

La misoginia online: metodi e tecniche

Lara Fontanella, Annalina Sarra

Dipartimento di Studi Socio-Economici, Gestionali e Statistici

Università G. D'Annunzio

Chieti-Pescara



ICOMIC: Identifying and Countering Online Misogyny

Project funded by EU Next Generation, MUR-Fondo Promozione e Sviluppo-DM 737/2021

Rassegne della letteratura su hate speech e misoginia online

Home > Scientometrics > Article

Thirty years of research into hate speech: topics of interest and their evolution

Open access | Published: 30 October 2020
Volume 126, pages 157–179, (2021) [Cite this article](#)

[Download PDF](#) You have full access to this open access article

Alice Tontodimamma, Eugenia Nissi, Annalina Sarra & Lara Fontanella

28k Accesses 8 Altmetric [Explore all metrics](#) →

Abstract

The exponential growth of social media has brought with it an increasing propagation of hate speech and hate based propaganda. Hate speech is commonly defined as any communication that disparages a person or a group on the basis of some characteristics such as race, colour, ethnicity, gender, sexual orientation, nationality, religion. Online hate diffusion has now developed into a serious problem and this has led to a number of international initiatives being proposed, aimed at qualifying the problem and developing effective counter-measures. The aim of this paper is to analyse the knowledge structure of hate speech literature and the evolution of related topics. We apply co-word analysis methods to identify different topics treated in the field. The analysed database was downloaded from Scopus, focusing on a number of publications during the last thirty years. Topic and network analyses of literature showed that the main research topics can be divided into three areas: "general debate hate speech versus freedom of expression", "hate-speech automatic detection and classification by machine-learning strategies", and "gendered hate speech and cyberbullying". The understanding of how research fronts interact led to stress the relevance of machine learning approaches to correctly assess hatred forms of online speech.

Humanities & Social Sciences Communications

Explore content About the journal Publish with us

nature > humanities and social sciences communications > review articles > article

Review Article | Open access | Published: 02 April 2024

How do we study misogyny in the digital age? A systematic literature review using a computational linguistic approach

Lara Fontanella, Berta Chulvi, Elisa Ignazzi, Annalina Sarra & Alice Tontodimamma

Humanities and Social Sciences Communications 11, Article number: 478 (2024) | [Cite this article](#)

28k Accesses 4 Altmetric [Metrics](#)

Abstract

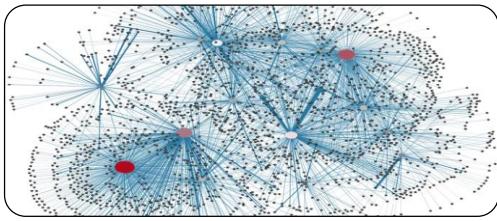
Nowadays, despite centuries of striving for equality, women still face higher levels of discrimination compared to men in nearly every aspect of life. Recently, this systemic inequality has manifested in cyberspace through the proliferation of abusive content that is even more aggressive than what one would expect in the 21st century. Various research disciplines are now attempting to characterise this new manifestation of misogyny. The endeavour to comprehend this phenomenon has resulted in a significant increase in publications from several fields, including Social Sciences, Arts and Humanities, Psychology, and Computer Science. This paper presents a systematic review of multidisciplinary research on misogyny from the years 1990 to 2022, encompassing a total of 2830 articles retrieved from the Scopus database as of December 31, 2022. The literature is thoroughly analysed using three approaches: bibliometric analysis, topic detection, and qualitative analysis of the documents. The findings suggest that the analysis of online misogyny has been the primary driver behind the exponential growth in publications in this field. Additionally, the results of the topic analysis and topic interaction reveal a limited connection between the areas of knowledge that are necessary to fully grasp this complex phenomenon.



Misogynistic speech detection

WPs

- Build a new **lexical resource** specifically designed to include terms expressing hatred towards women
- Create a **corpus** of comments shared on Twitter, Facebook, Instagram, YouTube, and Reddit **annotated for misogyny**.
- Compare the performance of **interpretable machine learning model**, exploiting lexical, lexicon-based, and sentiment analysis features, with xAI approaches.



Identification of producers of misogynistic content

WPs

- Create an extensive **corpus of tweets** addressing politically-active women, feminists, female entrepreneurs, journalists, influencers, and celebrities.
- Tweets automatic classification and identification of a **subset of Twitter users that are more likely to share misogynistic contents**.
- Build the **ego-networks** of the selected seed producers
- **Collective classification** of the network nodes to identify which nodes are more likely to share misogynistic contents.
- **Community detection** and **discursive critical analysis** of the textual content of the different communities.



Countering misogynistic speech

WPs

- Derive a **rough list of counter-actions** through a systematic review of the scientific literature
- Submit the list of counter-actions to a **panel of experts** to integrate and refine it.
- Development of **futures scenarios**.
- **Evaluation of the impacts of counter-actions** on future scenarios with a cross-impact approach.

Misogynistic speech detection



TEXTUAL
CORPUS



➤ Creazione di un lessico misogino

- predisposizione del lessico delle parole misogine a partire da revisedHurtlex; annotazione da parte di 6 annotatori; individuazione di uno score finale

➤ Creazione di un corpus di 13500 commenti annotati per misogynia

- Estrazione di commenti da Twitter, Facebook e Instagram
- Campionamento dei contenuti da annotare
- Predisposizione delle linee guida e formazione di 10 annotatori
- Predisposizione dei task di annotazione su LabelStudio
- Valutazione dell'agreement fra annotatori.

➤ Identificazione automatica di contenuti misogini

- applicazione di tecniche di xAI (ExPred)

Textual Data Collection

- **Social media:** Facebook, Instagram, and Twitter (*now rebranded X*)
- **Web scraping tools:** Facebook and Instagram <https://exportcomments.com/>,
Twitter <https://www.socialgrabber.net/>
- Comments related to:
 - **Political Women:** Chiara Appendino, Anna Ascani, Lucia Azzolina, Anna Maria Bernini, Laura Boldrini, Giulia Bongiorno, Maria Elena Boschi, Mara Carfagna, Marta Cartabia, Susanna Ceccardi, Monica Cirinnà, Ilaria Cucchi, Paola De Micheli, Michela Di Biase, Federica Gasbarro, Mariastella Gelmini, Barbara Lezzi, Sanna Marin, Giorgia Meloni, Alessia Morani, Paola Nugnes, Virginia Raggi, Elly Schlein, Valentina Vezzali
 - **Female television personality and Influencers:** Caterina Balivo, Ilary Blasi, Giulia De Lellis, Elodie, Elisa Esposito, Chiara Ferragni, Michelle Hunziker, Vanessa Incontrada, Elisa Isoardi, Miriam Leone, Emma Marrone, Aurora Ramazzotti, Belén Rodríguez
 - **Female Journalists:** Lucia Annunziata, Bianca Berlinguer, Luisella Costamagna, Ilaria d'Amico, Veronica Gentile, Diletta Leotta

Corpus construction and annotation

- **Annotated corpus:** 13,500 comments varying across social media platforms.

Annotation

- **Ten annotators** were trained and split into **5 groups** (1 woman, 1 man per group).
- Each group independently annotated a portion of the dataset using a misogyny **binary scheme**.

Agreement		Disagreement	Gwet's AC_1
non-misogynistic	misogynistic		
59.8%	26.8%	13.4%	
			0.76

- Following Zaidan et al. [2008], annotators **highlighted text snippets** to justify why a comment was deemed misogynistic → **annotated rationales**.

Creazione del corpus

Annotating comments with Label Studio

<https://labelstud.io/>

Label Studio Projects / Primo Round - Gruppo 1 EM

#45472340 2 of 5 < > ⌂ ⌃ ✖ ⌚ Skip Submit

Dispregiativo 1 Intimidatorio 2 Dominanza 3 Denigrazione 4 Sessismoben 5 Neosessismo 6 Giustificazione 7

SoloOffensivo 8

Details Annotation History #kmDeB

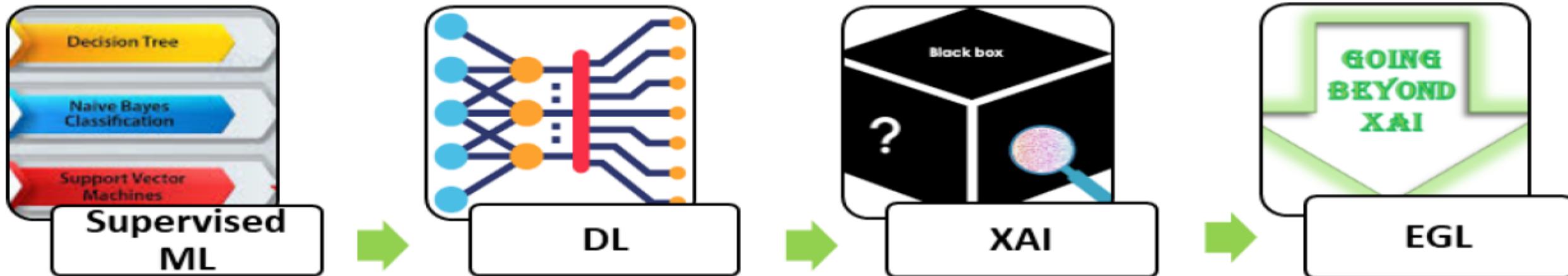
emideigo just now Draft

Relations (0)

Comments Add a comment ➤

Ma fatela stare zitta questa gallina

Approaches to Automatic Misogyny Identification



Machine Learning

- SVM
- Logistic Regression
- Naïve Bayes
- Random Forest

Deep Learning

- **Artificial Neural Networks** (CNNs, RNNs, LSTMs)
- **Transformer-based models** (e.g., BERT)
- **Large Language Models** (e.g., Chat-GPT)

Explainable AI

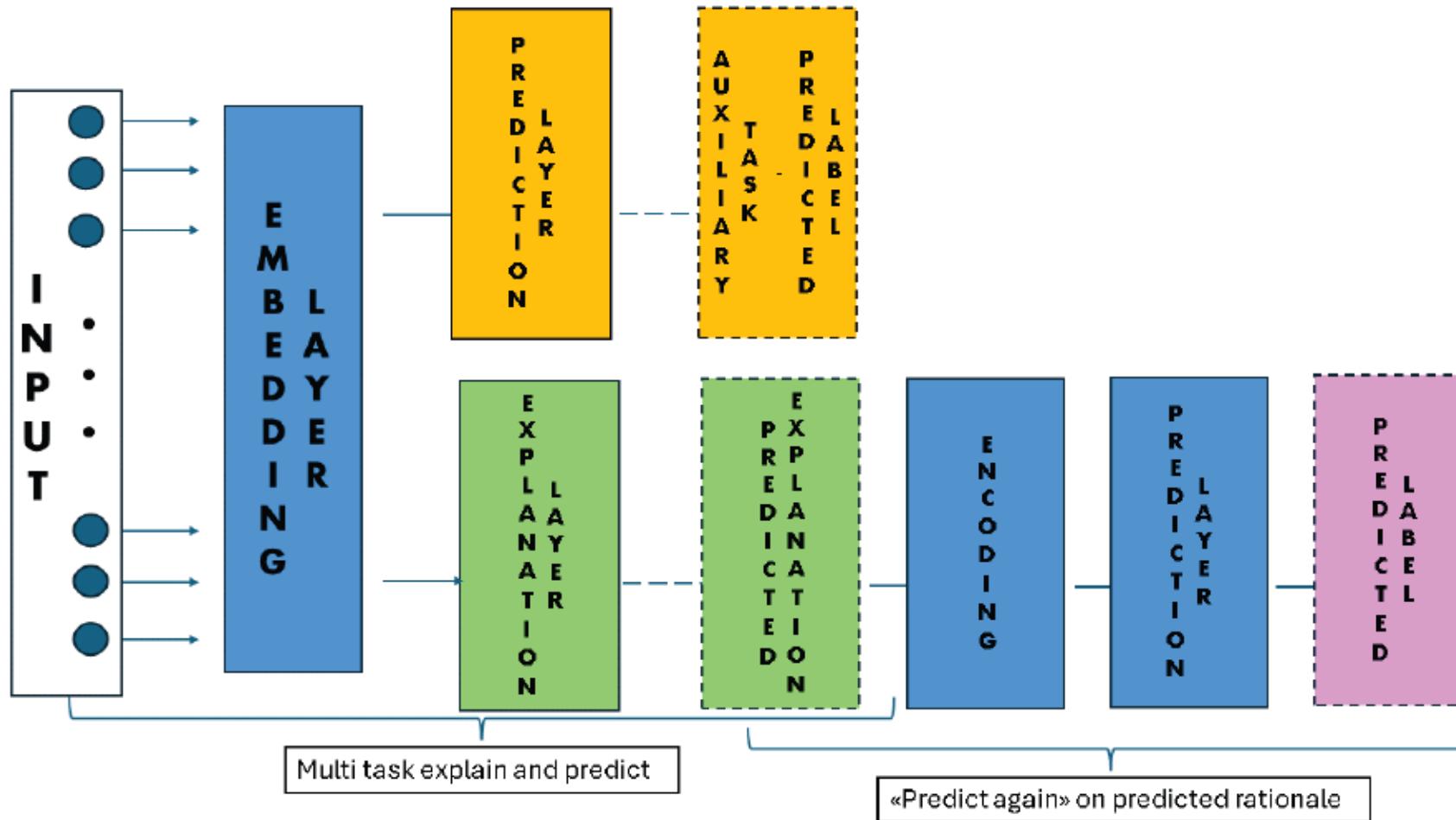
- **SHAP** (*Shapley Additive Explanations*)
- **SOC** (*Sampling-And-Occlusion*)
- **LIME** (*Local Interpretable Model-agnostic Explanations*)

Explanation Guided Learning

Also known as

- *Explanation Supervision*
- *Attention Supervision*
- *Explanation Alignment*
- *Learning from Explanation*

Explain-then-Predict-then-Predict Again [ExPred, Zhang et al., 2021]

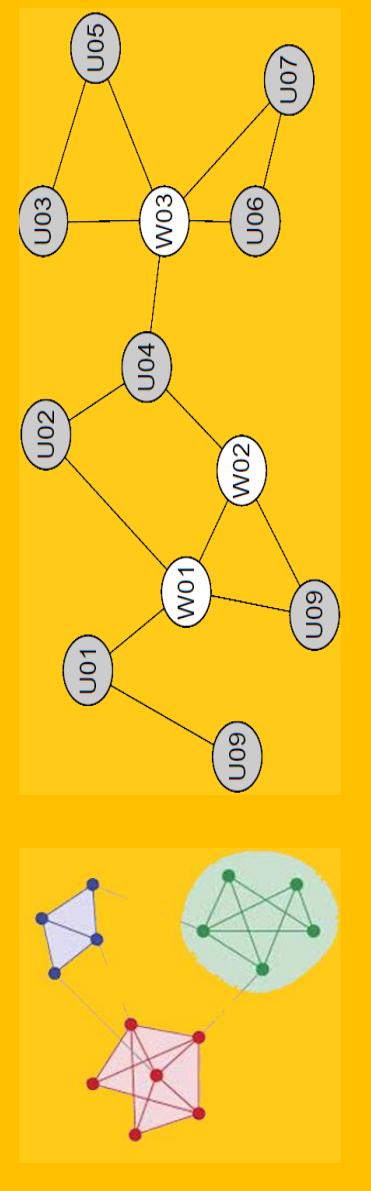


Model Performance

- **Python code:** <https://github.com/JoshuaGhost/expred>
- **Encoding:** BERT Italian version
<https://huggingface.co/dbmdz/bert-base-italian-xxl-uncased>

	NON-MISOGYNISTIC			MISOGYNISTIC			WEIGHTED AVERAGE		
	f1-score	precision	recall	f1-score	precision	recall	f1-score	precision	recall
$\lambda=5$	0.78	0.85	0.72	0.81	0.76	0.87	0.79	0.80	0.79
$\lambda=20$	0.77	0.77	0.77	0.77	0.77	0.77	0.77	0.77	0.77
$\lambda=50$	0.78	0.75	0.81	0.76	0.80	0.73	0.77	0.77	0.77
$\lambda=75$	0.77	0.72	0.82	0.74	0.80	0.69	0.75	0.76	0.76

Identification of producers of misogynistic content



➤ Raccolta di dati relazionali e testuali da Twitter

- Individuazione di una rete contenente 7,354 account (nodi) a cui corrispondono 82,807 tweets

➤ Annotazione di un sottoinsieme di nodi della rete (*annotazione a livello di utente*)

- Selezione e annotazione manuale di 937 nodi

➤ Classificazione automatica dell'intera rete

- Text Graph Convolutional Network

➤ Analisi di comunità e Critical Discourse Analysis

- Analisi della rete della manusfera su Facebook

Textual data collection

- **Social media:** Twitter (*now rebranded X*)
- **Temporal interval:** real time download from August 2022 to December 2022
- **Software:** Socialgrabber <https://www.socialgrabber.net/>
- **Keywords** related to:
 - **Political Women:** Chiara Appendino, Anna Ascani, Lucia Azzolina, Anna Maria Bernini, Laura Boldrini, Giulia Bongiorno, Maria Elena Boschi, Mara Carfagna, Marta Cartabia, Susanna Ceccardi, Monica Cirinnà, Ilaria Cucchi, Paola De Micheli, Michela Di Biase, Federica Gasbarro, Mariastella Gelmini, Barbara Lezzi, Sanna Marin, Giorgia Meloni, Alessia Morani, Paola Nugnes, Virginia Raggi, Elly Schlein, Valentina Vezzali
 - **Female television personality and Influencers:** Caterina Balivo, Ilary Blasi, Giulia De Lellis, Elodie , Elisa Esposito, Chiara Ferragni, Michelle Hunziker, Vanessa Incontrada, Elisa Isoardi, Miriam Leone, Emma Marrone, Aurora Ramazzotti, Belén Rodríguez
 - **Female Journalists:** Lucia Annunziata, Bianca Berlinguer, Luisella Costamagna, Ilaria d'Amico, Veronica Gentile, Diletta Leotta

Twitter accounts' textual data

Relational data collection

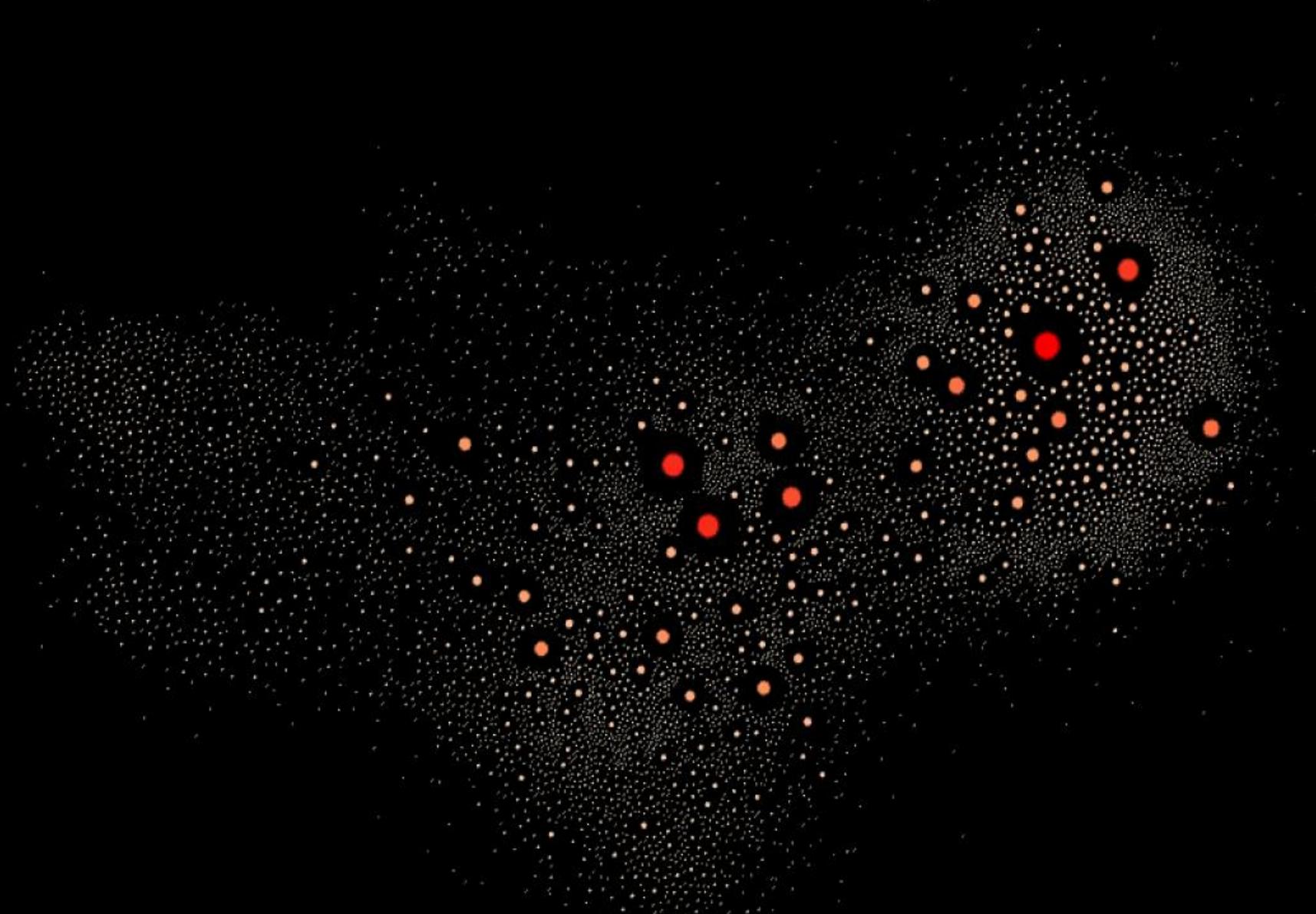
User selection rules and relational data collection:

- removal of accounts related to information providers (e.g., *newspapers, radio stations, television channels, and television programs, news aggregators*)
- removal of accounts with less than 5 tweets.
- removal of accounts with an outlier number of tweets.
- downsampling of accounts with tweets focused only on Giorgia Meloni
- removal of accounts no longer existing
- retrieval of friend/follower relations from Twitter using Socialgrabber

Number of accounts in the network: 7,354

Number of tweets: 82,807

Twitter accounts' friend-follower network



Training data

A subset of accounts was selected based on the following criteria:

- Node centrality measures: a well-spread sample on the network likely includes nodes with the highest degree and betweenness indexes [del Gobbo, 2021]
- Distribution of tweets by the women chosen for corpus construction
- Level of offensiveness: the revised Hurtlex dictionary [Tontodimamma et al., 2022] was used to derive an offensiveness score at the producer level

Number of sampled accounts: 937

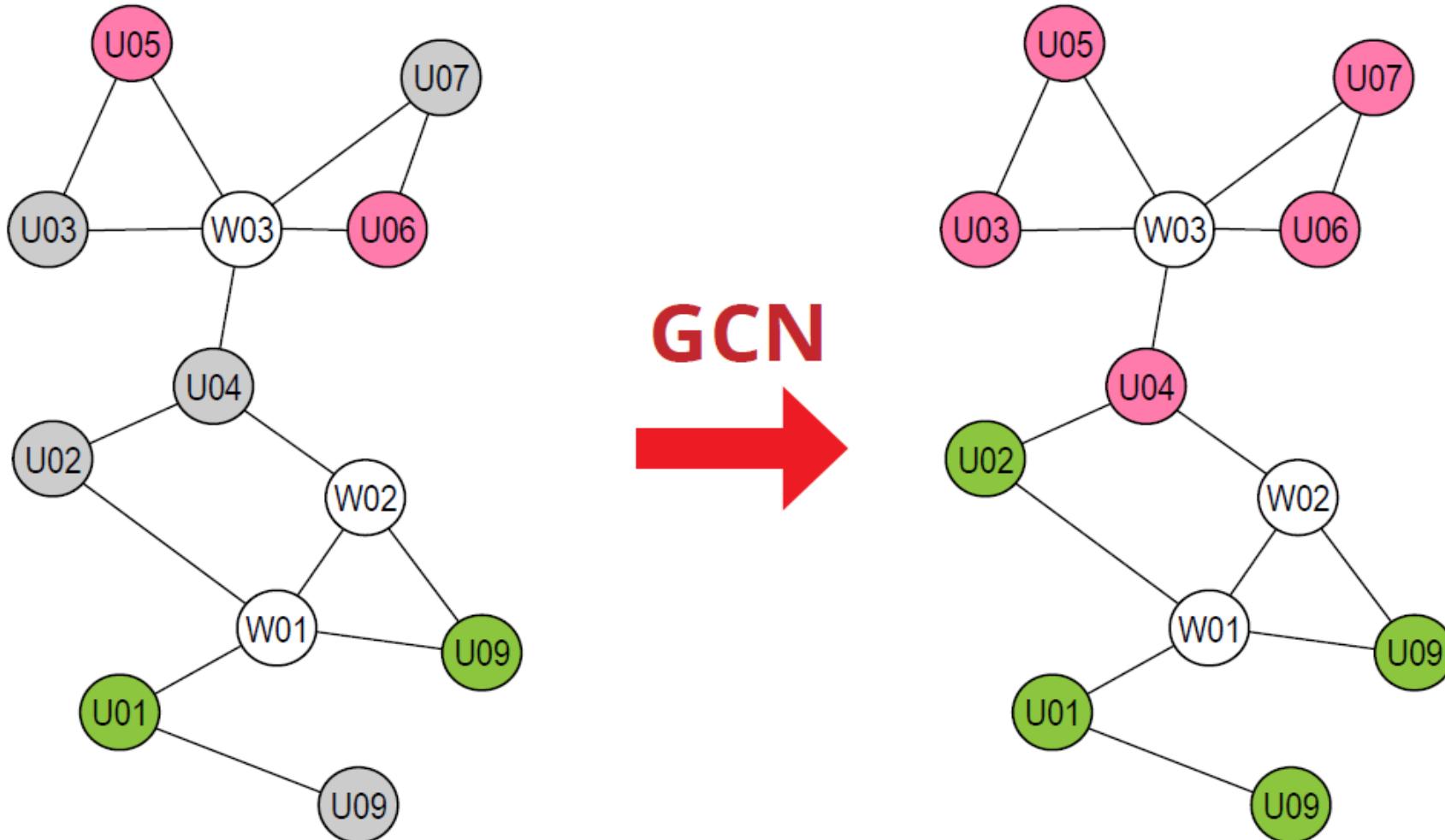
During their internship program, students were trained on the concept of misogyny and issues related to annotation.

They analysed in detail the textual content shared by the sampled users and manually classified them using a binary coding schema

Annotated accounts: 55.4% *non-misogynistic*; 44.6% *misogynistic*.

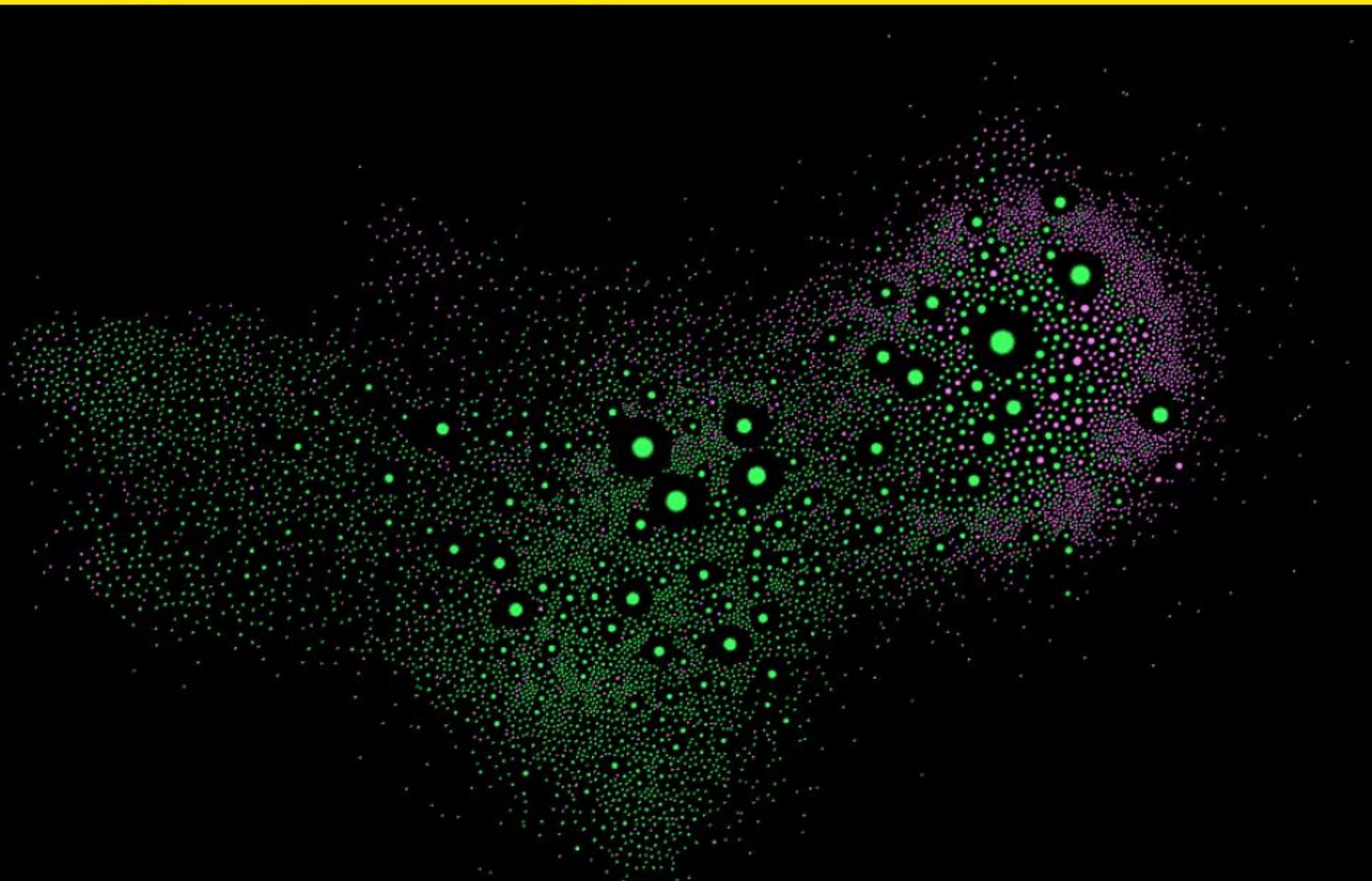
➤ Classificazione automatica dell'intera rete

Semi-supervised classification with GCN



Predicted misogynistic accounts on the Network

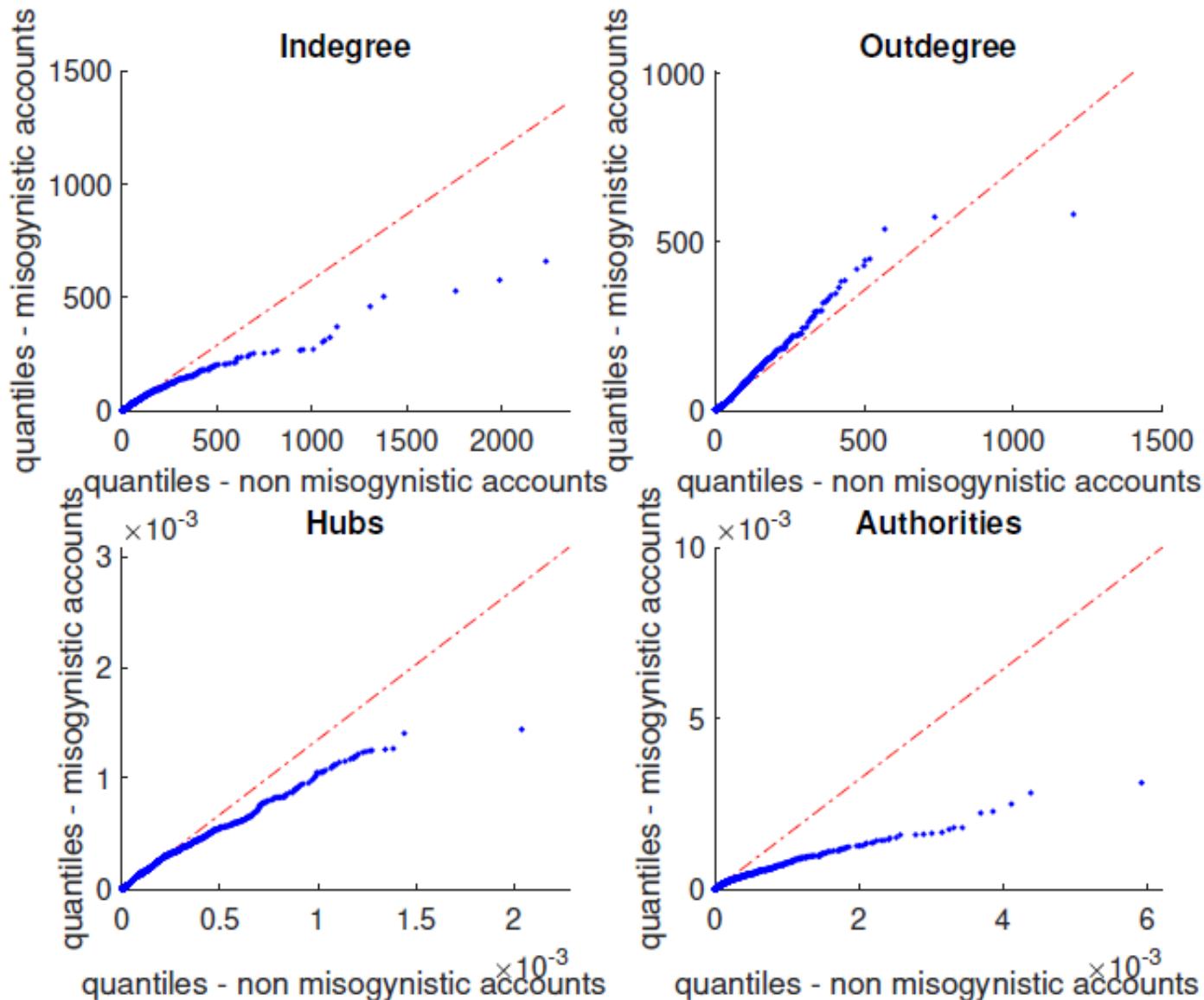
63.0% non-misogynistic; 37.0% misogynistic



Centrality measures on the friend directed network

Misogynistic Twitter accounts:

- are **clustered**
- tend to **follow more people**
- tend to **be followed by less people**
- tend to be of **less importance** in the network structure



Textual information comparison

Misogynistic Twitter

accounts:

- differ from other accounts in terms of their word usage.

misogynistic

The word cloud illustrates the frequency of specific words used in misogynistic vs. non-misogynistic Twitter accounts. The most prominent words are 'cazzo' (pink) and 'merda' (pink), which are repeated numerous times. Other significant pink words include 'signora', 'vai', 'schifo', 'farsi', 'inutile', 'ahahahah', 'vuoi', 'state', 'solita', 'anzi', 'deve', 'esperanza', 'lavorare', 'recarate', 'testa', 'ben', 'giornata', 'anno', 'nuovo', 'video', 'serata', 'ogni', 'dopo', 'vuole', 'raga', 'punto', 'cosa', 'molto', 'pare', 'giovani', 'intervista', 'figli', 'stesso', 'lavoro', 'cosa', 'insieme', 'cuore', 'storia', 'università', 'servizi', 'crisi', 'coalizione', 'momento', 'social', 'amore', 'attacco', 'modo', 'buon', 'voglio', 'capisco', 'stagione', 'credo', 'scusate', 'canzone', 'forza', 'foto', 'inizio', 'avere', 'nessun', 'realta', 'purtroppo', 'stasera', 'voleva', 'boh', 'nuovi', 'caso', 'studio', 'ieri', 'davy', 'ergo', 'idea', 'durante', 'storie', 'uscita', 'vabbè', 'tema', 'ecco', 'sotto', 'nuova', 'aborto', 'avere', 'E', 'donne', 'grazie', 'edizione'. Green words include 'chiama', 'nato', 'regime', 'meglio', 'basta', 'nemmeno', 'metti', 'neanche', 'scrive', 'solito', 'potere', 'frega', 'pescivendola', 'dovevi', 'vogliamo', 'almeno', 'sig.ra', 'figa', 'galera', 'boldrina', 'speriamo', 'maledetti', 'ciastroni', 'finta', 'stare', 'fanculo', 'sai', 'risorse', 'certo', 'qualcosa', 'rosica', 'pagliaccio', 'vergogna', 'vero', 'andate', 'caxxo', 'parli', 'ignorante', 'vecchia', 'guarda', 'dignità', 'cesso', 'gadesso', 'stronza', 'invece', 'male', 'cane', 'fogna', 'clandestini', 'idiota', 'allora', 'tacere', 'merde', 'dovrebbe', 'andare', 'magari', 'affanculo', 'vergogni', 'parassita', 'ragione', 'senza', 'nessuno', 'povera', 'puoi', 'rompere', 'quei', 'scuse', 'parla', 'carciofo', 'fara', 'palle', 'coglioni', 'cazzi', 'fatti', 'pensa', 'ridicola', 'niente', 'nulla', 'coglione', 'faccia', 'bene', 'button', 'cambia', 'vada', 'grande', 'rotto', 'culo', 'vai', 'signora', 'fate', 'fuori', 'taci', 'dovete', 'ahahahah', 'vuoi', 'solo', 'vergognati', 'state', 'pure', 'cagare', 'schifo', 'cazzo', 'cervello', 'altra', 'odio', 'solita', 'deve', 'esperanza', 'casacoglione', 'brutta', 'use', 'lavorare', 'recarate', 'testa', 'ben', 'merda', 'soldi', 'vaffanculo', 'bollettina', 'sembrala', 'voro', 'devianze', 'merito', 'twitter', 'sole', 'nuovo', 'vista', 'discorso', 'fine', 'nuovi', 'serata', 'ogni', 'dopo', 'qui', 'scuola', 'caso', 'studio', 'vuole', 'raga', 'punto', 'cosa', 'molto', 'ieri', 'davy', 'ergo', 'idea', 'durante', 'insomma', 'pare', 'giovani', 'intervista', 'figli', 'stesso', 'sera', 'storie', 'uscita', 'lavoro', 'cosa', 'insieme', 'cuore', 'storia', 'università', 'servizi', 'crisi', 'coalizione', 'momento', 'social', 'amore', 'attacco', 'modo', 'buon', 'voglio', 'capisco', 'stagione', 'credo', 'scusate', 'canzone', 'forza', 'foto', 'inizio', 'avere', 'nessun', 'realta', 'purtroppo', 'stasera', 'voleva', 'vabbè', 'tema', 'ecco', 'sotto', 'nuova', 'aborto', 'avere', 'E', 'donne', 'grazie', 'edizione'.

non misogynistic

Analisi di comunità e Critical Discourse Analysis: The Manosphere

Shared Ideology: misogynistic and anti-feminist gender ideology, promoting a toxic form of hegemonic masculinity.

Key Concepts:

- **Gynocentrism:** belief in a societal structure that favours women over men.
- **Misandry concept:** used by both extreme and moderate groups, enables these groups to adopt the language of identity politics, positioning men as the silenced victims of reverse discrimination in all aspects of political, economic, and social life and solidifying their sense of entitlement.
- **Red Pill Ideology:** Derived from *The Matrix*, represents awakening to "truths" about gender dynamics.

The Manosphere groups



Men's Rights Activists (MRAs)

Individuals and groups that believe men are systematically disadvantaged because of their gender

Involuntary Celibates (Incels)

Online subculture of primarily heterosexual men who identify as being unable to have romantic or sexual relationships.

Men Going Their Own Way (MGTOW)

Separatists, misogynists, anti-feminists. It is an extremely large group: 'mgtow.com' alone has 33,000 members. The site also lists 25 YouTube channels,

Pick-Up-Artists (PUA)

Self-proclaimed alpha males, who try to convince women to have sex with them through a mixture of flattery, psychological manipulation and coercion.

Network Analysis of the Italian Manosphere on Facebook: Methodological Approach

Network Construction

- **Strategy:** *Snowball Sampling* procedure:
 - **Seed Identification:** Select initial seed nodes (Facebook pages) known to belong to manosphere communities.
 - **Expansion:** Identify connected nodes by tracking interactions such as mentions, likes, comments, and page follows.
 - **Iteration:** Continue the sampling process until the network growth stabilizes or reaches a predefined limit.

Classification of the Facebook pages in the network

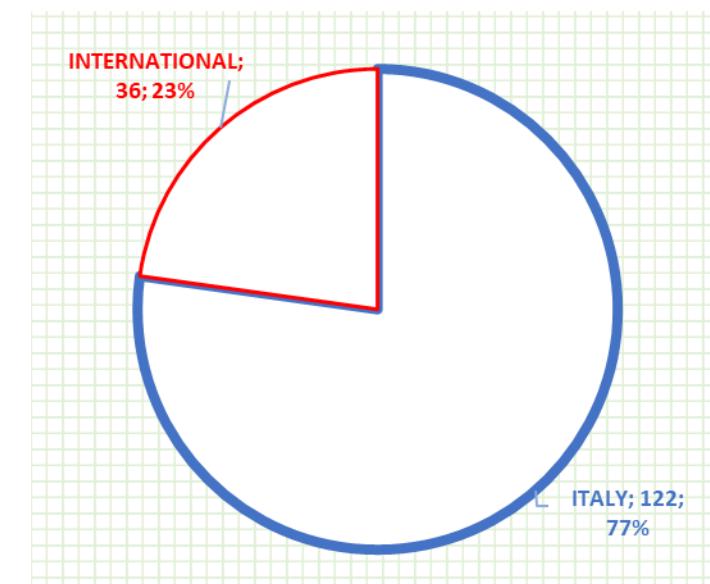
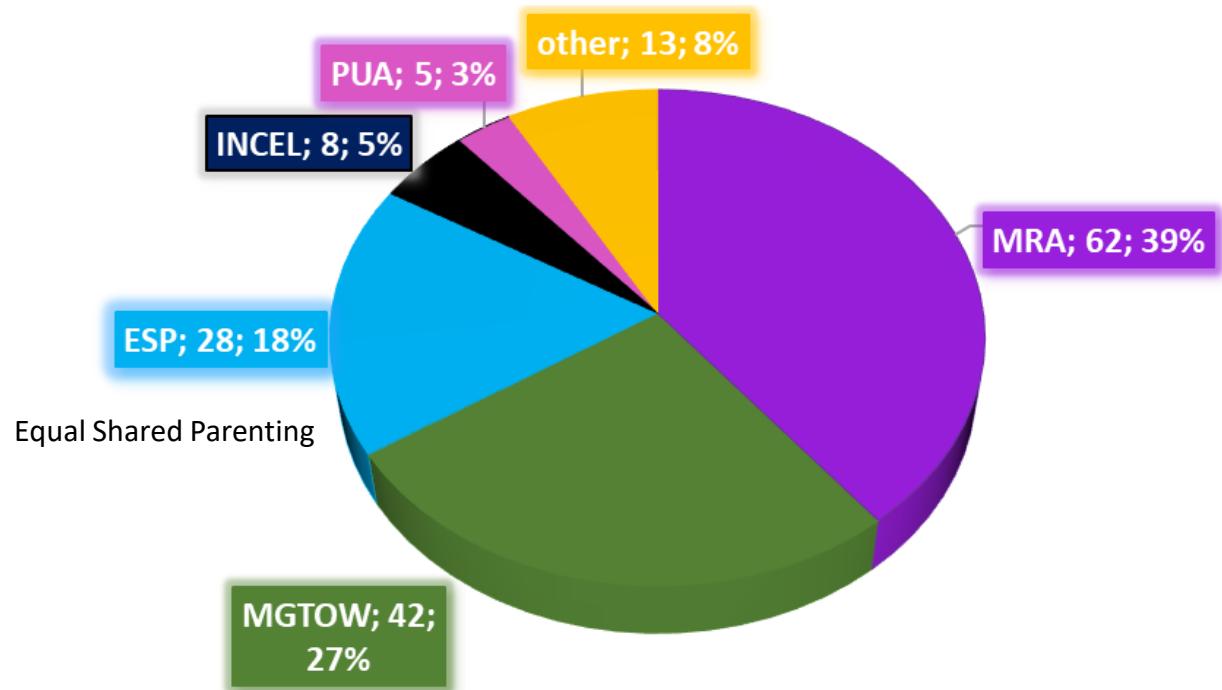
- **Categorization Method:** each Facebook page is assigned to one categories based on page content, descriptions, and primary topics.

Network

✓ Seed nodes:

- **INCEL**: *Il Redpillatore, Vita da Brutto - The Red Pill Never Dies*
- **MGTOW**: *MGTOW, Coscienza Maschile*
- **MRA**: *Diritti Maschili - Equità e Umanità, Resistenza Maschile*
- **PUA**: *Essere Uomo*

✓ Number of Facebook pages in the identified network: **158**



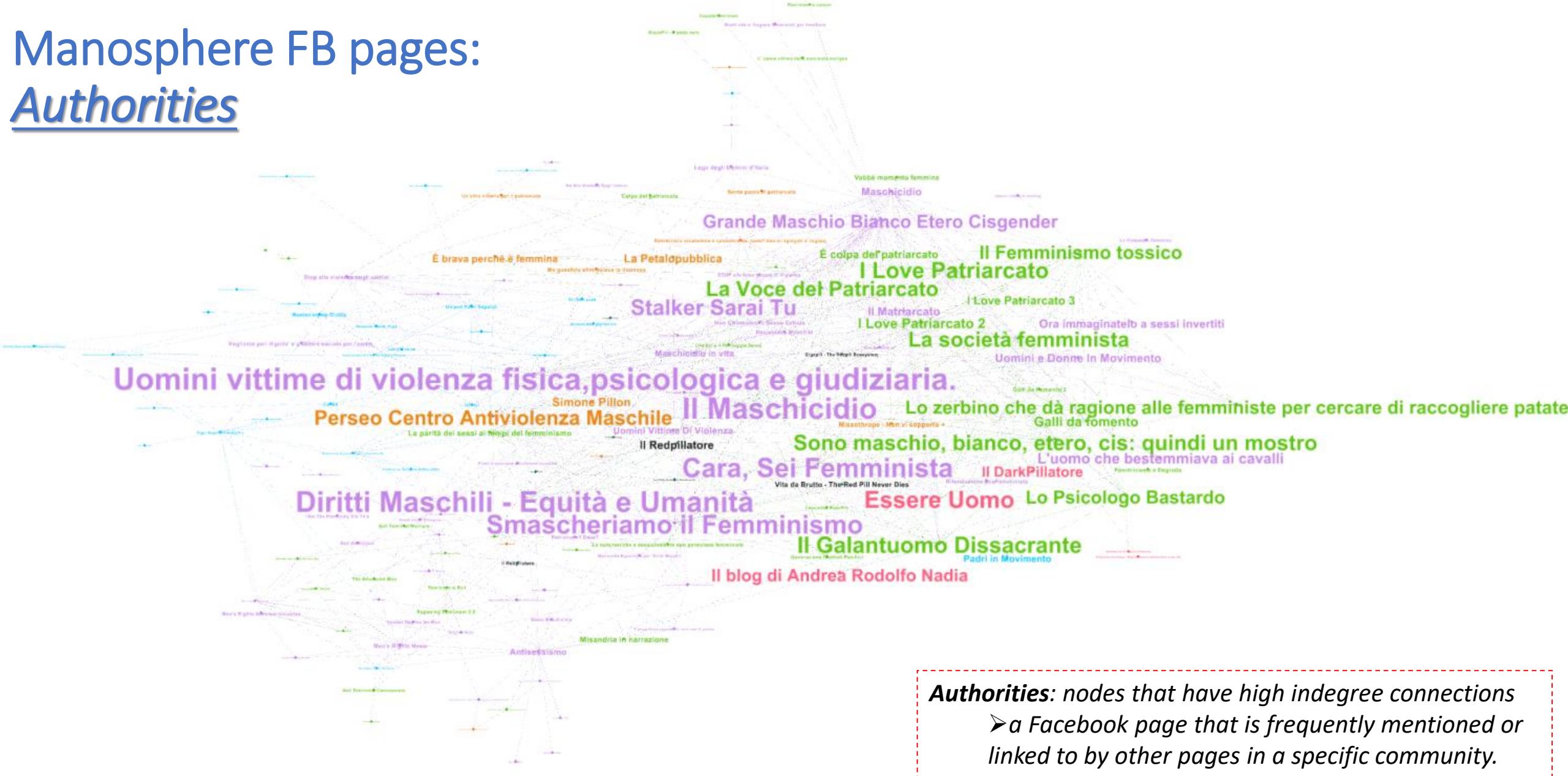
The FB manosphere network



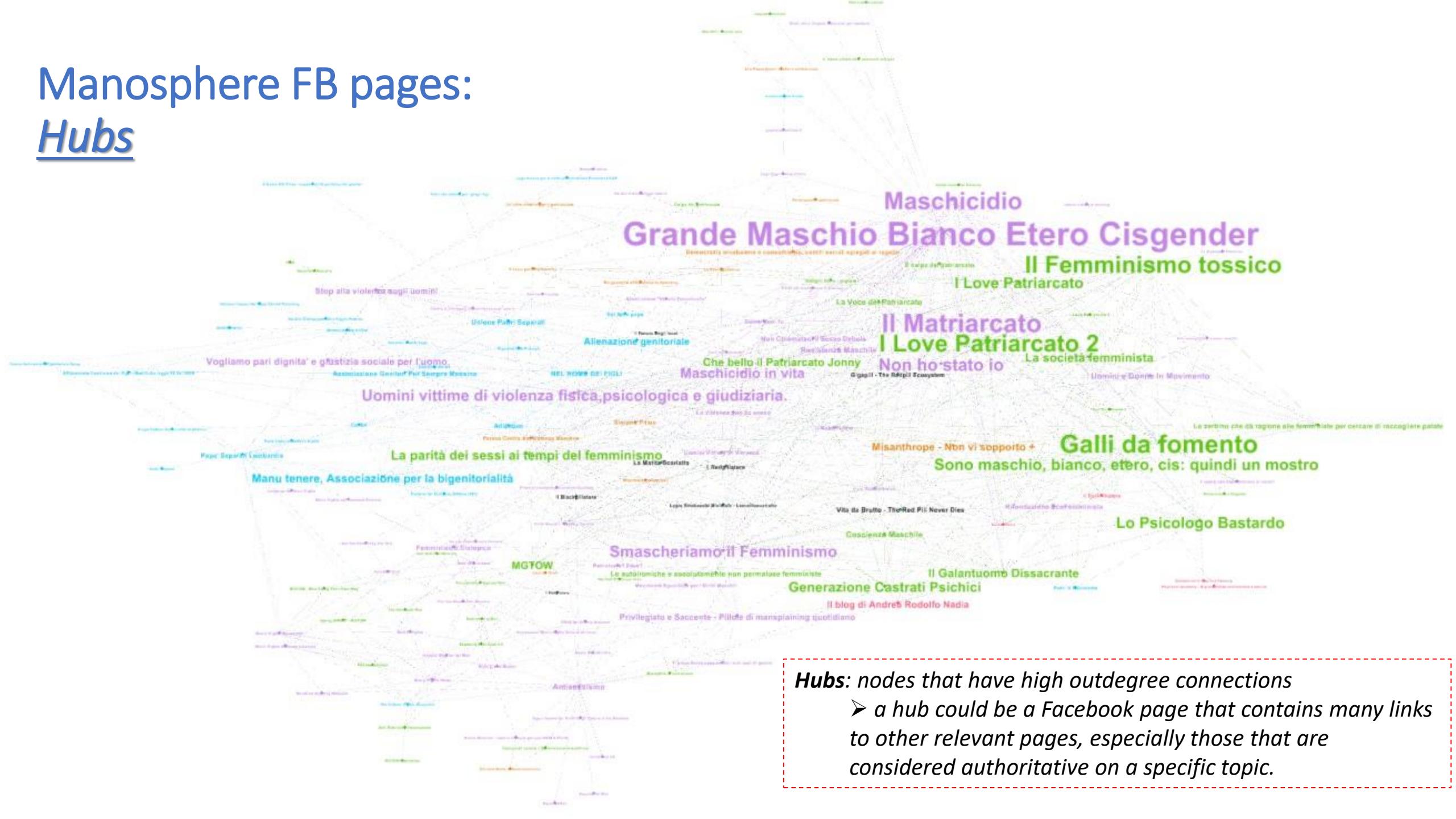
Italian and International FB pages



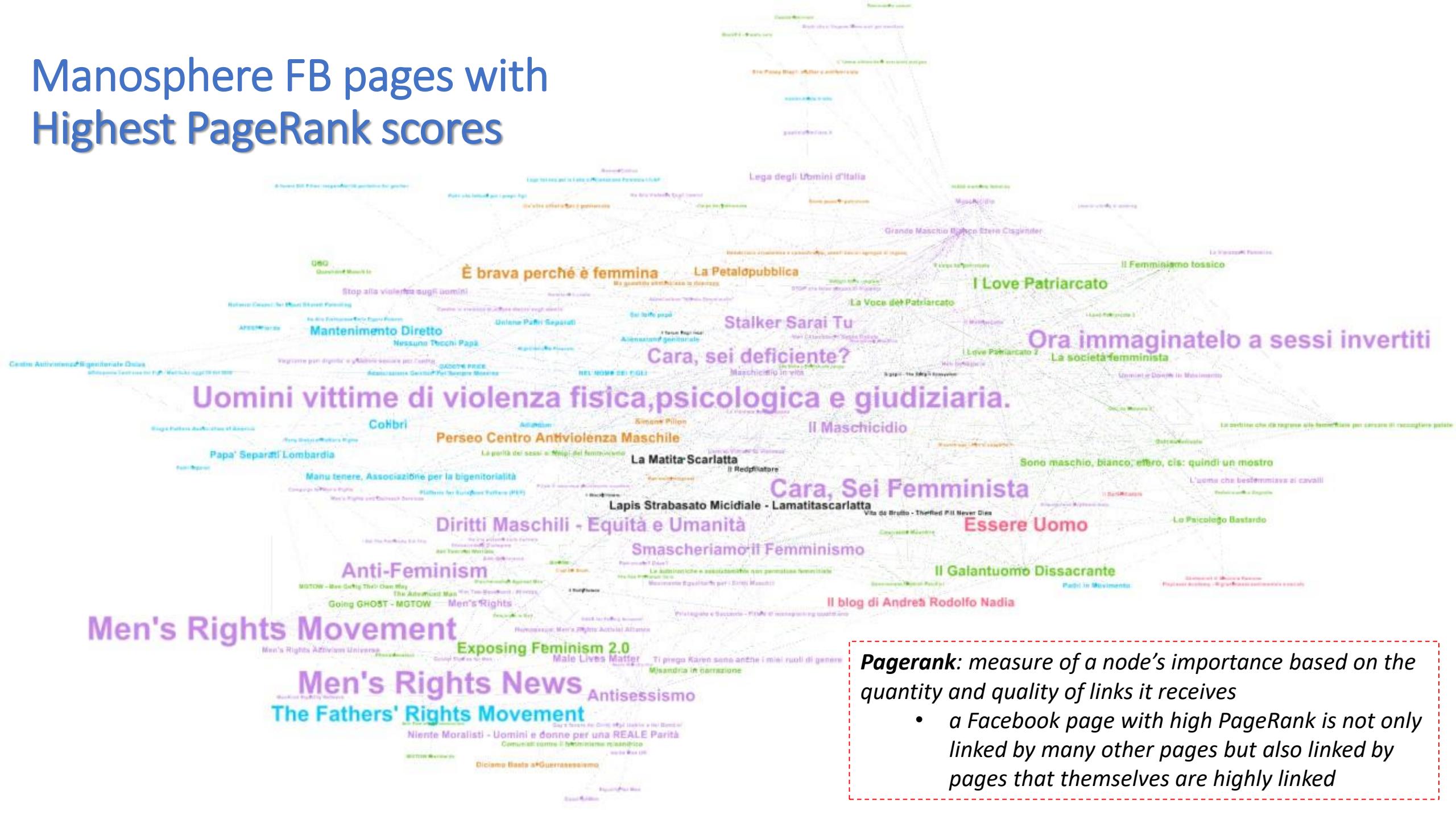
Manosphere FB pages: Authorities



Manosphere FB pages: Hubs



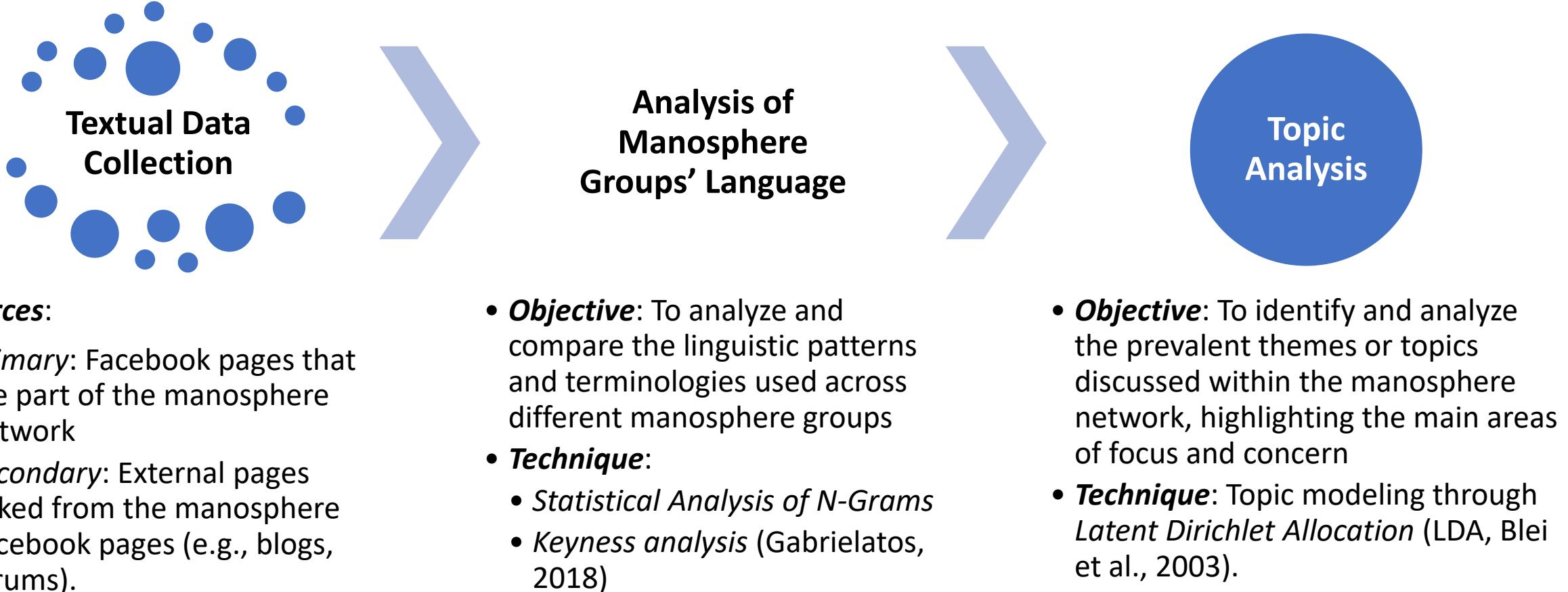
Manosphere FB pages with Highest PageRank scores



Pagerank: measure of a node's importance based on the quantity and quality of links it receives

- a Facebook page with high PageRank is not only linked by many other pages but also linked by pages that themselves are highly linked

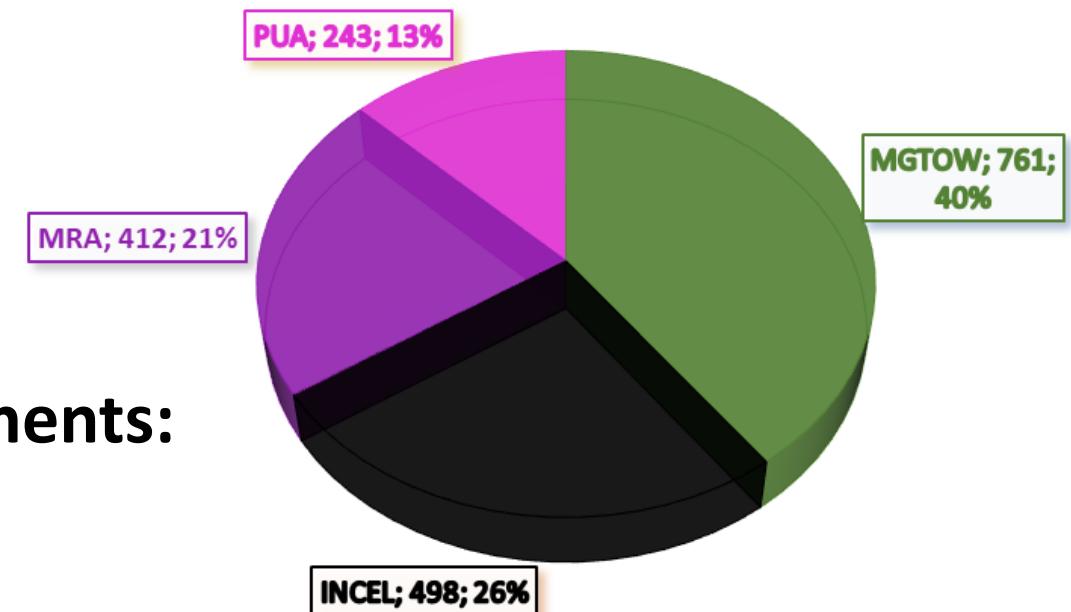
Content Analysis of Manosphere discourses: Methodological Approach



Textual data sources

MRA	412	MGTOW	761	INCEL	498	PUA	243
Affidamento Condiviso dei Figli - Modifiche legge 54 del 2006	3	Comunisti contro il femminismo misandrico	3	Class Alfa	18	Essere Uomo	13
Alternativa Antisessista (il nazifemminismo impera)	27	Coscienza Maschile	86	Il Forum degli Incel	335	Gentlemind di Maurizio Romano	22
Altrosenso	7	È colpa del patriarcato	7	Il Forum dei brutti	47	Il blog di Andrea Rodolfo Nadia	27
Antisessismo	40	Homodus	21	Il forum dei brutti (is)	28	Il DarkPillatore	21
Cara, sei deficiente?	15	Il Galantuomo Dissacrante	7	Il Redpillatore	33	PlayLover Academy	160
Cara, Sei Femminista	2	Lo Psicologo Bastardo	6	Un brutto forum	29		
Colibri	2	Mgtow Italia	518	Vita da Brutto - The Red Pill Never Dies	8		
Diritti Maschili - Equità e Umanità	113	Questione Maschile	107				
Femminismo Distopico	11	Redpill Italia - mgtow	6				
Grande Maschio Bianco Etero Cisgender	4						
I Love Patriarcato	3						
Il Maschicidio	17						
Il Matriarcato	20						
La società femminista	11						
La Violenza è Femmina	1						
Lega degli Uomini d'Italia	15						
L'uomo che bestemmiava ai cavalli	3						
Misandria in narrazione	3						
Smascheriamo il Femminismo	12						
Stalker Sarai Tu (La Fionda)	3						
Uomini e Donne In Movimento	100						

Textual documents:
1914



Unigrams

A dense cloud of Italian words related to gender and society, including 'uomini', 'donne', 'maschile', 'femminile', and various social and political terms.

Bigrams

A dense cloud of Italian bigrams, such as 'uomini donne', 'maggior_parte', 'punto_vista', and 'conquistare_ragazza'.

The vocabulary of the Manosphere

Trigrams

A dense cloud of Italian trigrams, including 'messaggio_sabato_sera', 'capiere_linguaggio_corpo', 'stragrande_maggioranza_donne', and 'teoria_dominazione_maschile'.

Four-grams

A dense cloud of Italian four-grams, including 'teoria_violenza_genere_dominazione', 'condizione_subordinazione_sociale_umana', and 'misandrico_parte_problema_credi'.

A dense cloud of Italian four-grams, including 'incitatrici_faide_sangue_donne', 'going_their_own_way', and 'vittime_maschili_violenza_femminile'.

linguaggio_corpo_segnali_attrazione

A dense cloud of Italian four-grams, including 'superare_delusione_d_amore', 'grande_uomo_grande_donna', and 'realmente_puoi_farne_meno'.

Identifying Salient Words and Phrases across Groups: Keyness analysis

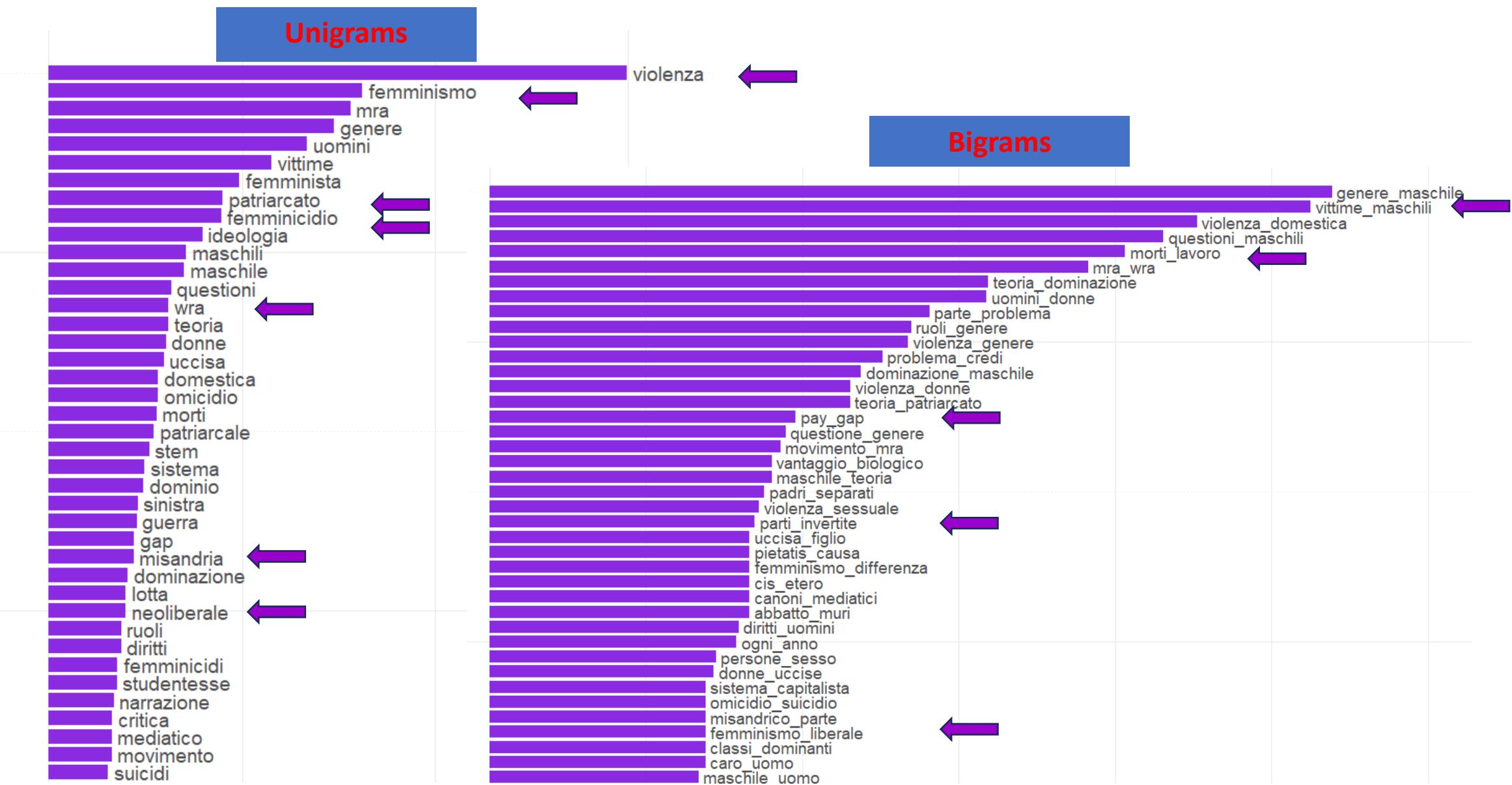
Objective: comparing the textual content of each manosphere subgroup against the remaining groups to identify terms that are uniquely salient in each subgroup.

Statistical Method: *Keyness analysis* (Gabrielatos, 2018) is a technique used to identify words or phrases (also known as "keywords") that occur significantly more frequently in one sub-corpus compared to the other sub-corpora.

Keyness Score: The keyness score is calculated using statistical measures such as chi-square test. These measures assess whether the observed frequency of a word in the target corpus is statistically different from what would be expected based on its frequency in the reference corpus.

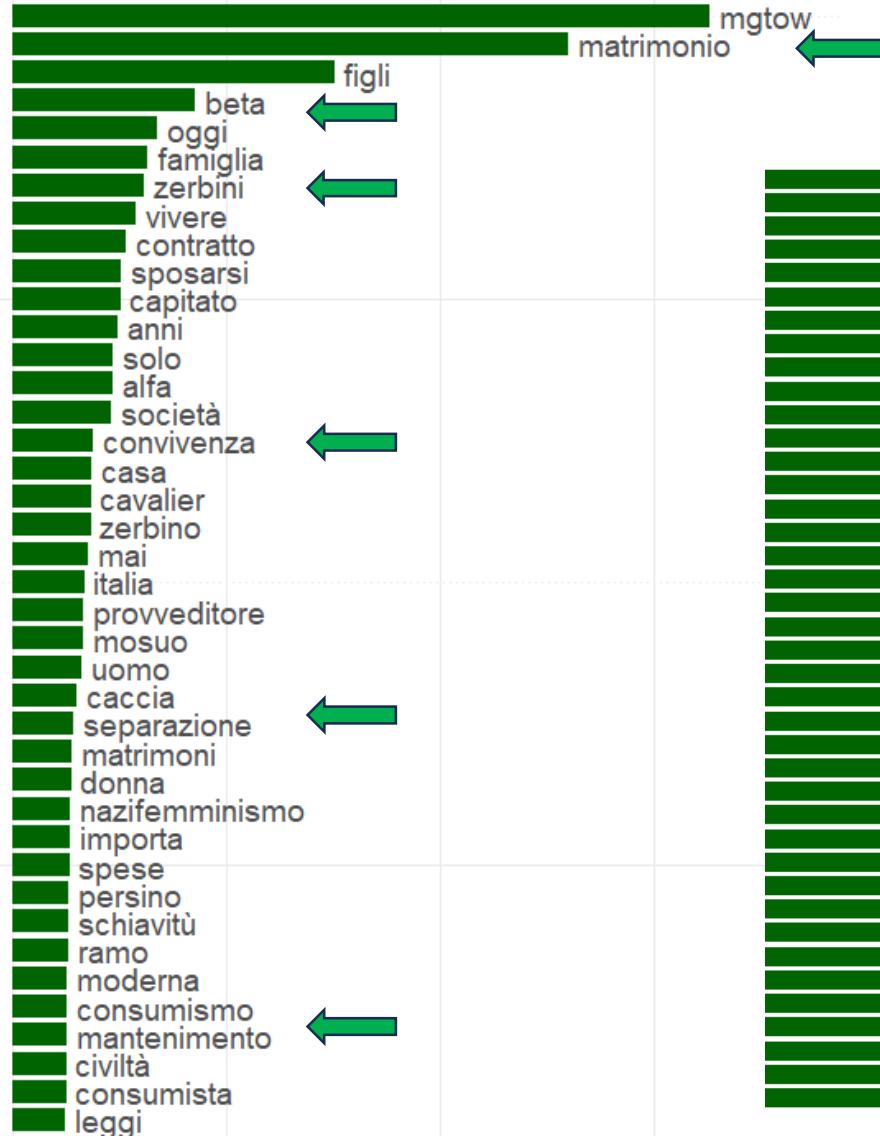
Statistical Significance: a statistical threshold (e.g. value for the chi-square test) is applied to determine which words or phrases are significantly overrepresented in each subgroup.

Keyness: MRA

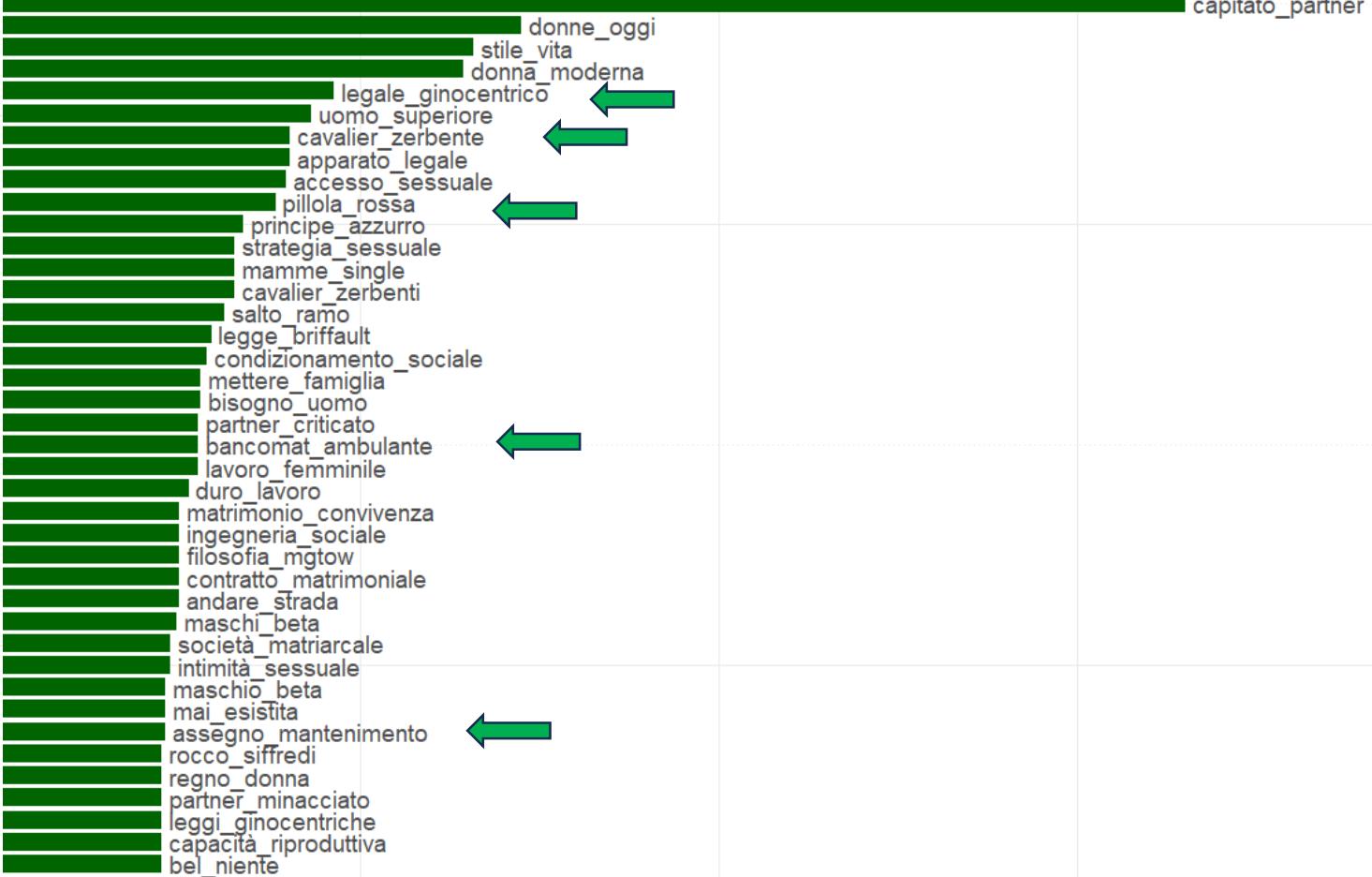


Keyness: MGTOW

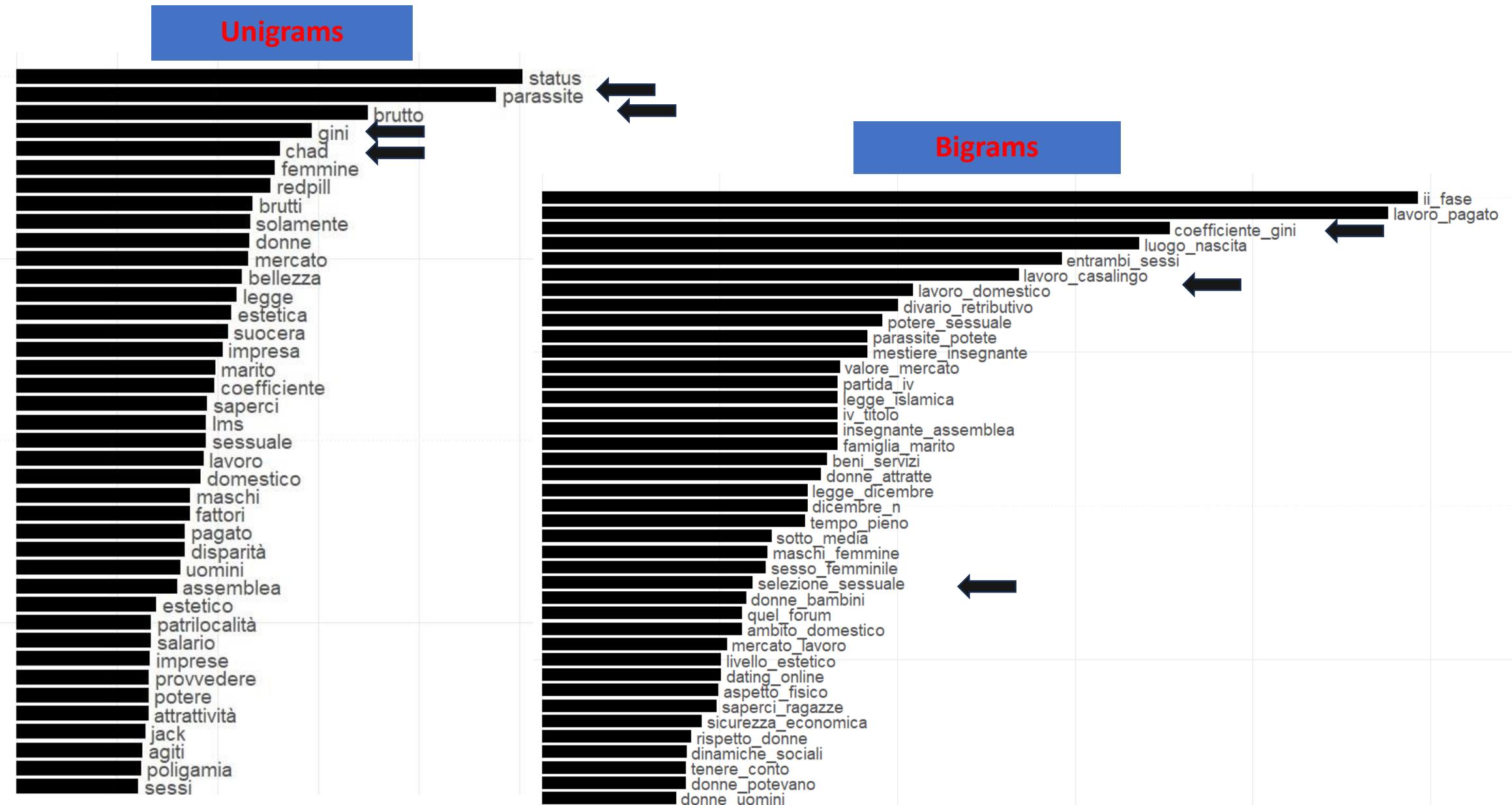
Unigrams



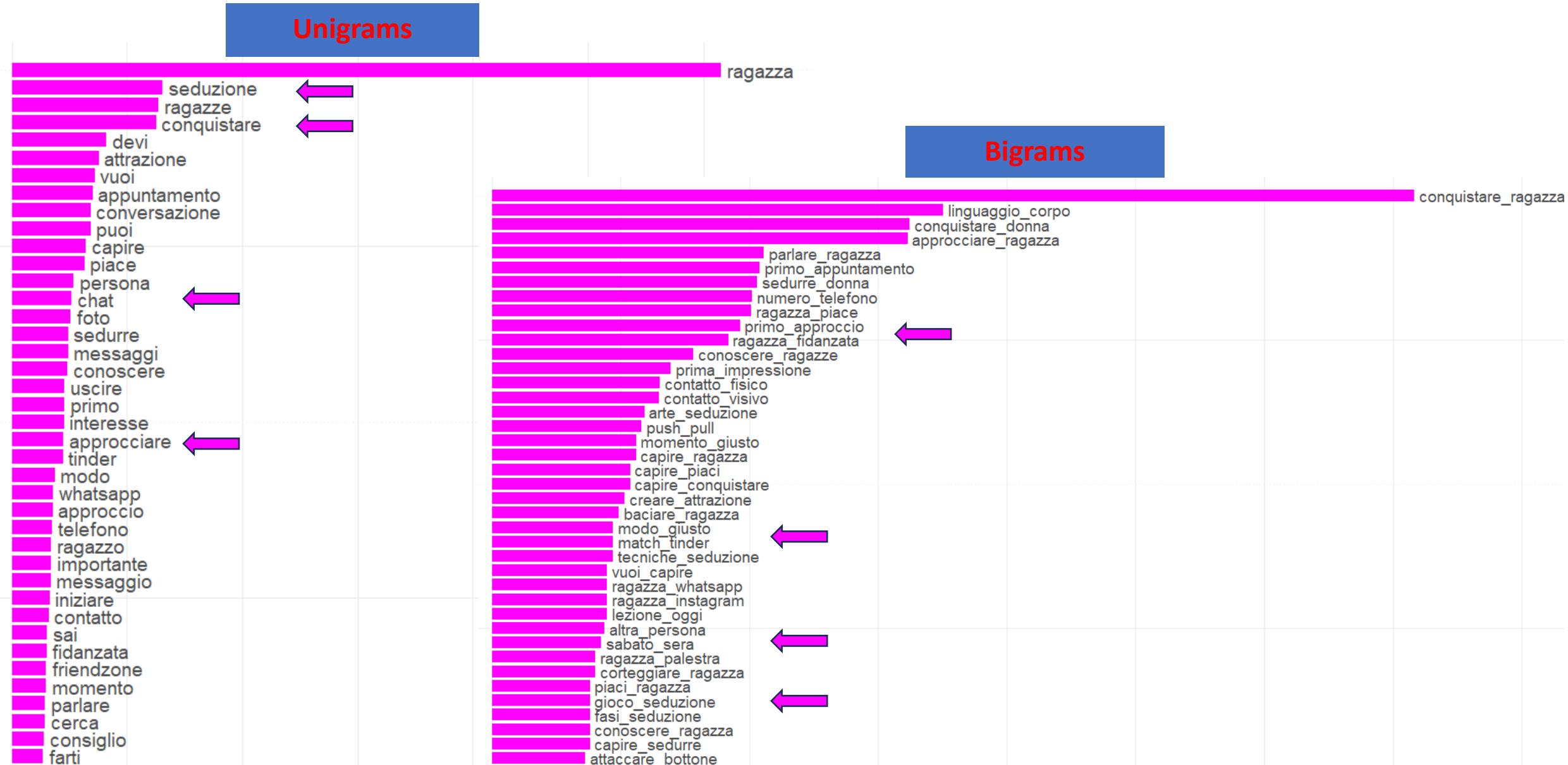
Bigrams



Keyness: INCEL



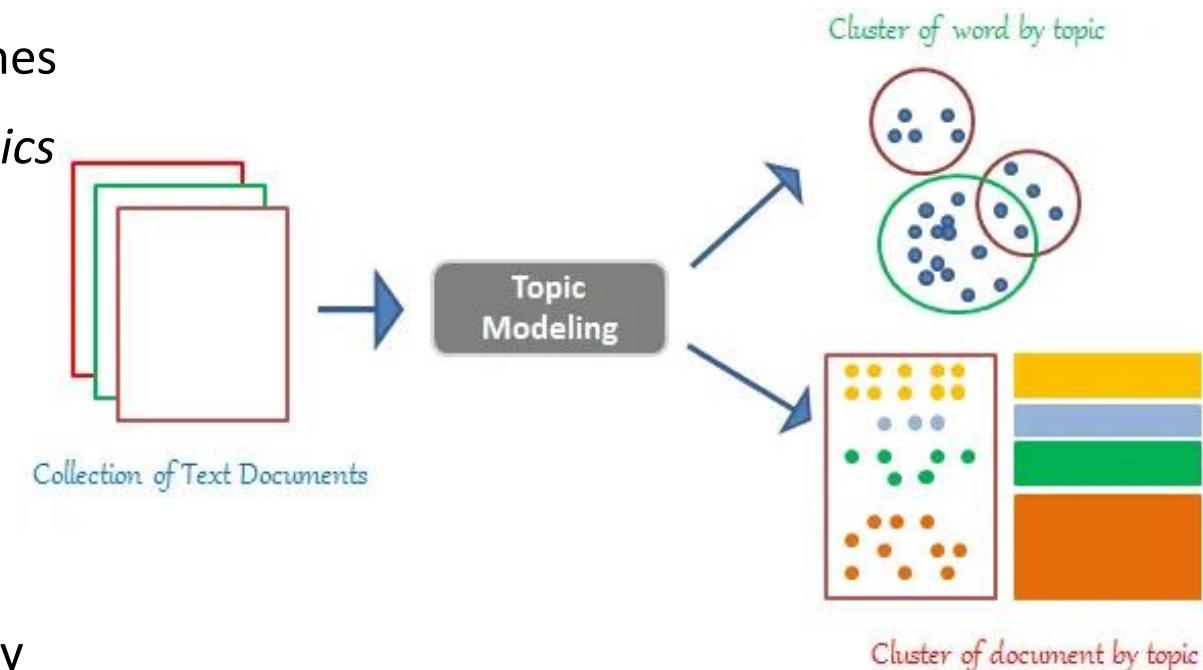
Keyness: PUA

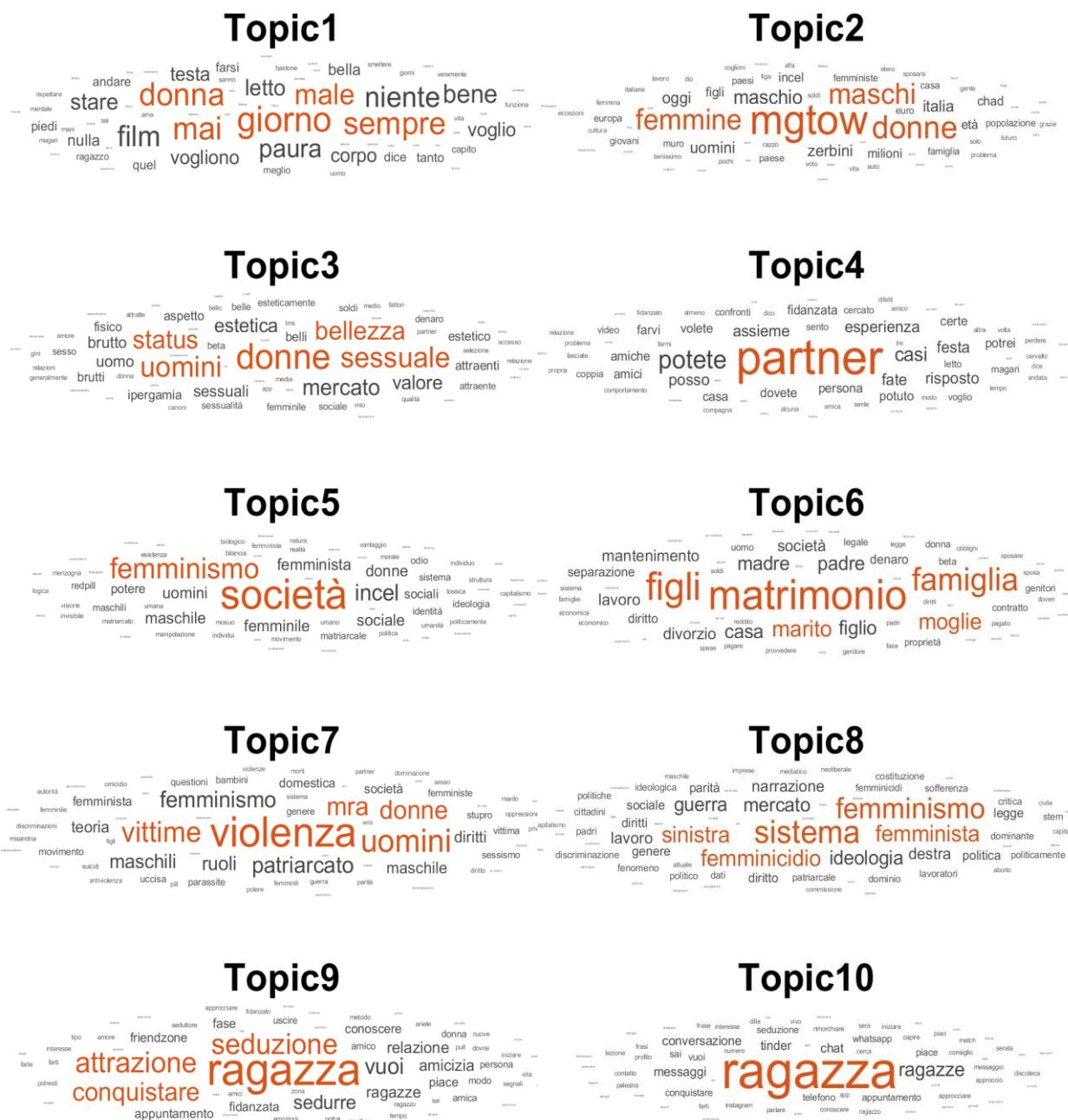


Topic analysis

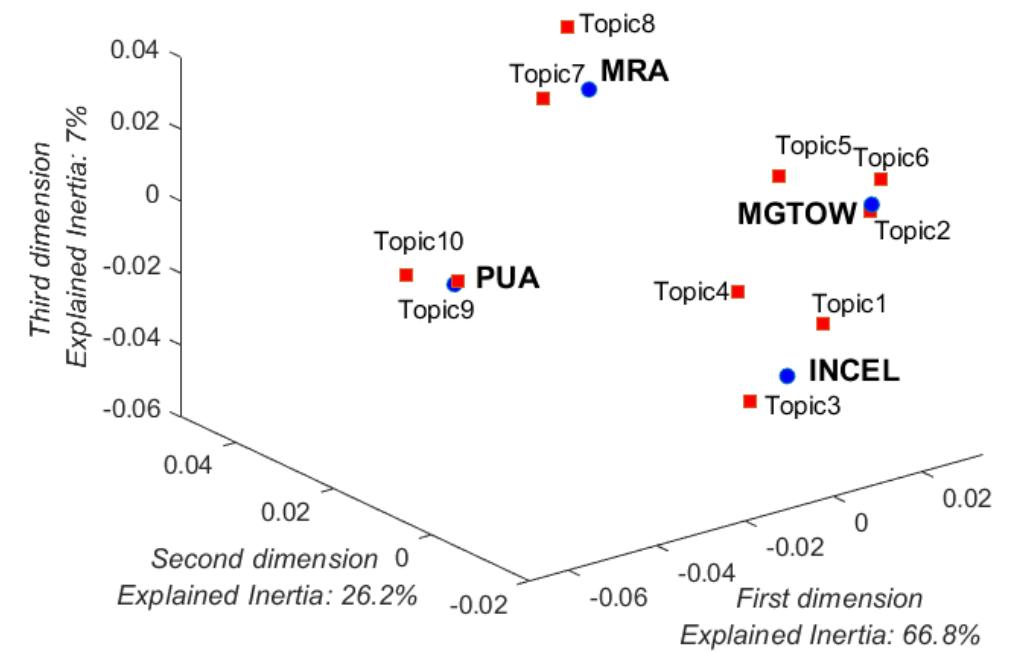
Latent Dirichlet Allocation

- LDA Latent Dirichlet Allocation (LDA) is a statistical method used in natural language processing (NLP) and machine learning for uncovering **hidden thematic structures (*TOPIC*)** In large text corpora, making it easier to understand and analyze unstructured data.
- LDA is a type of generative probabilistic model that assumes
 - *each document is viewed as a mixture of multiple topics*
 - *each topic is a mixture of words.*
- LDA infers
 - which topics are present in the documents
 - which words are most associated with each topic.
- Algorithms like Gibbs sampling or variational inference, which iteratively estimate the topic distributions until they best explain the observed data.





Identified topics and association to manosphere groups

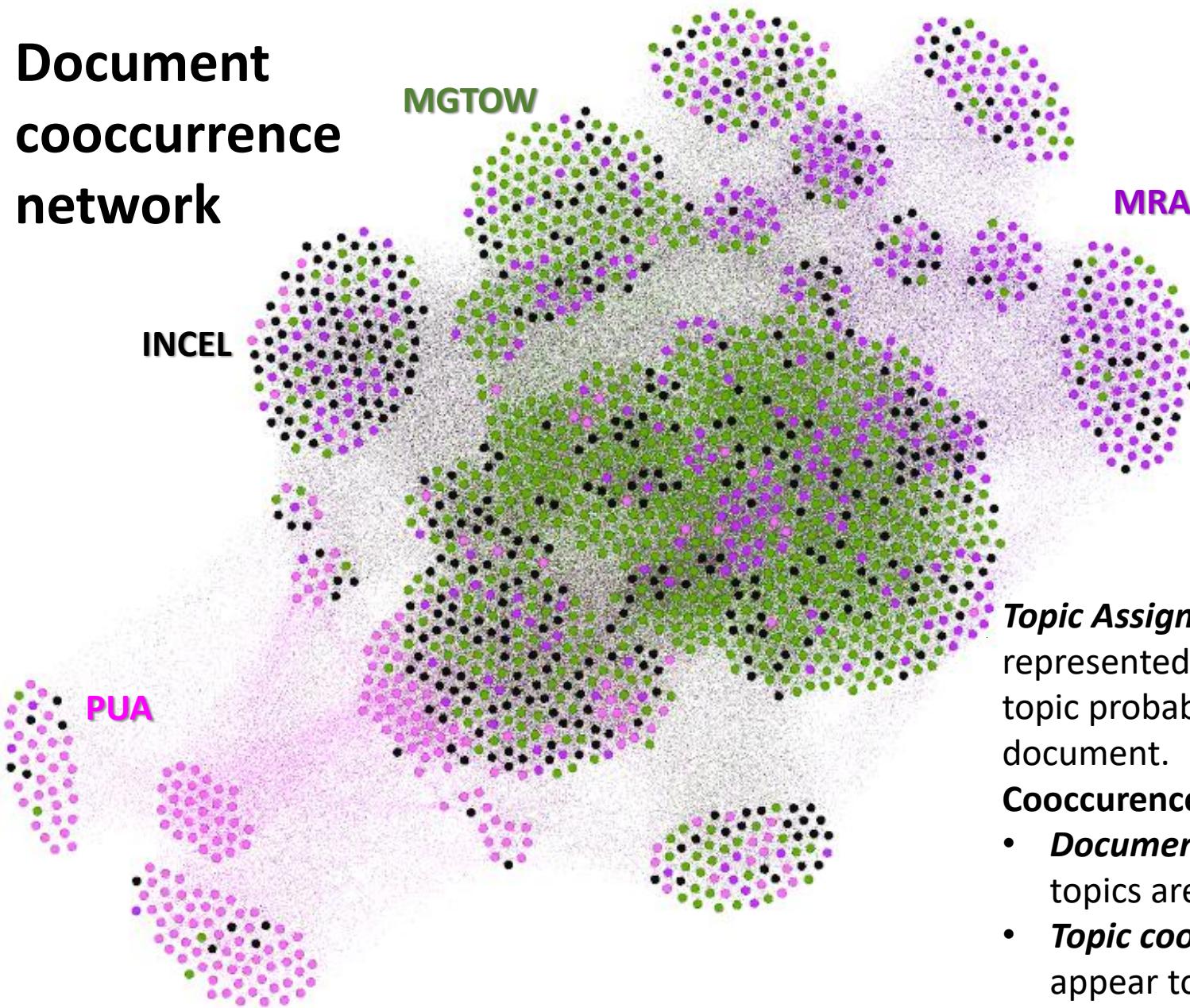


Methodology

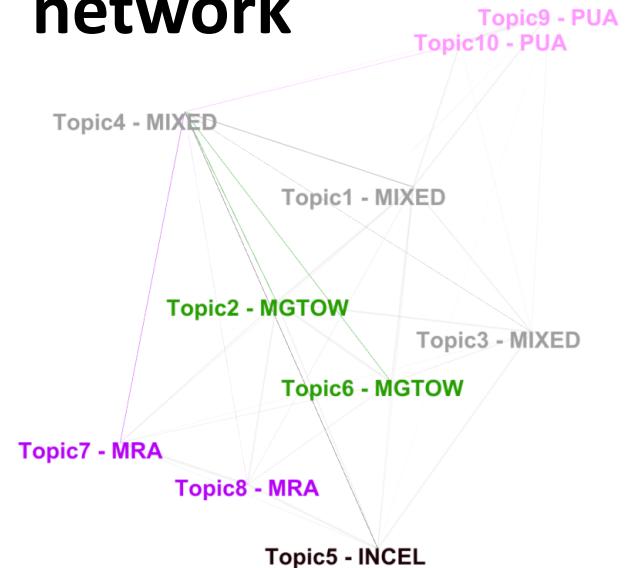
Topic Assignment: The topic with the highest document-topic probability was assigned to each document, ensuring that each document is linked to its most relevant topic.

Correspondence Analysis: A correspondence analysis (Greenacre, 2007) was conducted on the contingency table of manosphere groups and topics to examine the relationships and associations between specific topics and the different manosphere groups.

Document cooccurrence network



Topic cooccurrence network



Topic Assignment: Given that in LDA each document is represented as a mixture of topics, topics with a document-topic probability greater than 0.25 were assigned to each document.

Cooccurrence networks:

- **Document cooccurrence network:** illustrates how many topics are shared by each pair of documents.
- **Topic cooccurrence network:** illustrates how often topics appear together within the same documents, allowing for the identification of patterns

Pubblicazioni

- Tontodimamma A., Anzani S., Stranisci M.A., Basile V., Ignazzi E., Fontanella L. (2023) *An experimental annotation task to investigate annotators' subjectivity in a Misogyny dataset*. In ASA 2022 Data-Driven Decision Making, Firenze University Press, Genova University Press.
- Tontodimamma A., Ignazzi E., Anzani S., Fontanella L., Di Zio S. (2023), *An Experimental Annotation Task Investigating Annotator Agreement Within a Misogynistic Dictionary and Corpus*. In "Statistical methods for evaluation and quality: techniques, technologies and trends (T3)". Book of Short Papers IES 2023, Il Viandante Edizioni.
- Ignazzi E., Sarra A., Fontanella L. (2023) *Exploring Misogyny through Time: From Historical Origins to Modern Complexities*. In Boero M., Boffa D. "Gender Differences in Language, Rights and Society", Series Philosophies of Communication, Il Sileno Edizioni.
- Ignazzi E., Coll M., Del Gobbo E., Chulvi B., Rosso P., Fontanella L. (2024) *Characterizing Misogyny in Italian Online Discourse: Consensus and Disagreement in a New Dataset of Social Media Comments*. In Proceedings of the 17th Journées Internationales d'Analyse Statistique des Données Textuelles (JADT24), Presses universitaires de Louvain.
- Ignazzi E., Maretti M., Fontanella L. (2025) *The Italian Manosphere: Composition, Structure, and Functions of a Digital Network*. Social Inclusion 2025, Volume 13, Article 9341.
- Del Gobbo E., Ignazzi E., Cafarelli B., Fontanella L. (2025) *Misogynistic content detection over social networks*. In "Methodological and Applied Statistics and Demography III", Springer
- Fontanella L., Cucco A., del Gobbo E., Ignazzi E., Fontanella S. (forthcoming) *Improving online misogyny detection: the role of human annotation and explanation-guided learning*, Statistica Applicata - Italian Journal of Applied Statistics
- Del Gobbo E., Cucco A., Fontanella L. (forthcoming) *Identification of Misogynistic Accounts on Twitter through Graph Convolutional Networks*. Springer.