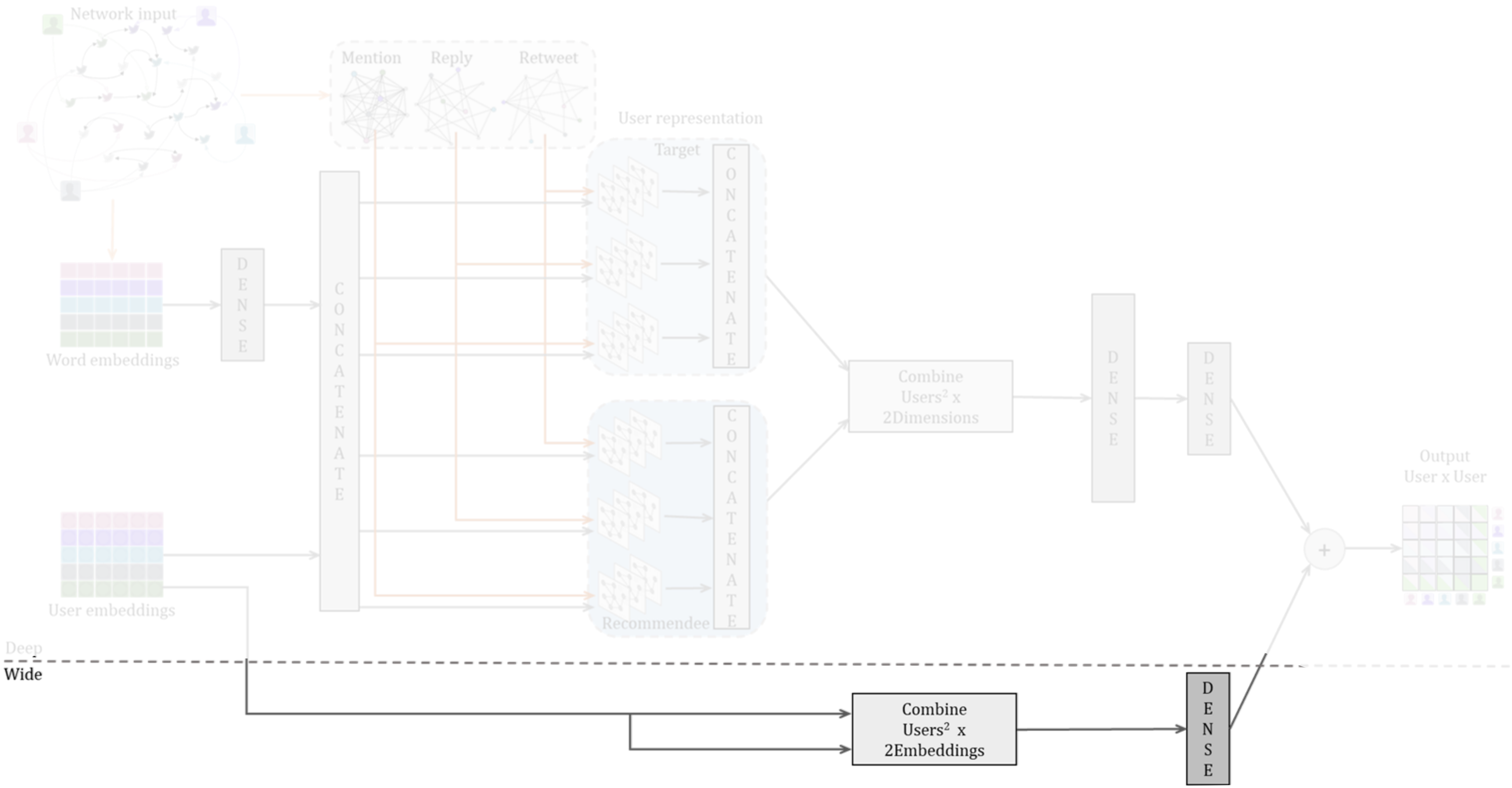
FRediECH

Friend REcommenDer for breakIng Echo CHambers



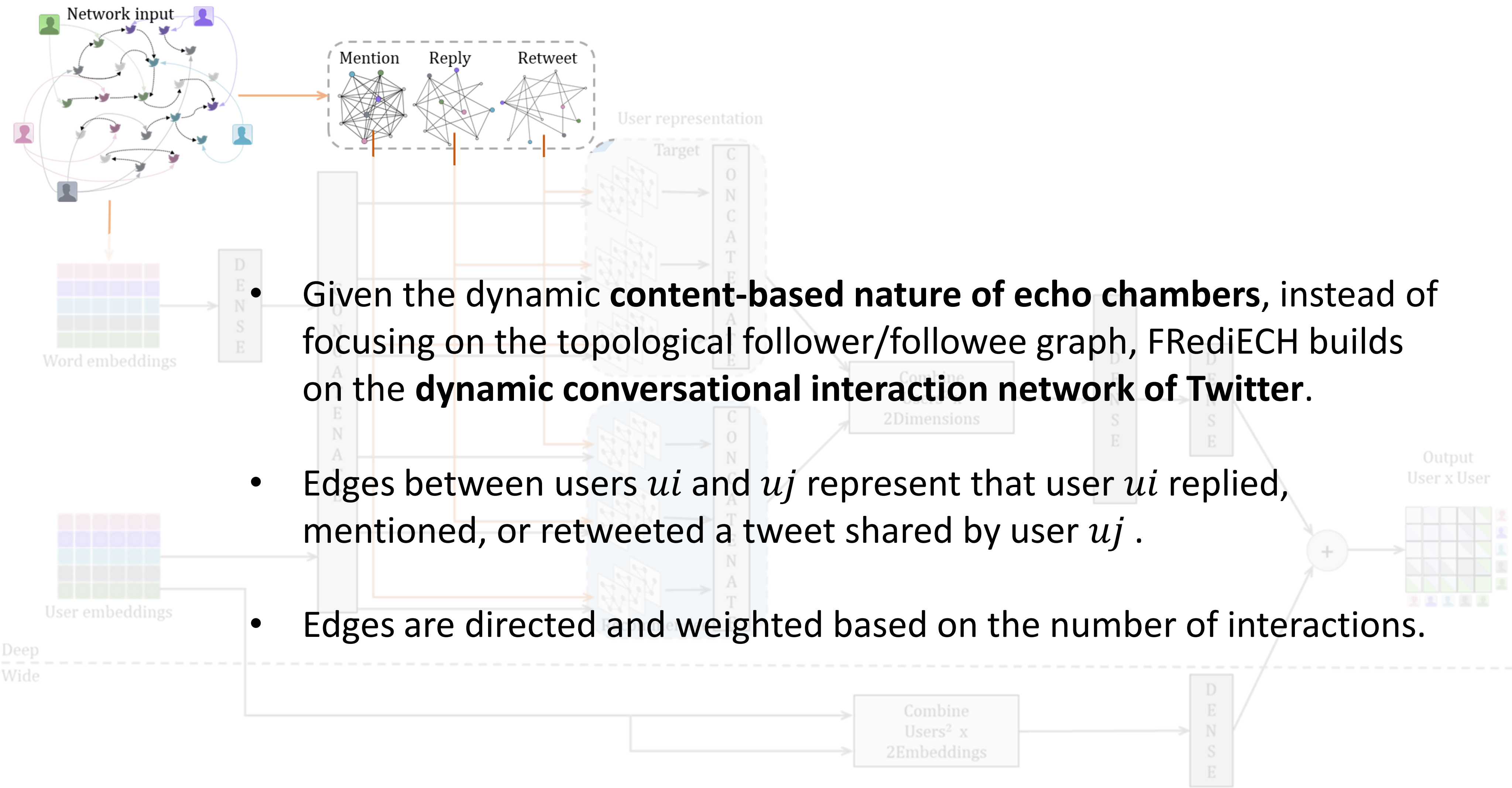
FRediECH

Friend REcommenDer for breakIng Echo CHambers



FRediECH

Friend REcommenDer for breakIng Echo CHambers



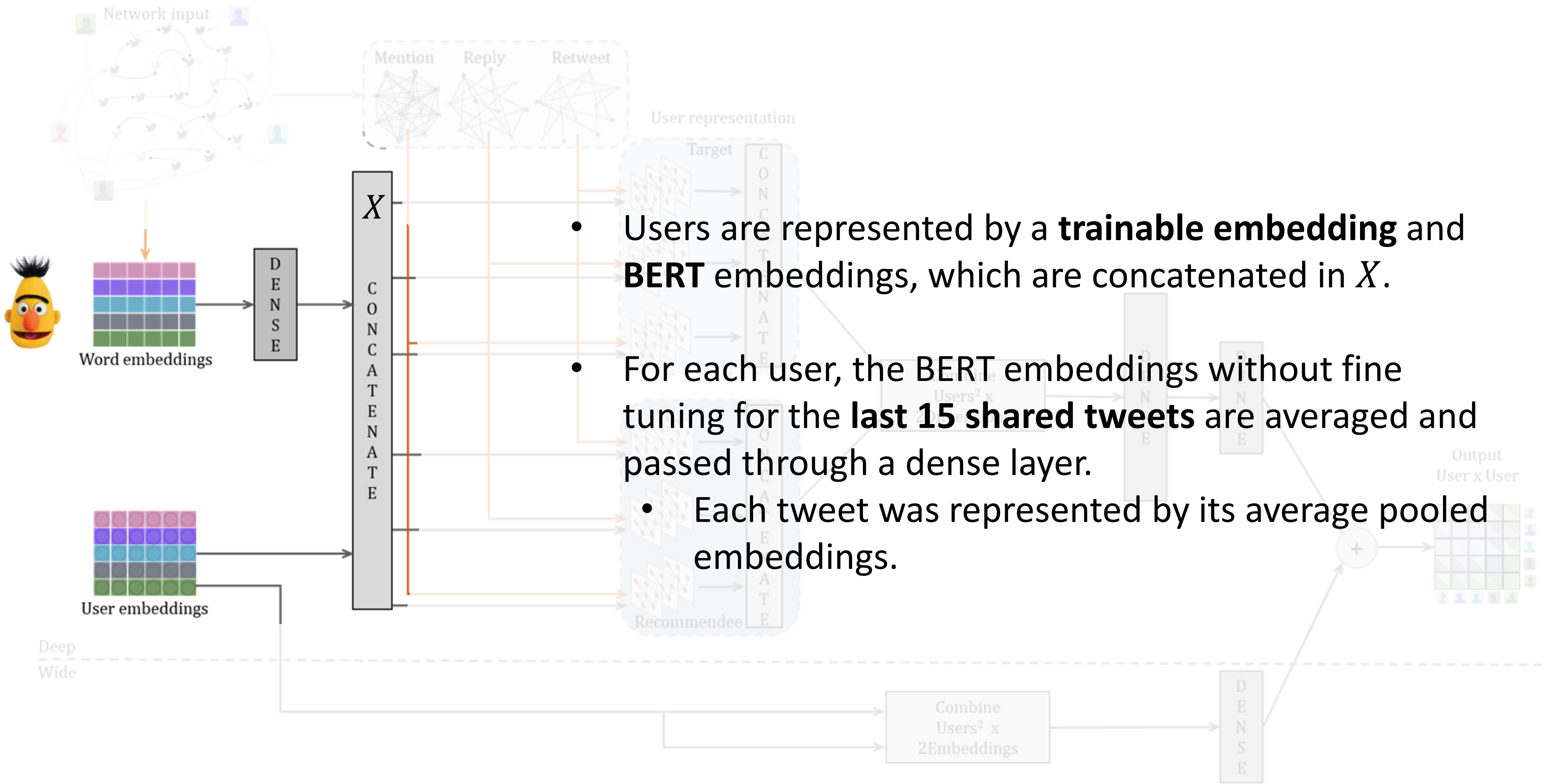
- Given the dynamic **content-based nature of echo chambers**, instead of focusing on the topological follower/followee graph, FRediECH builds on the **dynamic conversational interaction network of Twitter**.

- Edges between users ui and uj represent that user ui replied, mentioned, or retweeted a tweet shared by user uj .

- Edges are directed and weighted based on the number of interactions.

FRediECH

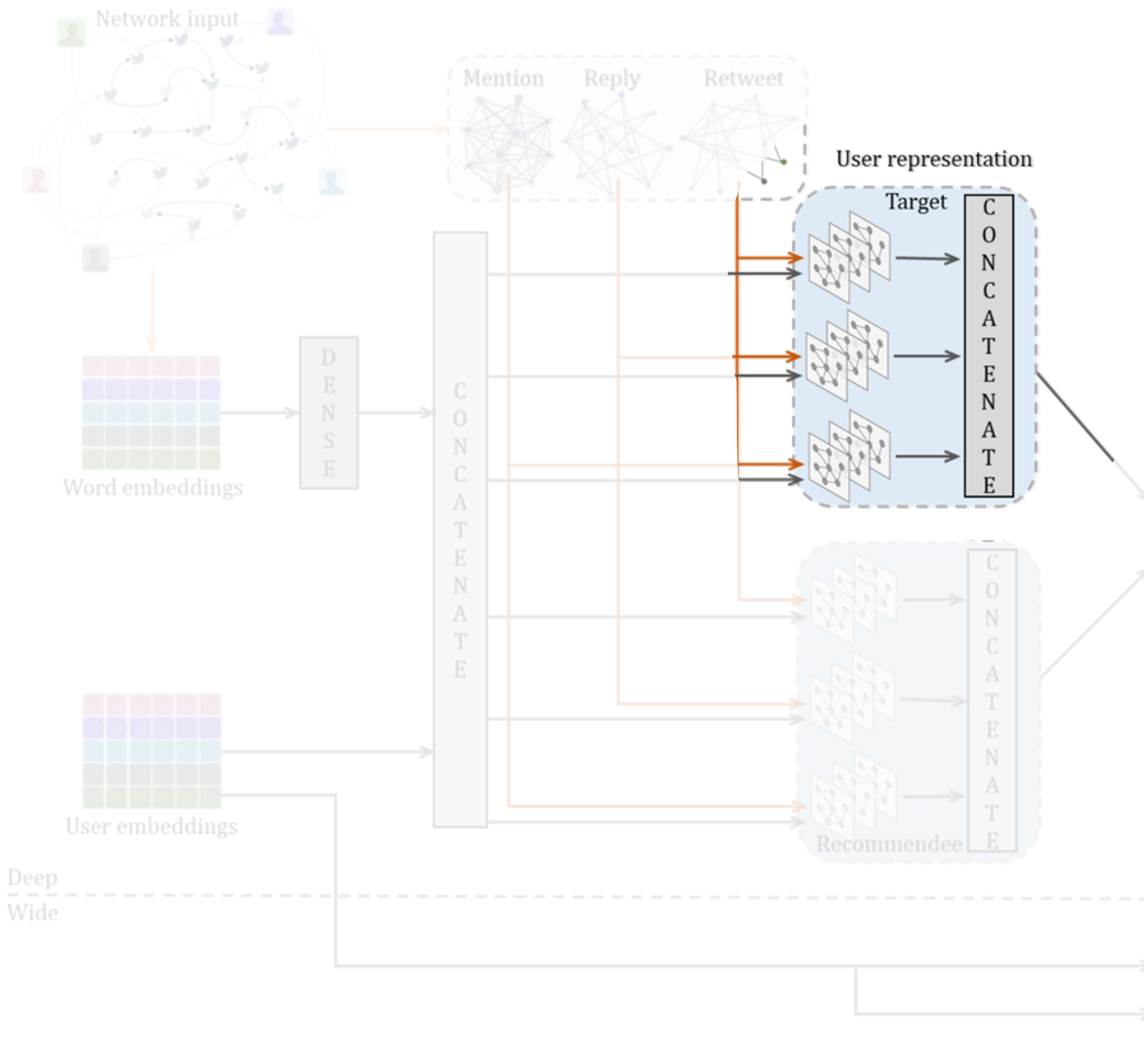
Friend REcommenDer for breakIng Echo CHambers



- Users are represented by a **trainable embedding** and **BERT** embeddings, which are concatenated in X.
- For each user, the BERT embeddings without fine tuning for the **last 15 shared tweets** are averaged and passed through a dense layer.
  - Each tweet was represented by its average pooled embeddings.

FRediECH

Friend REcommenDer for breakIng Echo CHambers



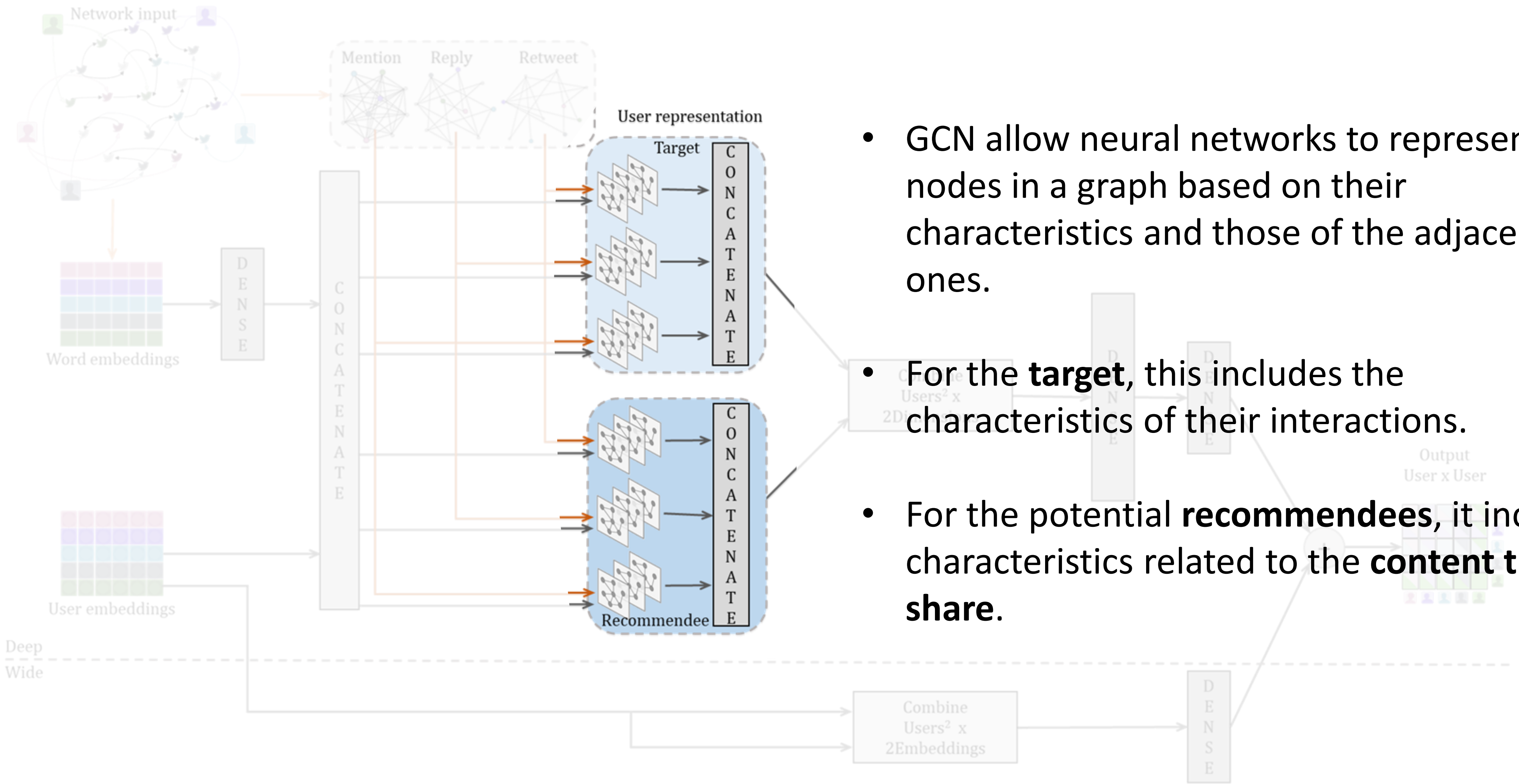
- To learn users’ latent characteristics based on their interactions, we defined **three parallels GCNs**.

- Each GCN allows learning the **specific weights** of each interaction type.

- GCNs’ outputs are **concatenated** to generate an intermediate user representation based on the combined interaction types.

FRediECH

Friend REcommenDer for breakIng Echo CHambers



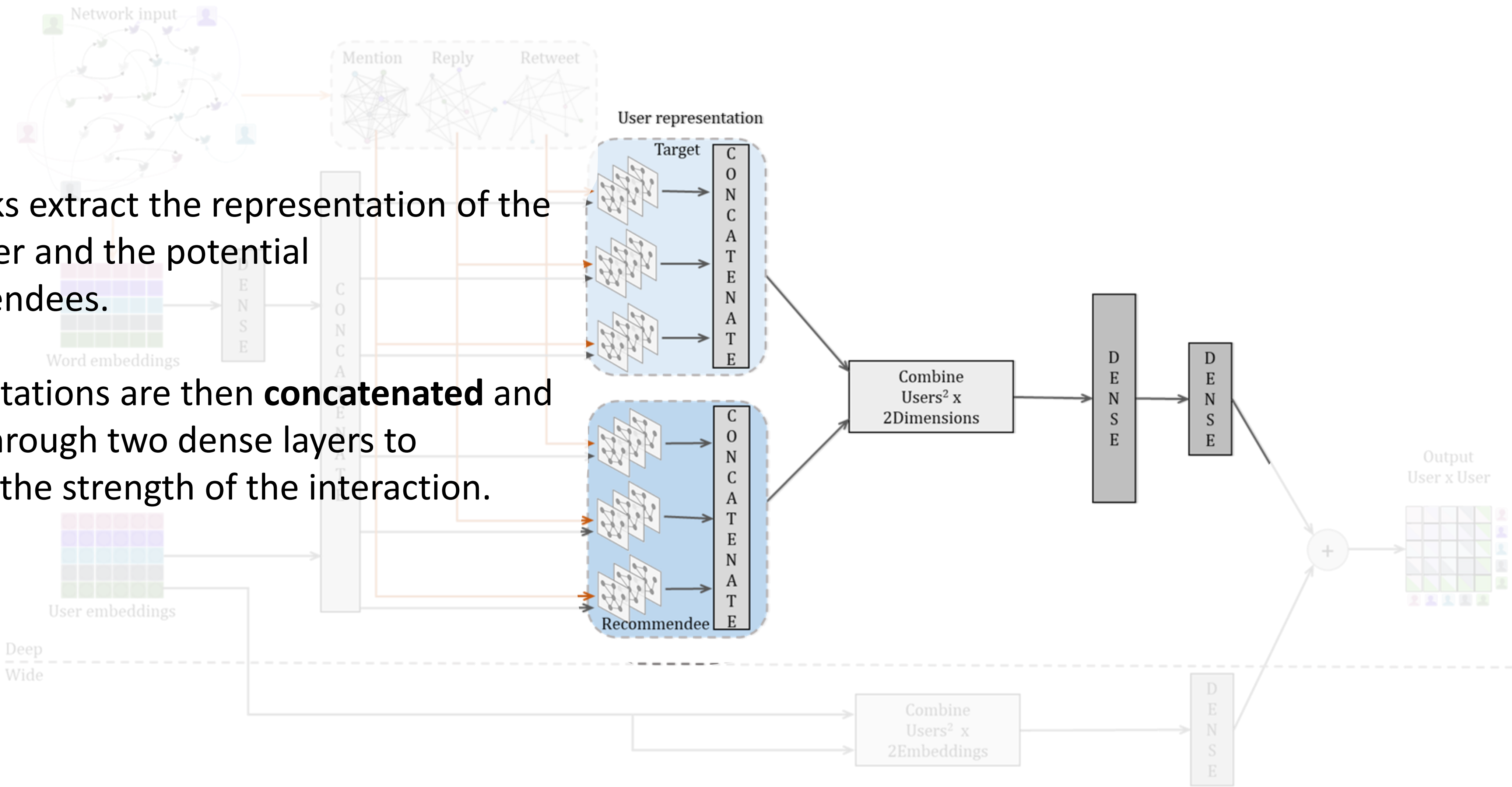
- GCN allow neural networks to represent nodes in a graph based on their characteristics and those of the adjacent ones.
- For the **target**, this includes the characteristics of their interactions.
- For the potential **recommendees**, it includes characteristics related to the **content they share**.

FRediECH

Friend REcommenDer for breakIng Echo CHambers

- The blocks extract the representation of the target user and the potential recommendees.

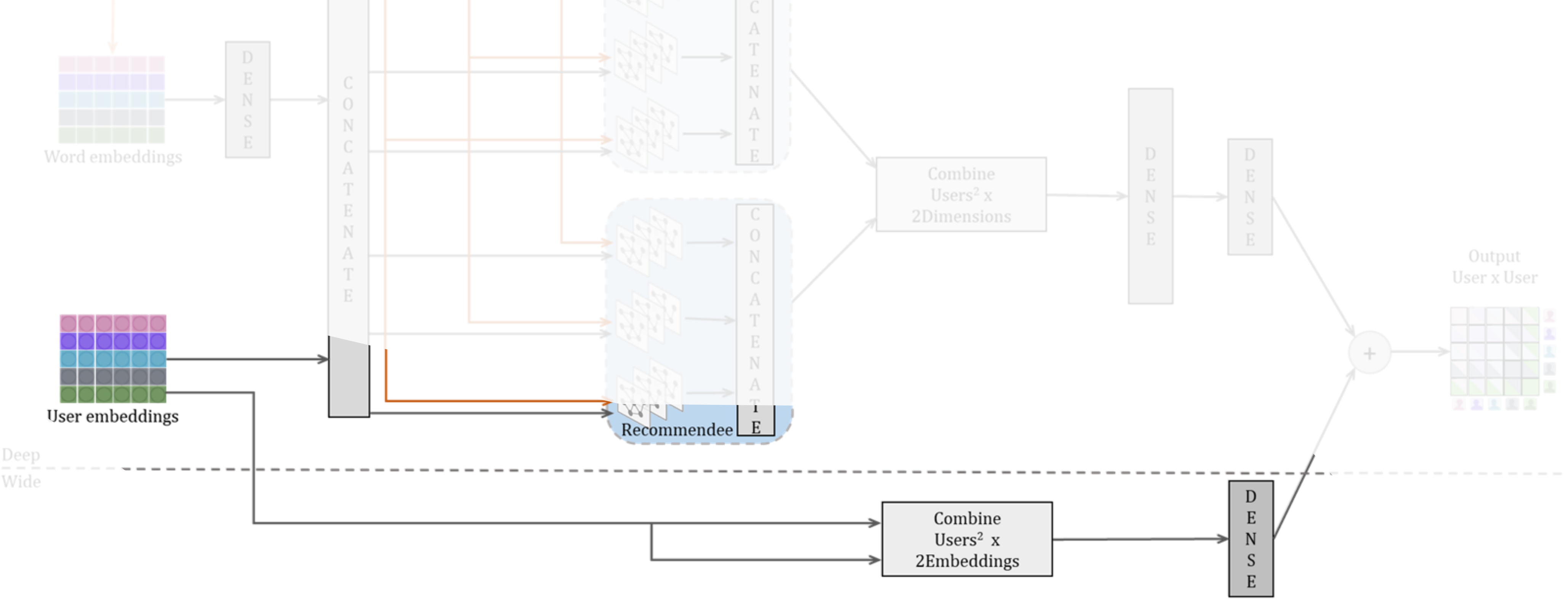
- Representations are then **concatenated** and passed through two dense layers to estimate the strength of the interaction.



FRediECH

Friend REcommenDer for breakIng Echo CHambers

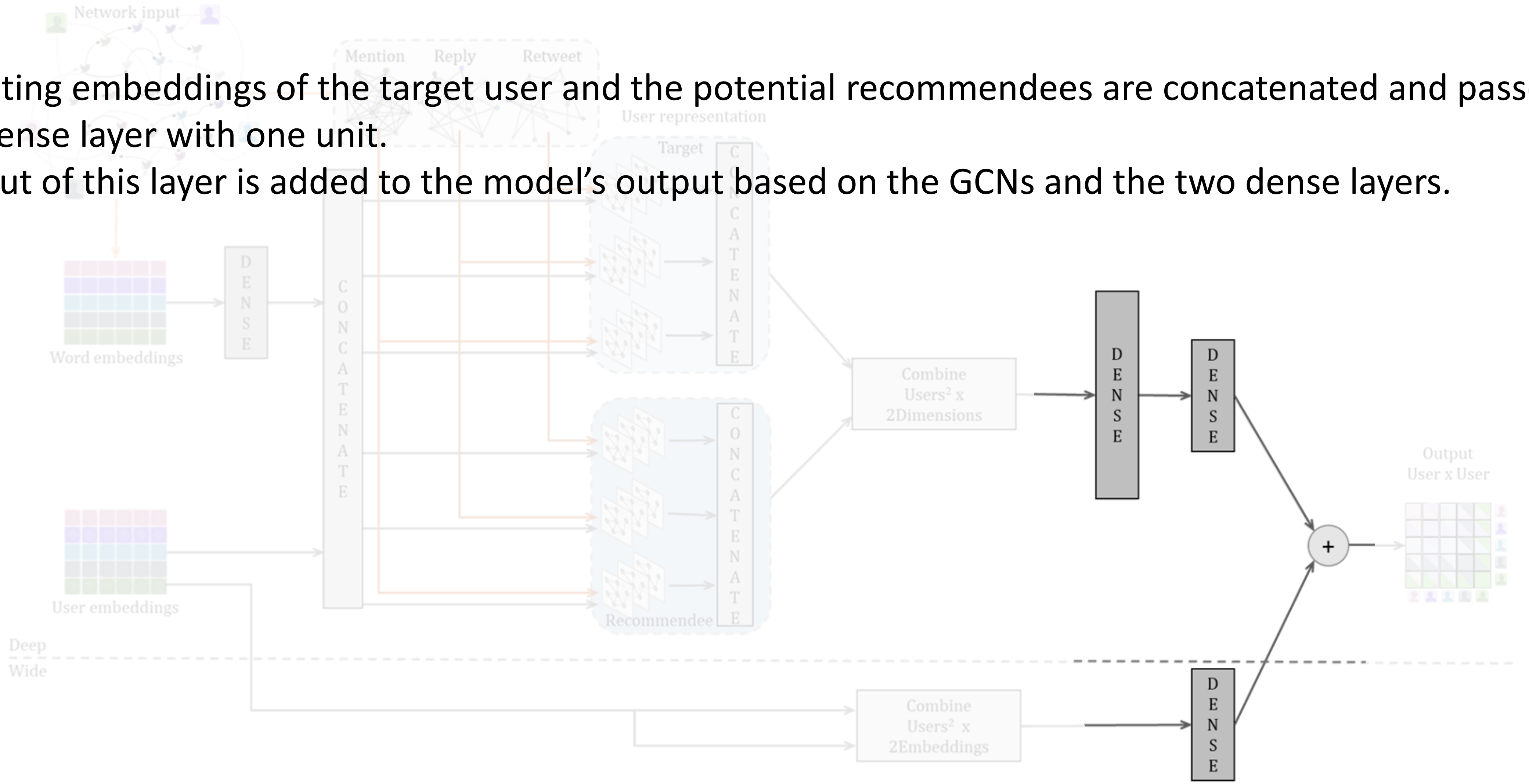
- The resulting embeddings of the target user and the potential recommendees are concatenated and passed through a dense layer with one unit.
- The output of this layer is added to the model's output based on the GCNs and the two dense layers.



FRediECH

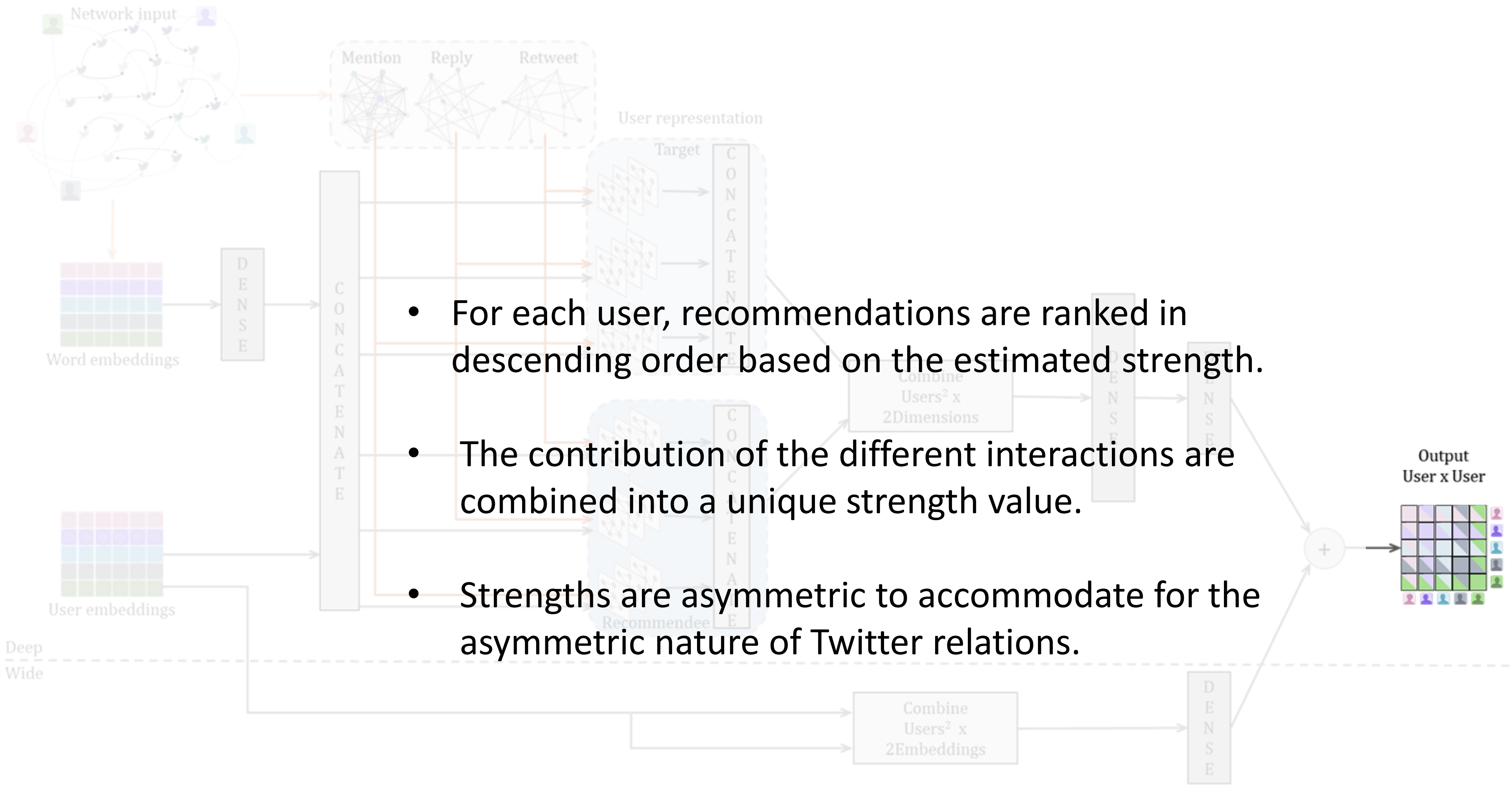
Friend REcommenDer for breakIng Echo CHambers

- The resulting embeddings of the target user and the potential recommendees are concatenated and passed through a dense layer with one unit.
- The output of this layer is added to the model's output based on the GCNs and the two dense layers.



FRediECH

Friend REcommenDer for breakIng Echo CHambers



- For each user, recommendations are ranked in descending order based on the estimated strength.
- The contribution of the different interactions are combined into a unique strength value.
- Strengths are asymmetric to accommodate for the asymmetric nature of Twitter relations.

FRediECH

Model training

- Interactions between users belonging to **different echo chambers carry a higher weight** than **interactions** between users in the **same echo chamber**.

- The goal is to favour the diversity of recommendations by **learning the structure of echo chambers** without explicitly finding them.
  - This allows for more freedom in the echo chamber definition and more sensitivity to changes in the network.

FRediECH

Model training

$$L(Y,\hat{Y}) = \frac{\sum_{i,j} d(u_i,u_j)\left(\widehat{Y_{ij}} - \log_2(2Y_{ij})\right)^2}{|E|}$$

- Interactions between users belonging to **different echo chambers carry a higher weight** than **interactions** between users in the **same echo chamber**.

- The goal is to favour the diversity of recommendations by **learning the structure of echo chambers** without explicitly finding them.
  - This allows for more freedom in the echo chamber definition and more sensitivity to changes in the network.

- We define a loss function based on the distance between users ($d\ (ui\ ,uj\ )$) and the number of interactions ($Yi\ j$).
  - The logarithm reduces the influence of users with many interactions.
  - The scalar prevents an interaction with a weight of 1 to become zero.
  - β allows tuning the preference of whether recommendations belong to the same group.

FRediECH

Model training

$$L(Y,\hat{Y}) = \frac{\sum_{i,j} d(u_i,u_j)\left(\widehat{Y_{ij}} - \log_2(2Y_{ij})\right)^2}{|E|}$$

- Interactions between users belonging to **different echo chambers carry a higher weight** than **interactions** between users in the **same echo chamber**.

- The goal is to favour the diversity of recommendations by **learning the structure of echo chambers** without explicitly finding them.
  - This allows for more freedom in the echo chamber definition and more sensitivity to changes in the network.

- We define a loss function based on the distance between users ($d\,(ui\,,uj\,)$) and the number of interactions ($Yi\,j\,$).
  - The logarithm reduces the influence of users with many interactions.
  - The scalar prevents an interaction with a weight of 1 to become zero.
  - β allows tuning the preference of whether recommendations belong to the same group.

- $d(u_i,u_j)$ is based on the cosine similarity over a new 10-dimensional embedding (e_i) representing users.
  - These embeddings were defined to capture the implicit community structure and were trained before the main model.
  - Users with similar interaction patterns will be represented by similar embeddings.
  - This loss function was based on GloVe.

Experimental evaluation

RQ1. How does FRediECH perform when compared with other tecniques?

RQ2. Hos do the key components of FRediECH contribute to the recommendations' performance?

Experimental evaluation

Data

- We used the obamacare data collection.
  - Tweets related to the `#obamacare` and `#aca` hashtags in Twitter.
  - Spans between May 2008 and October 2017.
  - It includes estimated user polarity.
- Tweets were retrieved using the Faking it! tool.
- We retrieved approximately 8 million public tweets belonging to 8,164 users, and 585,524 adjacent users.
- We kept **6,442 users** with at least one relation and that belonged to the largest connected component of the retrieved interaction graph.
  - This selection ensures that each user can be both source and destination of information content.

	Avg (± std)
#users	6,442
#tweets	7,016,552
Tweets per user	1089 (± 1413)
Relations per user	680 (± 1071)
Mentions per user	460 (± 733)
Replies per user	87 (± 190)
Retweets per user	399 (± 353)

Experimental evaluation

Data

- We used the obamacare data collection.
  - Tweets related to the `#obamacare` and `#aca` hashtags in Twitter.
  - Spans between May 2008 and October 2017.
  - It includes estimated user polarity.

- Tweets were retrieved using the **Faking it!** tool.

- We retrieved approximately 8 million public tweets belonging to 8,164 users, and 585,524 adjacent users.

- We kept **6,442 users** with at least one relation and that belonged to the largest connected component of the retrieved interaction graph.
  - This selection ensures that each user can be both source and destination of information content.

- We **verified** the existence of echo chambers.

	Avg (± std)
#users	6,442
#tweets	7,016,552
Tweets per user	1089 (± 1413)
Relations per user	680 (± 1071)
Mentions per user	460 (± 733)
Replies per user	87 (± 190)
Retweets per user	399 (± 353)

Experimental evaluation
Data

	Avg (± std)
#users	6,442
#tweets	7,016,552
Tweets per user	1089 (± 1413)
Relations per user	680 (± 1071)
Mentions per user	460 (± 733)
Replies per user	87 (± 190)
Retweets per user	399 (± 353)

- We used the obamacare data collection.
  - Tweets related to the `#obamacare` and `#aca` hashtags in Twitter.
  - Spans between May 2008 and October 2017.
  - It includes estimated user polarity.
- Tweets were retrieved using the **Faking it!** tool.
- We retrieved approximately 8 million public tweets belonging to 8,164 users, and 585,524 adjacent users.
- We kept **6,442 users** with at least one relation and that belonged to the largest connected component of the retrieved interaction graph.
  - This selection ensures that each user can be both source and destination of information content.
- We **verified** the existence of echo chambers.

Experimental evaluation
Data

- We used the obamacare data collection.
  - Tweets related to the `#obamacare` and `#aca` hashtags in Twitter.
  - Spans between May 2008 and October 2017.
  - It includes estimated user polarity.

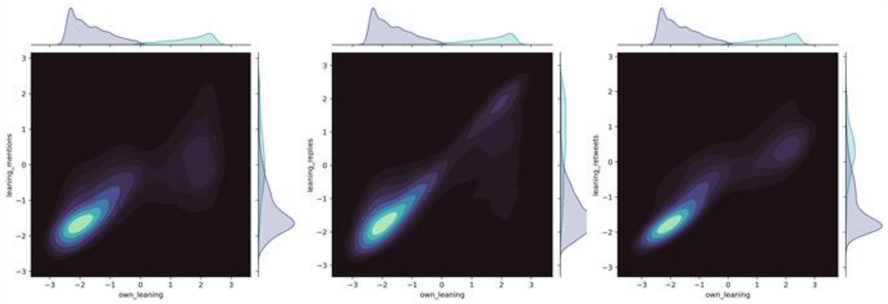
- Tweets were retrieved using the **Faking it!** tool.

- We retrieved approximately 8 million public tweets belonging to 8,164 users, and 585,524 adjacent users.

- We kept **6,442 users** with at least one relation and that belonged to the largest connected component of the retrieved interaction graph.
  - This selection ensures that each user can be both source and destination of information content.

- We **verified** the existence of echo chambers.

	Avg (± std)
#users	6,442
#tweets	7,016,552
Tweets per user	1089 (± 1413)
Relations per user	680 (± 1071)
Mentions per user	460 (± 733)
Replies per user	87 (± 190)
Retweets per user	399 (± 353)



Experimental evaluation
Baselines

Trivial, non-personalized and traditional recommenders.

Random | Popularity | Topology | Content

Adapted traditional and state-of-the-art user-item recommendation techniques.

ImplicitMF | NeuralCF | GraphRec | Diffnet | Mult-VAE

Techniques focused on enhancing the structural diversity of recommendations to mitigate filter bubbles.

SCC | CAM

Experimental evaluation

Evaluation

Relevance

- Precision@k
- Recall@k
- DCG@k

Diversity

- Variations of intra-list dissimilarities were used to assess:
  - **Diversity**, **novelty**.
  - Individuals and groups.
- Euclidean distance over structural and content-based representations.

- All evaluations were performed over the **same data partitions** and evaluated using the same set of metrics.
- We selected the top-10 recommended users (50% of users have 10 or more interactions).
- Recommendations were considered correct if they appeared in the test set.

- Training set: interactions before August 30 2017 (80% of all interactions)
- Test set: remaining interactions.

Evaluation results

	Precision	Recall	nDCG	Structural dissimilarities			
				Ind Diversity	Ind Novelty	Group Diversity	Group Novelty
FRediECH	0.152	0.183	**0.685**	<u>0.888</u>	**0.992**	**0.927**	**0.938**
Random	0.113**	0.053**	0.459**	0.732	0.699**	0.726	0.797
Popularity	**0.281**	0.22	**0.686**	0.369**	0.559**	0.391	0.673
Topology-based Adamic-Adar	<u>0.27</u>	**0.285**	0.632	0.359**	0.431**	0.517	0.653
Topology-based Jaccard	0.191	0.249	0.567	0.364**	0.453**	0.592	0.667
Topology-based RA	<u>0.272</u>	<u>0.27</u>	0.642	0.367**	0.436	0.573	0.643
Topology-based CN	0.259	0.302	0.619	0.356**	0.424**	0.564	0.633
Content-based Full Tweets	0.115**	0.053**	0.439**	0.726	0.698**	0.727	0.797
Content-based 15 Tweets	0.246	0.22	0.584	0.428**	0.491**	0.629**	0.69
SCC	0.259	0.252	0.597	0.35**	0.496**	0.469	0.621
CAM	0.228	0.158	0.513	0.345**	0.424**	0.53	0.647
Implicit	<u>0.271</u>	0.252	<u>0.654</u>	0.401**	0.435**	0.559	0.643
NeuralCF	0.251	0.262	0.579	0.351**	0.419**	0.566	0.647
GraphRec	0.103**	0.183**	0.389**	**0.935**	<u>0.842</u>**	<u>0.739</u>	<u>0.895</u>
Mult-VAE	0.26	0.254	0.627	0.413**	0.433**	0.607	0.637
Original graph	-	-	-	0.325	0.418	0.581	0.603

** indicates statistically significant differences favouring FRediECH

Evaluation results

	Precision	Recall	nDCG	Structural dissimilarities			
				Ind Diversity	Ind Novelty	Group Diversity	Group Novelty
FRediECH	0.152	0.183	**0.685**	0.888	**0.992**	**0.927**	**0.938**
Random	0.113**	0.053**	0.459**	0.732	0.699**	0.726	0.797
Popularity	**0.281**	0.22	**0.686**	0.369**	0.559**	0.391	0.673

- Random recommender achieved high diversity and novelty results, but recommendations were less relevant.
- There is a trade-off between the relevance, and diversity and novelty.
- Techniques achieving high relevance also achieved low diversity and novelty scores.
- Statistically significant differences were observed regarding diversity and novelty when compared to all techniques but GraphRec and FRediECH.

** indicates statistically significant differences favouring FRediECH

Evaluation results

	Precision	Recall	nDCG	Structural dissimilarities			
				Ind Diversity	Ind Novelty	Group Diversity	Group Novelty
FRediECH	0.152	0.183	**0.685**	0.888	**0.992**	**0.927**	**0.938**
Topology-based Adamic-Adar	0.27	**0.285**	0.632	0.359**	0.431**	0.517	0.653
Topology-based Jaccard	0.191	0.249	0.567	0.364**	0.453**	0.592	0.667
Topology-based RA	0.272	0.27	0.642	0.367**	0.436	0.573	0.643
Topology-based CN	0.259	0.302	0.619	0.356**	0.424**	0.564	0.633
Content-based Full Tweets	0.115**	0.053**	0.439**	0.726	0.698**	0.727	0.797
Content-based 15 Tweets	0.246	0.22	0.584	0.428**	0.491**	0.629**	0.69

- Topological baselines achieved high precision and low diversity, which is expected as recommendations are based on user neighbourhood.
- Diversity and novelty differences were significant and favoured FRediECH.

- While considering the full tweet set increased the diversity of recommendations, using only the last 15 increased their relevance.
    - These observations could relate to the broad period covered by the data collection, in which conversation topics (and user interests) could have shifted.

Evaluation results

	Precision	Recall	nDCG	Structural dissimilarities			
				Ind Diversity	Ind Novelty	Group Diversity	Group Novelty
FRediECH	0.152	0.183	**0.685**	<u>0.888</u>	**0.992**	**0.927**	**0.938**
GraphRec	0.103**	0.183**	0.389**	**0.935**	<u>0.842</u>**	<u>0.739</u>	<u>0.895</u>
Original graph	-	-	-	0.325	0.418	0.581	0.603

- FRediECH achieved the highest diversity and novelty results, followed by GraphRec.
  - For individual diversity, in which GraphRec outperformed FRediECH.
- In terms of relevance, FRediECH also significantly outperformed GraphRec.
- Most of the differences favouring FRediECH were statistically significant.
- Despite lower precision and recall that other techniques, nDCG results showed that even when recommending non relevant users, the relevant ones were ranked high.

** indicates statistically significant differences favouring FRediECH

Evaluation results

RQ1. Comparison with state-of-the-art techniques

	Precision	Recall	nDCG	Structural dissimilarities			
				Ind Diversity	Ind Novelty	Group Diversity	Group Novelty
FRediECH	0.152	0.183	**0.685**	0.888	**0.992**	**0.927**	**0.938**
SCC	0.259	0.252	0.597	0.35**	0.496**	0.469	0.621
CAM	0.228	0.158	0.513	0.345**	0.424**	0.53	0.647
Original graph	-	-	-	0.325	0.418	0.581	0.603

- SCC achieved higher relevance and structural novelty.
- CAM achieved higher content diversity.

- Diversity and novelty of both techniques were close to those of the original network, thus failing to significantly improve the quality of the network.

** indicates statistically significant differences favouring FRediECH

Evaluation results

	Precision	Recall	nDCG	Structural dissimilarities			
				Ind Diversity	Ind Novelty	Group Diversity	Group Novelty
FRediECH	0.152	0.183	**0.685**	<u>0.888</u>	**0.992**	**0.927**	**0.938**
Random	0.113**	0.053**	0.459**	0.732	0.699**	0.726	0.797
Popularity	**0.281**	0.22	**0.686**	0.369**	0.559**	0.391	0.673
Topology-based Adamic-Adar	<u>0.27</u>	**0.285**	0.632	0.359**	0.431**	0.517	0.653
Topology							0.667
Topology							0.643
Topology							0.633
Content-							0.797
Content-based 15 Tweets	0.246	0.22	0.584	0.428**	0.491**	0.629**	0.69
GraphRec	0.103**	0.183**	0.389**	**0.935**	<u>0.842**</u>	<u>0.739</u>	<u>0.895</u>
Mult-VAE	0.26	0.254	0.627	0.413**	0.433**	0.607	0.637
Original graph	-	-	-	0.325	0.418	0.581	0.603

	Traditional	State-of-the-art	Original structure
Avg. Improvements	47%	44%	
Maximum	60% (individual novelty)		67% (individual novelty)

- In general, the novelty of recommendations was higher than their diversity.
- Novelty was higher for the structural distance, which implies that recommended users belong to other communities, but still shared similar content.

** indicates statistically significant differences favouring FRediECH

Evaluation results

	Precision	Recall	nDCG	Structural dissimilarities			
				Ind Diversity	Ind Novelty	Group Diversity	Group Novelty
FRediECH	0.152	0.183	**0.685**	<u>0.888</u>	**0.992**	**0.927**	**0.938**
Random	0.113**	0.053**	0.459**	0.732	0.699**	0.726	0.797
Popularity	**0.281**	0.22	**0.686**	0.369**	0.559**	0.391	0.673
Topology-based Adamic-Adar	<u>0.27</u>	**0.285**	0.632	0.359**	0.431**	0.517	0.653
Topology-based Jaccard	0.191	0.249	0.567	0.364**	0.453**	0.592	0.667
Topology-based RA	<u>0.272</u>	<u>0.27</u>	0.642	0.367**	0.436	0.573	0.643
Topology-based CN	0.259	0.302	0.619	0.356**	0.424**	0.564	0.633
Content[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Content[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
SCC	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
CAM	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Implicit	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
NeuralCF	0.251	0.262	0.579	0.351**	0.419**	0.566	0.647
GraphRec	0.103**	0.183**	0.389**	**0.935**	<u>0.842</u>**	<u>0.739</u>	<u>0.895</u>
Mult-VAE	0.26	0.254	0.627	0.413**	0.433**	0.607	0.637
Original graph	-	-	-	0.325	0.418	0.581	0.603

Results showed that FRediECH (despite the trade-off with precision) satisfactorily increased the diversity and novelty of recommendations, when measured in terms of individual users and the communities they belong to.

** indicates statistically significant differences favouring FRediECH

Evaluation results

RQ2. Ablation Study

$\texttt{FRediECH}_{\texttt{NO-NS}}$	Remove the negative sampling from the described model.
$\texttt{FRediECH}_{\texttt{NO-WIDE}}$	Remove the wide component of the architecture.
$\texttt{FRediECH}_{\texttt{NO-WIDE-NO-NS}}$	Remove the wide component of the architecture and the negative sampling.
$\texttt{FRediECH}_{\texttt{DUAL}}$	Different embeddings are used for representing the target and recommended users, which are processed by different GCNs.
$\texttt{FRediECH}_{\texttt{NO-BERT}}$	Remove the textual embeddings from the described model.
$\texttt{FRediECH}_{\texttt{MENTION}}$ $\texttt{FRediECH}_{\texttt{REPLY}}$ $\texttt{FRediECH}_{\texttt{RETWEET}}$	Only one interaction type is considered.
$\texttt{FRediECH}_{\texttt{MENTION-REPLY}}$ $\texttt{FRediECH}_{\texttt{MENTION-RETWEET}}$ $\texttt{FRediECH}_{\texttt{REPLY-RETWEET}}$	The described model includes pairs of interactions.

- Relations were removed from both the training and test sets.
- A new model was trained from scratch for each evaluation.

Evaluation results

RQ2. Ablation Study

	Precision	Recall	nDCG	Structural dissimilarities			
				Ind Diversity	Ind Novelty	Group Diversity	Group Novelty
FRediECH	0.152	0.183	0.685	**0.888**	**0.992**	**0.927**	0.938
$\text{FRediECH}_{\text{NO-NS}}$	0.149**	0.172	0.553**	0.726	0.82**	0.845	0.852
$\text{FRediECH}_{\text{NO-WIDE}}$	0.152	0.189	0.685	0.888	**0.993**	0.845	0.966
$\text{FRediECH}_{\text{NO-WIDE-NO-NS}}$	0.134	0.172	0.609	0.597**	0.82**	0.728	0.852
$\text{FRediECH}_{\text{DUAL}}$	**0.169**	0.192	0.561	0.73	0.937**	0.762	0.912
$\text{FRediECH}_{\text{NO-BERT}}$	**0.16**	0.193	0.56	0.596**	0.97**	0.708	0.936
$\text{FRediECH}_{\text{MENTION}}$	0.14	0.182	0.544	0.541**	**0.993**	0.698	0.93
$\text{FRediECH}_{\text{REPLY}}$	0.103**	**0.203**	**0.732**	0.509**	**0.99**	0.643	**0.99**
$\text{FRediECH}_{\text{RETWEET}}$	0.146	0.193	0.567	0.646**	**0.99**	0.724	0.941
$\text{FRediECH}_{\text{MENTION-REPLY}}$	0.136	0.176	0.547	0.651	**0.99**	0.741	0.932
$\text{FRediECH}_{\text{MENTION-RETWEET}}$	0.159	0.184	0.542	0.627**	0.96	0.732	0.916
$\text{FRediECH}_{\text{REPLY-RETWEET}}$	**0.162**	0.183	0.55	0.69	0.947**	0.762	0.909

Evaluation results

RQ2. Ablation Study

	Precision	Recall	nDCG	Structural dissimilarities			
				Ind Diversity	Ind Novelty	Group Diversity	Group Novelty
FRediECH	0.152	0.183	0.685	**0.888**	**0.992**	**0.927**	0.938
$\text{FRediECH}_{\text{NO-NS}}$	0.149**	0.172	0.553**	0.726	0.82**	0.845	0.852
$\text{FRediECH}_{\text{NO-WIDE}}$	0.152	0.189	0.685	0.888	**0.993**	0.845	0.966
$\text{FRediECH}_{\text{NO-WIDE-NO-NS}}$	0.134	0.172	0.609	0.597**	0.82**	0.728	0.852
$\text{FRediECH}_{\text{DUAL}}$	**0.169**	0.192	0.561	0.73	0.937**	0.762	0.912
$\text{FRediECH}_{\text{NO-BERT}}$	**0.16**	0.193	0.56	0.596**	0.97**	0.708	0.936
$\text{FRediECH}_{\text{MENTION}}$	0.14	0.182	0.544	0.541**	**0.993**	0.698	0.93
$\text{FRediECH}_{\text{REPLY}}$	0.103					0.643	**0.99**
$\text{FRediECH}_{\text{RETWEET}}$	0.14					0.724	0.941
$\text{FRediECH}_{\text{MENTION-REPLY}}$	0.13					0.741	0.932
$\text{FRediECH}_{\text{MENTION-RETWEET}}$	0.15					0.732	0.916
$\text{FRediECH}_{\text{REPLY-RETWEET}}$	**0.16**					0.762	0.909

Architectural changes

- Relevance was not greatly affected.
- Diversity/novelty showed more variability.
- Differences favouring the original FRediECH were statistically significant.

Evaluation results

RQ2. Ablation Study

	Precision	Recall	nDCG	Ind Di			
FRediECH	0.152	0.183	0.685	**0.**			
FRediECH$_{\text{NO-NS}}$	0.149**	0.172	0.553**	0.			
FRediECH$_{\text{NO-WIDE}}$	0.152	0.189	0.685	0.			
FRediECH$_{\text{NO-WIDE-NO-NS}}$	0.134	0.172	0.609	0.597**	0.82**	0.728	0.852
FRediECH$_{\text{DUAL}}$	**0.169**	0.192	0.561	0.73	0.937**	0.762	0.912
FRediECH$_{\text{NO-BERT}}$	**0.16**	0.193	0.56	0.596**	0.97**	0.708	0.936
FRediECH$_{\text{MENTION}}$	0.14	0.182	0.544	0.541**	**0.993**	0.698	0.93
FRediECH$_{\text{REPLY}}$	0.103**	**0.203**	**0.732**	0.509**	**0.99**	0.643	**0.99**
FRediECH$_{\text{RETWEET}}$	0.146	0.193	0.567	0.646**	**0.99**	0.724	0.941
FRediECH$_{\text{MENTION-REPLY}}$	0.136	0.176	0.547	0.651	**0.99**	0.741	0.932
FRediECH$_{\text{MENTION-RETWEET}}$	0.159	0.184	0.542	0.627**	0.96	0.732	0.916
FRediECH$_{\text{REPLY-RETWEET}}$	**0.162**	0.183	0.55	0.69	0.947**	0.762	0.909

Data available to the model

- Including content allowed to significantly increase the novelty and diversity of recommendations.
- Only considering one interaction significantly decreased diversity and novelty.
- Interactions might carry different weights.

Evaluation results

RQ2. Ablation Study

	Precision	Recall	nDCG	Structural dissimilarities			
				Ind Diversity	Ind Novelty	Group Diversity	Group Novelty
FRediECH	0.152	0.183	0.685	**0.888**	**0.992**	**0.927**	0.938
FRediECH$_{NO\text{-}NS}$	0.149**	0.172	0.553**	0.726	0.82**	0.845	0.852
FRediECH$_{NO\text{-}WIDE}$	0.152	0.189	0.685	0.888	**0.993**	0.845	0.966
FRediECH$_{NO\text{-}WIDE\text{-}NO\text{-}NS}$	0.134	0.172	0.609	0.597**	0.82**	0.728	0.852
FRediECH$_{DUAL}$	**0.169**	0.192	0.561	0.73	0.937**	0.762	0.912
FRediECH	**0.16**	0.193	0.56	0.596**	0.97**	0.708	0.936
FRe[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
FRe[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
FRediECH$_{RETWEET}$	0.146	0.193	0.567	0.646	**0.99**	0.724	0.941
FRediECH$_{MENTION\text{-}REPLY}$	0.136	0.176	0.547	0.651	**0.99**	0.741	0.932
FRediECH$_{MENTION\text{-}RETWEET}$	0.159	0.184	0.542	0.627**	0.96	0.732	0.916
FRediECH$_{REPLY\text{-}RETWEET}$	**0.162**	0.183	0.55	0.69	0.947**	0.762	0.909

In summary, results showed that each component of FRediECH significantly contributed to its performance.

Summary (II)

We developed `FRediECH` inspired by a **graph convolutional network** and a **Deep & Wide** architecture, coupling echo chamber awareness and user representations to balance the <u>relevance</u>, <u>diversity</u> and <u>novelty</u> of friend recommendations.

`FRediECH` allows recommending **users who are different among them and from the already known ones**, thus effectively **helping to reduce the echo chamber effect**.

`FRediECH` produced **similarly relevant recommendations** to those of the selected baselines while **increasing their diversity and novelty**.

What are we doing about it?

How misinformation spreads?

How to avoid misinformation derived effects?

How to identify users propagating misinformation?

Antonela Tommasel, Juan Manuel Rodriguez, Filippo Menczer. “Following the trail of fake news spreaders in social media: A deep learning model”. In 30th ACM Conference on User Modeling, Adaptation and Personalization. Barcelona, España. DOI: 10.1145/3511047.3536410

Identifying fake news spreaders

Users play a fundamental role as **creators and disseminators** of fake content.

It is **essential to detect both fake content and the users spreading it**, as the latter will provide **valuable information** for the design of **mitigation or intervention strategies** to rapidly contain the spreading.

Identifying fake news spreaders

Users play a fundamental role as **creators and disseminators** of fake content.

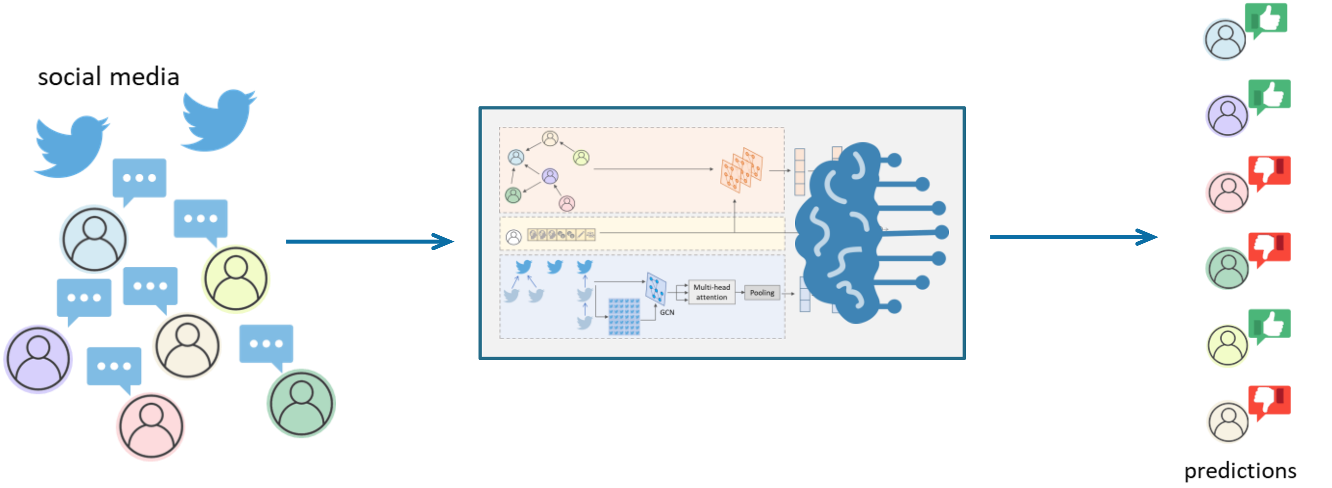
It is **essential to detect both fake content and the users spreading it**, as the latter will provide **valuable information** for the design of **mitigation or intervention strategies** to rapidly contain the spreading.

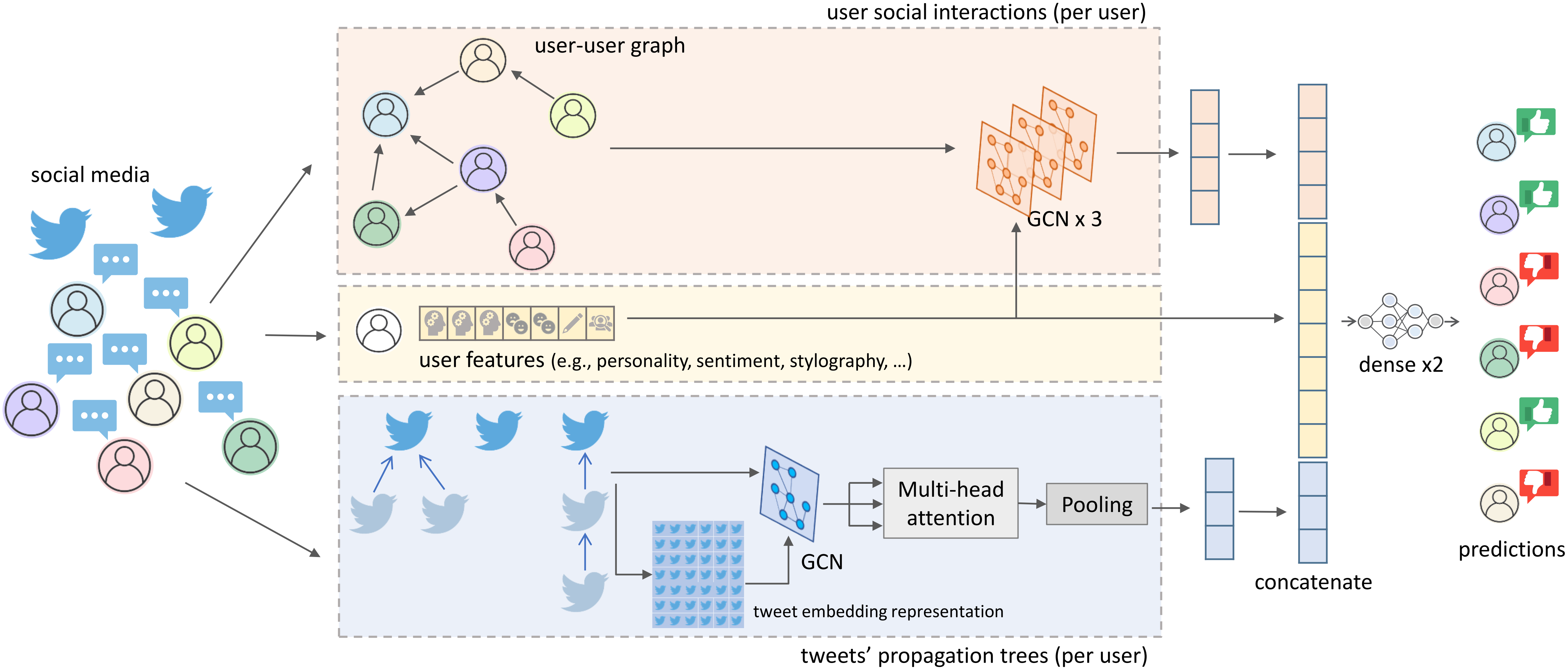
How can we effectively detect fake news spreaders in social media?

Identifying fake news spreaders

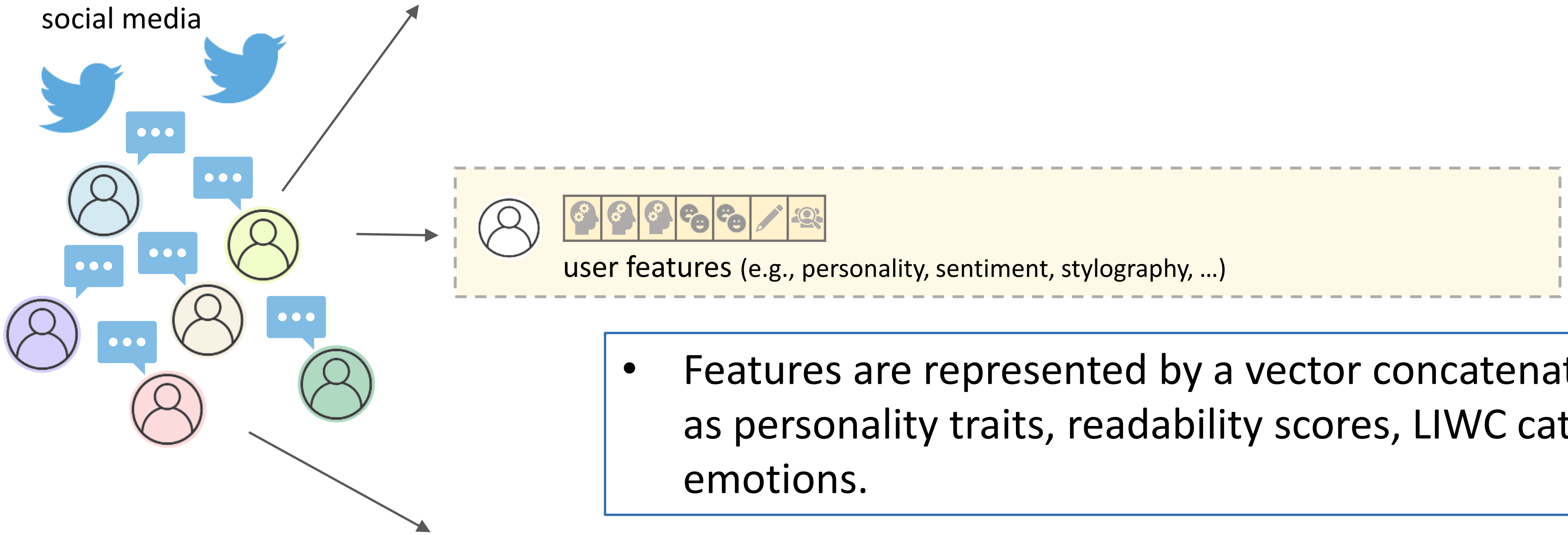
How can we effectively detect fake news spreaders in social media?

For a given user u_i and their **social interactions**, the **shared content** and the **content propagation trees**, the goal is to learn a function $F \rightarrow \{1, -1\}$, such that 1 indicates that u_i is a fake news spreaders, and -1 otherwise.



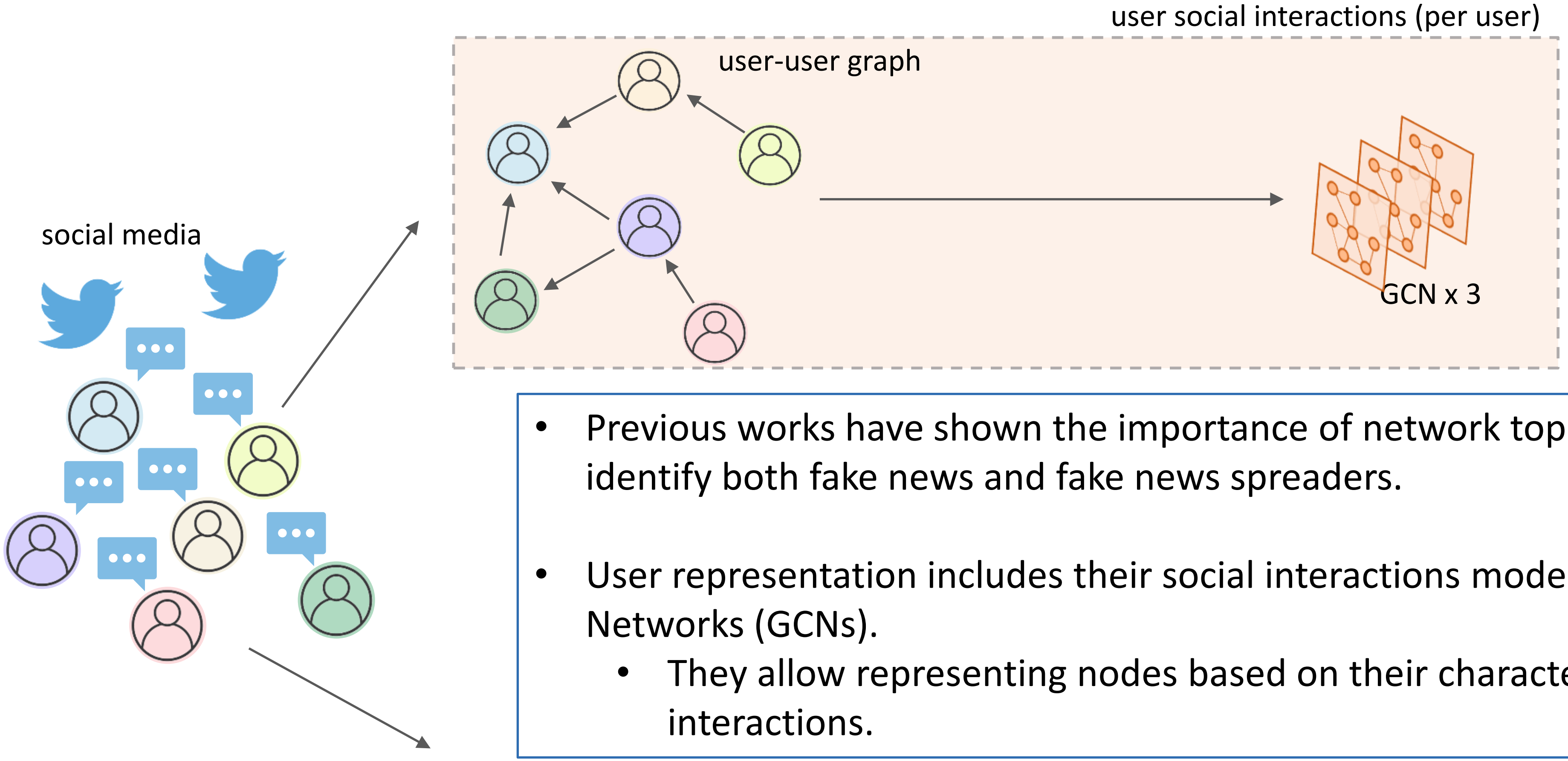

user social interactions (per user)
user-user graph
social media
GCN x 3
user features (e.g., personality, sentiment, stylography, …)
dense x2
Multi-head attention
Pooling
GCN
tweet embedding representation
concatenate
predictions
tweets’ propagation trees (per user)

Model overview



- Features are represented by a vector concatenating characteristics such as personality traits, readability scores, LIWC categories, sentiment and emotions.

Model overview



- Previous works have shown the importance of network topology and interactions to identify both fake news and fake news spreaders.

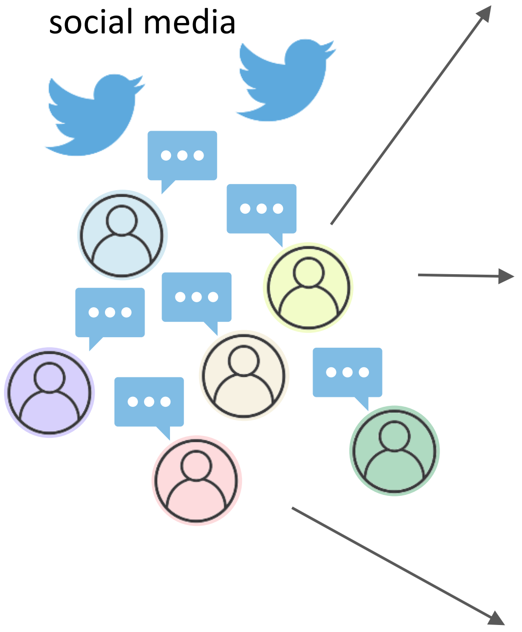
- User representation includes their social interactions modeled by Graph Convolutional Networks (GCNs).
  - They allow representing nodes based on their characteristics and those of their interactions.

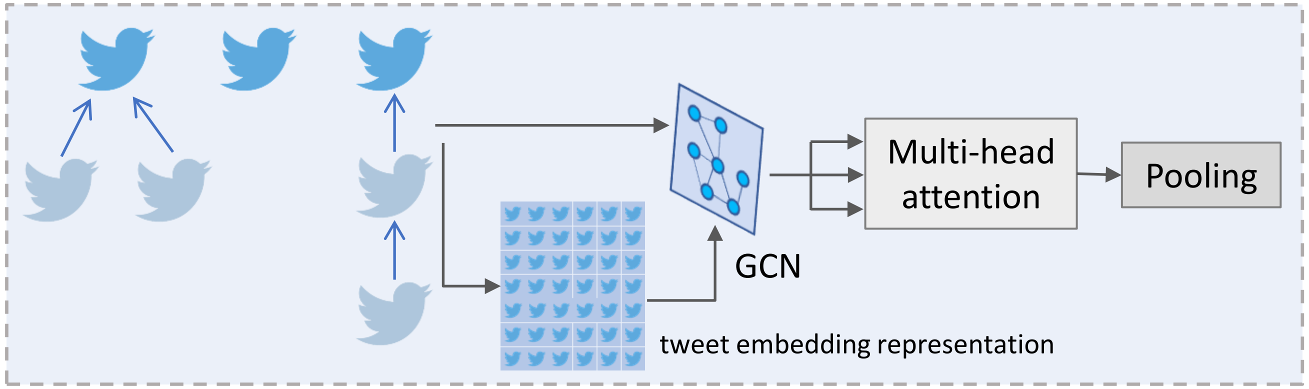
Model overview

- To capture the semantics of propagation patterns, we represent each shared tweet based on a **propagation tree derived from the triggered replies**.

- Each tweet is represented by its propagation tree, its pooled BERT embeddings, and the pooled BERT embeddings of the tweets in its propagation tree.

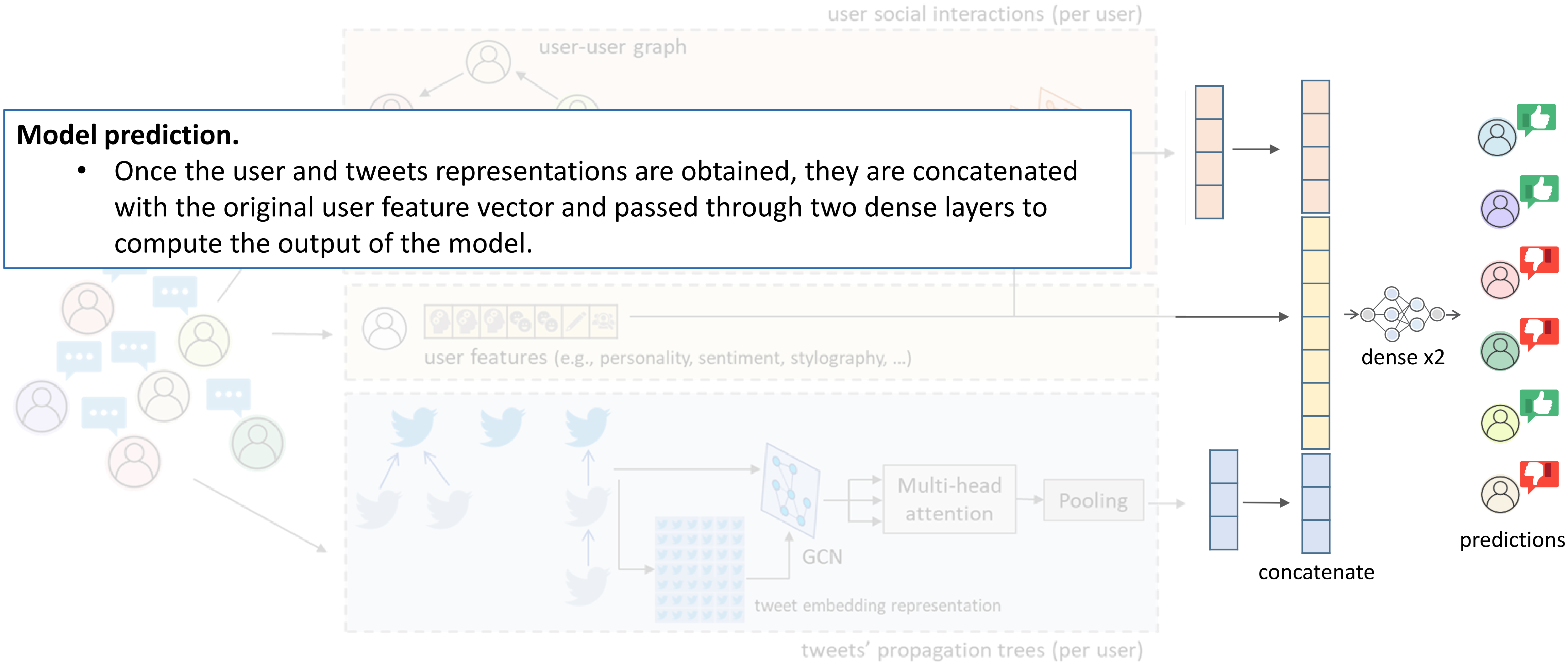
- The propagation trees and the embeddings are fed to a single GCN.
- The multi-headed self-attention mechanism, to better learn tweet interactions.





tweets' propagation trees (per user)

Model overview



Model prediction.

- Once the user and tweets representations are obtained, they are concatenated with the original user feature vector and passed through two dense layers to compute the output of the model.

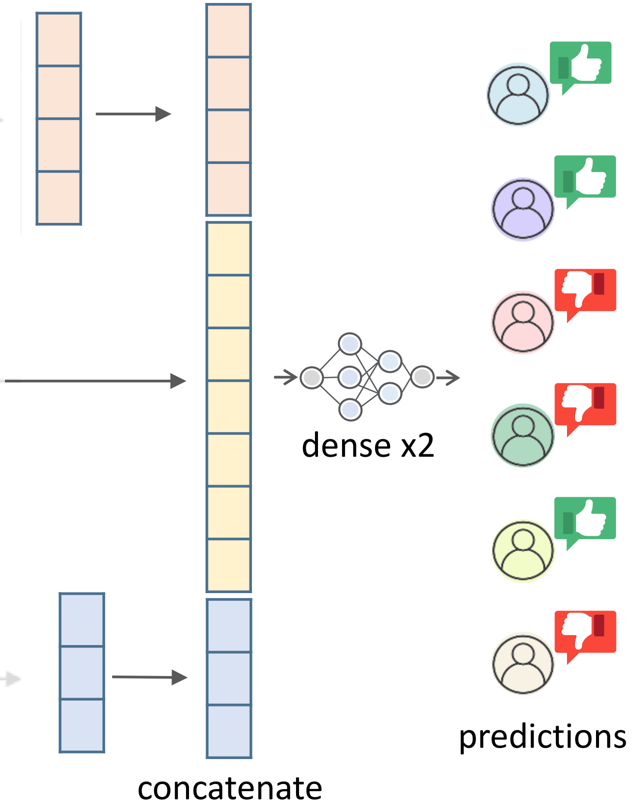
Model overview

Model prediction.

- Once the user and tweets representations are obtained, they are concatenated with the original user feature vector and passed through two dense layers to compute the output of the model.

Model training.

- The model was trained using a Hinge loss function with class weights.
- Class weights aimed to compensate for unbalanced class distribution in training, which would induce a majority class prediction.



Experimental evaluation

Data

- We used the **FibVid** data collection.
  - Tweets related to the COVID-19 pandemic.
  - The collection is based on news claims appearing in **Politifact** and **Snopes**.

- Tweets were retrieved using the **Faking it!** tool.

- The collection comprised **772** COVID-related news claims and **112k** relevant tweets belonging to **24k** users, which were shared during 2020.

- Tweets have a authentic/fake label based on Politifact and Snopes.
  - 26% authentic content, 74% fake content.
  - Labels were used to determine whether users were fake news spreaders.
  - Users were deemed as spreaders if the **proportion of shared fake content was higher than a certain threshold** (0.5).

Experimental evaluation

Baselines

Traditional

- Based on hand-crafted feature sets.
  - Tweet/user stats (popularity, screenname length, account age, …).
  - LIWC.
  - Personality traits.
  - Readability.
  - Content-based embeddings.
  - Topology-based embeddings.

State-of-the-art

- All based on deep-learning models.
- Mostly based on content-based information.

Experimental evaluation

Evaluation

Data split

- All evaluations were performed over the **same data partitions** and evaluated using the same set of metrics.
- Temporal data split.
  - Training set: first 70% users sorted according to the date of their first interaction.
  - Test set: remaining users.

Evaluation metrics

- Binary/weighted precision and recall.
  - More importance to recall.
- AUC-ROC.

Results

	Binary (spreaders class)		Weighted		AUC-ROC
	precision	recall	precision	recall	
our model	**0.851**	**0.840**	**0.839**	**0.838**	**0.878**
tweet stats	<u>0.833</u>	<u>0.669</u>	<u>0.769</u>	<u>0.755</u>	<u>0.840</u>
user stats & tweet stats	0.803	0.623	0.737	0.721	0.726
readability	0.573	0.448	0.542	0.534	0.539
LIWC	0.555	0.443	0.526	0.520	0.520
personality	0.537	0.413	0.511	0.504	0.509
node2vec	0.660	0.544	0.620	0.612	0.658
GloVe	0.539	0.436	0.512	0.507	0.510
BERT	0.534	0.411	0.508	0.501	0.506
Sharma and Sharma	0.634	0.239	0.571	0.527	0.713
Sansonetti	0.467	0.642	0.397	0.426	0.688
CheckerOrSpreader	0.809	0.119	0.662	0.521	0.544
Shrestha and Spezzano	0.634	0.239	0.571	0.527	0.713
FacTweet	0.436	0.176	0.523	0.561	0.506

Best results in **bold**.

Results

		Binary (spreaders class)		Weighted		AUC-ROC
		precision	recall	precision	recall	
	our model	**0.851**	**0.840**	**0.839**	**0.838**	**0.878**
Simple baselines	tweet stats	0.833	0.669	0.769	0.755	0.840
	user stats & tweet stats	0.803	0.623	0.737	0.721	0.726
	readability	0.573	0.448	0.542	0.534	0.539
	LIWC	0.555	0.443	0.526	0.520	0.520
	personality	0.537	0.413	0.511	0.504	0.509
	node2vec	0.660	0.544	0.620	0.612	0.658
	GloVe	0.539	0.436	0.512	0.507	0.510
	BERT	0.534	0.411	0.508	0.501	0.506
State-of-the-art baselines	Sharma and Sharma	0.634	0.239	0.571	0.527	0.713
	Sansonetti	0.467	0.642	0.397	0.426	0.688
	CheckerOrSpreader	0.809	0.119	0.662	0.521	0.544
	Shrestha and Spezzano	0.634	0.239	0.571	0.527	0.713
	FacTweet	0.436	0.176	0.523	0.561	0.506

Best results in **bold**.

Results

	Binary (spreaders class)		Weighted		AUC-ROC
	precision	recall	precision	recall	
tweet stats	<u>0.833</u>	<u>0.669</u>	<u>0.769</u>	<u>0.755</u>	<u>0.840</u>
user stats & tweet stats	0.803	0.623	0.737	0.721	0.726
readability	0.573	0.448	0.542	0.534	0.539
LIWC	0.555	0.443	0.526	0.520	0.520
personality	0.537	0.413	0.511	0.504	0.509
node2vec	0.660	0.544	0.620	0.612	0.658
GloVe	0.539	0.436	0.512	0.507	0.510
BERT	0.534	0.411	0.508	0.501	0.506

Simple baselines

- Best baselines results were obtained with simple user/tweet features.
    - High precision, but relatively low recall.
- Hand-crafted content features achieved similar results than considering embeddings.
    - Text have similar content (expected) and similar style.
    - Differences might be too subtle for the embeddings to detect them.
- Network topology (node2vec) seems to be more useful than content.

Results

	Binary (spreaders class)		Weighted		AUC-ROC
	precision	recall	precision	recall	
tweet stats	0.833	0.669	0.769	0.755	0.840
user stats & tweet stats	0.803	0.623	0.737	0.721	0.726

Simple baselines

- Best baselines results were obtained with simple user/tweet features.
  - High precision, but relatively low recall.
- Hand-crafted content features achieved similar results than considering embeddings.
  - Text have similar content (expected) and similar style.
  - Differences might be too subtle for the embeddings to detect them.
- Network topology (node2vec) seems to be more useful than content.

Results

	Binary (spreaders class)		Weighted		AUC-ROC
	precision	recall	precision	recall	
readability	0.573	0.448	0.542	0.534	0.539
LIWC	0.555	0.443	0.526	0.520	0.520
personality	0.537	0.413	0.511	0.504	0.509
GloVe	0.539	0.436	0.512	0.507	0.510
BERT	0.534	0.411	0.508	0.501	0.506

Simple baselines

- Best baselines results were obtained with simple user/tweet features.
  - High precision, but relatively low recall.
- Hand-crafted content features achieved similar results than considering embeddings.
  - Text have similar content (expected) and similar style.
  - Differences might be too subtle for the embeddings to detect them.
- Network topology (node2vec) seems to be more useful than content.

Results

	Binary (spreaders class)		Weighted		AUC-ROC
	precision	recall	precision	recall	
node2vec	0.660	0.544	0.620	0.612	0.658

Simple baselines

- Best baselines results were obtained with simple user/tweet features.
  - High precision, but relatively low recall.
- Hand-crafted content features achieved similar results than considering embeddings.
  - Text have similar content (expected) and similar style.
  - Differences might be too subtle for the embeddings to detect them.
- Network topology (node2vec) seems to be more useful than content.

Results

	Binary (spreaders class)		Weighted		AUC-ROC
	precision	recall	precision	recall	
Sharma and Sharma	0.634	0.239	0.571	0.527	0.713
Sansonetti	0.467	0.642	0.397	0.426	0.688
CheckerOrSpreader	0.809	0.119	0.662	0.521	0.544
Shrestha and Spezzano	0.634	0.239	0.571	0.527	0.713
FacTweet	0.436	0.176	0.523	0.561	0.506

State-of-the-art baselines

- Similar precision than simpler baselines, but much lower recall.
- Sansonetti et al. achieved the best recall results.
  - Confirm the relevance of user/tweet stats for the detection.
- FacTweet lowest precision and recall.
  - Dividing user timeline in chunks to classify them independently might not be the best alternative with small datasets.

Results

	Binary (spreaders class)		Weighted		AUC-ROC
	precision	recall	precision	recall	
Sansonetti	0.467	0.642	0.397	0.426	0.688

State-of-the-art baselines

- Similar precision than simpler baselines, but much lower recall.
- Sansonetti et al. achieved the best recall results.
  - Confirm the relevance of user/tweet stats for the detection.
- FacTweet lowest precision and recall.
  - Dividing user timeline in chunks to classify them independently might not be the best alternative with small datasets.

Results

	Binary (spreaders class)		Weighted		AUC-ROC
	precision	recall	precision	recall	
FacTweet	0.436	0.176	0.523	0.561	0.506

- Similar precision than simpler baselines, but much lower recall.
- Sansonetti et al. achieved the best recall results.
  - Confirm the relevance of user/tweet stats for the detection.
- FacTweet lowest precision and recall.
  - Dividing user timeline in chunks to classify them independently might not be the best alternative with small datasets.

State-of-the-art baselines

Experimental evaluation

Results - Highlights

	Traditional	**State-of-the-art**
Avg. precision Improvements	43%	54%
Avg. recall improvements	61%	184%
Avg. AUC-ROC improvements	51%	42%

- Best baselines results were obtained with simple user/tweet features.
  High precision, but relatively low recall.
- Hand-crafted content features achieved similar results than considering content embeddings.
- Network topology seemed to be more useful than content.

- Our model achieved the highest results.
  - Better balance between precision and recall than the evaluated baselines.
- Some baselines achieved similar precision to our model, but lower recall.

Summary (III)

We presented **a model for identifying fake news spreaders in social media** by combining content and user features, the induced propagation trees, and features learned from user interactions.

A preliminary evaluation showed the **models' potential for accurately detecting fake news spreaders** and the i**mportance of combining the different aspects of user representation** to achieve a more effective characterization of spreaders.

Ethical statement

Disclaimer

- In all cases, research was based on publicly available Twitter data originally collected and tagged by third parties.
- No user personal information was included in the analysis, and no user identity is being disclosed.
- As per Twitter TOS, the shared data only includes user and tweet IDs and aggregated content features.
- The presented works aimed to provide users with tools to identify and quantify unwanted phenomena and, in the long term, to help users increase their decision awareness.
    - Results of the studies should not be used to publicly criticize any user, regardless of their behaviour.
- The studies can suffer from bias from the data collection and tagging process.

Quick recap

What have we done?

How misinformation spreads?

How to avoid misinformation derived effects?

How to identify users propagating misinformation?

We presented a **preliminary exploration** to better understand how **user recommenders affect network dynamics** in terms of **misinformation spreader distribution and influence**.

Quick recap

What have we done?

How misinformation spreads?

We presented a **preliminary exploration** to better understand how **user recommenders affect network dynamics** in terms of **misinformation spreader distribution and influence**.

How to avoid misinformation derived effects?

We developed `FRediECH` inspired by a **graph convolutional network** and a **Deep & Wide** architecture, coupling echo chamber awareness and user representations to balance the relevance, diversity and novelty of friend recommendations.

How to identify users propagating misinformation?

Quick recap

What have we done?

How misinformation spreads?	We presented a **preliminary exploration** to better understand how **user recommenders affect network dynamics** in terms of **misinformation spreader distribution and influence**.
How to avoid misinformation derived effects?	We developed `FRediECH` inspired by a **graph convolutional network** and a **Deep & Wide** architecture, coupling echo chamber awareness and user representations to balance the relevance, diversity and novelty of friend recommendations.
How to identify users propagating misinformation?	We presented **a model for identifying fake news spreaders in social media** by combining content and user features, the induced propagation trees, and features learned from user interactions.

Quick recap

What do we want to do in the future?

ACADEMIC HARDWARE GRANT PROGRAM

How misinformation spreads?

- Perform a more extensive evaluation with other data collections, recommenders and opinion models.
- Explore recommendations and their effect on misinformation spreading.
- Consider different follow/unfollow dynamics and densification scenarios.

How to avoid misinformation derived effects?

How to identify users propagating misinformation?

Quick recap

What do we want to do in the future?

ACADEMIC HARDWARE GRANT PROGRAM

How misinformation spreads?

- Perform a more extensive evaluation with other data collections, recommenders and opinion models.
- Explore recommendations and their effect on misinformation spreading.
- Consider different follow/unfollow dynamics and densification scenarios.

How to avoid misinformation derived effects?

- Evaluations over different datasets varying the domain and time period to truly assess usefulness and generalizability.
- Analyses regarding the relevance of each type of interaction, and their contribution to the final recommendations.
- Explanations to better guide users in broadening their interactions.

How to identify users propagating misinformation?

Quick recap

What do we want to do in the future?

ACADEMIC HARDWARE GRANT PROGRAM

How misinformation spreads?

- Perform a more extensive evaluation with other data collections, recommenders and opinion models.
- Explore recommendations and their effect on misinformation spreading.
- Consider different follow/unfollow dynamics and densification scenarios.

How to avoid misinformation derived effects?

- Evaluations over different datasets varying the domain and time period to truly assess usefulness and generalizability.
- Analyses regarding the relevance of each type of interaction, and their contribution to the final recommendations.
- Explanations to better guide users in broadening their interactions.

How to identify users propagating misinformation?

- Evaluate with other data collections varying scale and domain.
- Explore the representation of user relations.
- Explore the temporal relation of tweets.
- Perform an ablation study.

Thanks!

Questions?

antonela.tommasel@isistan.unicen.edu.ar

Recommender systems and misinformation
What can we do about them?

ANTONELA TOMMASEL

CONICET
ISISTAN