ISSN 2704-601X (PRINT) - ISSN 2704-5846 (ONLINE)

***Scientific Program Committee***

Enrico di Bella (University of Genoa, co-chair)
Luigi Fabbris (ASA, Tolomeo Studi e Ricerche, Padua, co-chair)
Corrado Lagazio (University of Genoa, co-chair)
Fabrizio Antolini (University of Teramo)
Rossella Berni (University of Florence)
Bruno Bertaccini (University of Florence)
Silvia Maria Biasotti (CNR)
Gian Lorenzo Boracchia (Liguria Region)
Eugenio Brentari (University of Brescia)
Maurizio Carpita (University of Brescia)
Giulia Cavrini (University of Bolzano-Bozen)
Alessandro Celegato (AICQ-AISS, PSV Project Service and Value)
Giuliana Coccia (ASVIS)
Cristina Davino (Federico II University of Naples)
Giulia De Candia (ISTAT)
Loretta Degan (Galgano Group)
Tonio Di Battista (G. D'Annunzio University of Chieti and Pescara)
Angela Maria Digrandi (CNR-National Council for Research)
Simone Di Zio (G. D'Annunzio University of Chieti and Pescara)
Patrizia Elli (Assirm)
Benito Vittorio Frosini (Sacred Heart Catholic University of Milan)
Antonio Giusti (University of Florence)
M. Gabriella Grassia (Federico II University of Naples)
Salvo Ingrassia (University of Catania)
Enrico Ivaldi (University of Genoa)
Paolo Mariani (Bicocca University of Milan)
Stefania Mignani (University of Bologna)
Marta Nai Ruscone (University of Genoa)
Chiara Parretti (University Guglielmo Marconi of Rome)
Francesco Palumbo (Federico II University of Naples)
Francesco Porro (University of Genoa)
Fabio Rapallo (University of Genoa)
Alessandra Petrucci (University of Florence)
Alfonso Piscitelli (Federico II University of Naples)
Alessandra Righi (Istat)
Fabrizio Ruggeri (CNR-National Research Council)
Giorgio Tassinari (University of Bologna)
Nicola Tedesco (University of Cagliari)
Venera Tomaselli (University of Catania)
Domenico Vistocco (Federico II University of Naples)
Mariangela Zenga (Bicocca University of Milan)

***Local Program Committee***

Enrico di Bella (University of Genoa, co-chair)
Corrado Lagazio (University of Genoa, co-chair)
Gian Lorenzo Boracchia (Regione Liguria)
Fabrizio Culotta (University of Genoa)
Giulia De Candia (ISTAT)
Enrico Ivaldi (University of Genoa)
Marta Nai Ruscone (University of Genoa)
Fabio Rapallo (University of Genoa)
Eva Riccomagno (University of Genoa)
Francesco Porro (University of Genoa)
Sara Preti (University of Genoa)

# ASA 2022
# Data-Driven Decision Making

BOOK OF SHORT PAPERS

edited by
Enrico di Bella
Luigi Fabbris
Corrado Lagazio

FIRENZE UNIVERSITY PRESS | GENOVA UNIVERSITY PRESS
2023

ASA 2022 Data-Driven Decision Making : book of short papers / edited by Enrico di Bella, Luigi Fabbris, Corrado Lagazio. – Firenze : Firenze University Press ; Genova : Genova University Press, 2023.
(Proceedings e report ; 134)

https://books.fupress.com/isbn/9791221501063

ISSN 2704-601X (print)
ISSN 2704-5846 (online)
ISBN 979-12-215-0106-3 (PDF)
ISBN 979-12-215-0107-0 (XML)
DOI 10.36253/979-12-215-0106-3

Cover graphic design: Lettera Meccanica SRLs
Front cover image: Orizzonte del centro di Genova, Italia © sepavo|123rf.com

Volume published with the support of Università di Genova, Scuola di Scienze Sociali - Dipartimento di Scienze Politiche e Internazionali.

Partners

*Peer Review Policy*
Peer-review is the cornerstone of the scientific evaluation of a book. All FUP publications undergo a peer-review process by external experts under the responsibility of the Editorial Board and the Scientific Boards of each series (DOI 10.36253/fup_best_practice.3).

*Referee List*
In order to strengthen the network of researchers supporting FUP evaluation process, and to recognise the valuable contribution of referees, a Referee List is published and constantly updated on FUP - Genova University Press's website (DOI 10.36253/fup_referee_list).

*Firenze University Press Editorial Board*
M. Garzaniti (Editor-in-Chief), M.E. Alberti, F. Vittorio Arrigoni, E. Castellani, F. Ciampi, D. D'Andrea, A. Dolfi, R. Ferrise, A. Lambertini, R. Lanfredini, D. Lippi, G. Mari, A. Mariani, P.M. Mariano, S. Marinai, R. Minuti, P. Nanni, A. Orlandi, I. Palchetti, A. Perulli, G. Pratesi, S. Scaramuzzi, I. Stolzi.

The online digital edition is published in Open Access on www.fupress.com.

Content license: except where otherwise noted, the present work is released under Creative Commons Attribution 4.0 International license (CC BY 4.0: http://creativecommons.org/licenses/by/4.0/legalcode). This license allows you to share any part of the work by any means and format, modify it for any purpose, including commercial, as long as appropriate credit is given to the author, any changes made to the work are indicated and a URL link is provided to the license.

Metadata license: all the metadata are released under the Public Domain Dedication license (CC0 1.0 Universal: https://creativecommons.org/publicdomain/zero/1.0/legalcode).

© 2023 Author(s)

Published by Firenze University Press and Genova University Press

Firenze University Press
Università degli Studi di Firenze
via Cittadella, 7, 50144 Firenze, Italy
www.fupress.com

*This book is printed on acid-free paper*
*Printed in Italy*

# Table of contents

Assessing the predictive capability of Invalsi tests on high school final mark 11
*Silvia Bacci, Bruno Bertaccini, Alessandra Petrucci, Valentina Tocchioni*

Profiling students' satisfaction towards university courses with a latent class approach 17
*G. Damiana Costanzo, Michelangelo Misuraca, Angela Cosca*

The relation between students' educational performances and their access test results: a focus on an Italian case 23
*Matteo Corsi, Luca Persico, Sara Preti, Agnese Sechi*

Structure and dynamics of immigration in the municipalities of northwestern Italy 29
*Simona Ballabio, Arianna Carra, Flavio Verrecchia, Alberto Vitalini*

Are Italian youngsters adequately equipped for an after-pandemic upswing? 35
*Luigi Bollani, Simone Di Zio, Luigi Fabbris*

How Italians coped with COVID-19 lockdown: evidence from a survey promoted through social networks 41
*Margherita Silan, Riccardo Bellide*

Official statistics for measuring the sustainability of tourism: the UNWTO initiative 47
*Emanuela Recchini*

Misinformation and Disinformation in Statistical Methodology for Social Sciences: causes, consequences, and remedies 53
*Giulio Giacomo Cantone, Venera Tomaselli*

The impact of economic insecurity on life satisfaction among German citizens 59
*Demetrio Panarello, Gennaro Punzo*

Cultural and sensorial correlates of Trebbiano wine consumption 65
*Luigi Fabbris, Alfonso Piscitelli*

Tourism and territorial economy: beyond satellite accounting 71
*Fabrizio Antolini, Antonio Giusti, Francesca Petrei*

Short-term forecasts on time series for tourism in Lombardy 77
*Andrea Marletta, Roberta Rossi, Elena Diceglie*

Interventions for non-self-sufficiency – Focus on care and social policies in South Tyrol 83
*Giulia Cavrini, Nadia Paone, Evan Tedeschi*

The territorialisation of the 2030 Agenda: a multilevel approach 89
*Raffaele Attanasio, Manlio Calzaroni, Alessandro Ciancio, Federico Olivieri, Giovanni Siciliano*

Sustainable development goals: classifying European countries through self-organizing maps 95
*Cristina Davino, Nicola D'Alesio*

FUP Best Practice in Scholarly Publishing (DOI 10.36253/fup_best_practice)

Enrico di Bella, Luigi Fabbris, Corrado Lagazio (edited by), *ASA 2022 Data-Driven Decision Making. Book of short papers*, © 2023 Author(s), CC BY 4.0, published by Firenze University Press and Genova University Press, ISBN 979-12-215-0106-3, DOI 10.36253/979-12-215-0106-3

Individual and social aspects of after-Covid-19 pandemic depression 101
*Pasquale Anselmi, Daiana Colledani, Simone Di Zio, Luigi Fabbris, Egidio Robusto*

Spread of Covid-19 epidemic in Italy between March 2020 and February 2021: empirical evidence at provincial level 107
*Fabrizio Antolini, Samuele Cesarini, Francesco Giovanni Truglia*

New perspectives for the quality of sub-municipal data with the Italian permanent population and housing census 113
*Giancarlo Carbonetti, Stefano Daddi, Giampaolo De Matteis, Marco Di Zio, Davide Fardelli, Raffaele Ferrara, Fabio Lipizzi, Enrico Orsini*

The Land Cover/Use Code of the new Istat Census cartography[1] 119
*Stefano Mugnoli, Alberto Sabbi, Fabio Lipizzi*

Trusted smart statistics: new statistics for decision makers. Istat's experience 125
*Massimo De Cubellis, Gerarda Grippo*

Quality of life in Health Care: focus on patients 129
*Laura Benedan, Angela Digrandi, Paolo Mariani, Cinzia Pilo, Mariangela Zenga*

Access to emergency care services and inequalities in living standards: some evidence from two Italian northern regions 135
*Andrea Marino, Marco Pesce, Raffaella Succi*

Population ageing and sustainability in South Tyrol: measuring the economic implications of an ageing society 141
*Giulia Cavrini, Elisa Cisotto, Alex Weissensteiner*

Job loss and financial struggle among the older age groups in 2021: Lessons from the European Union 145
*Demetrio Panarello, Giorgio Tassinari*

On the use of auxiliary information in spatial sampling 151
*Chiara Bocci, Emilia Rocco*

Measures of interrater agreement when each target is evaluated by a different group of raters 157
*Giuseppe Bove*

A Natural Language Processing approach to measuring expertise in the Delphi-based scenarios 163
*Yuri Calleo, Simone Di Zio, Francesco Pilla*

Exploring Globalization with Cosmopolitics 169
*Maria Serena Causo, Erika Cerasti, Fabrizio De Fausti, Monica Scannapieco*

Professional choices and personal values: Similarities and differences between Schein's career anchors and Schwartz basic values 175
*Maria Cristiana Martini, Aldo Arra*

Factors affecting tertiary education decisions of immigrants in Italy 181
*Michele Lalla, Patrizio Frederic*

Internet use, feeling of unacceptance and Loneliness: immigrants of first and second generation in Italy 187
*Giovanni Busetta, Maria Gabriella Campolo, Antonia Cava*

A composite indicator to measure regional investment policies on R&D and innovation 193
*Sergio Salamone, Alessandro Faramondi, Stefania Della Queva*

Assessing intimate partner violence in African countries through a model-based composite indicator 197
*Anna Maria Parroco, Micaela Arcaio*

Students' feedback on the digital ecosystem: a structural topic modeling approach 203
*Annalina Sarra, Adelia Evangelista, Tonio Di Battista*

The digitization of the private sector. A non-aggregative method to monitor the NRRP agenda at macro-area level 209
*Susanna Traversa, Enrico Ivaldi*

Digital.VET: an innovative approach for teaching and training 215
*Teresa Maltese, Maria Santarcangelo, Vito Santarcangelo, Diego Sinitò, Aneta Poniszewska-Marańda, Jure Šuligoj, Alcidio Jesus, Elisardo Sanchis*

The joint estimation of accuracy and speed: An application to the INVALSI data 221
*Luca Bungaro, Marta Desimoni, Mariagiulia Matteucci, Stefania Mignani*

Ammonia emissions and fine particulate matter: some evidence in Lombardy 227
*Alessandro Fusta Moro, Matteo Salis, Andrea Zucchi, Michela Cameletti, Natalia Golini, Rosaria Ignaccolo*

On the utility of treating a vineyard against *Plasmopara viticola*: a Bayesian analysis 233
*Lorenzo Valleggi, Federico Mattia Stefanini*

Trust and security in Italy 239
*Silvia Golia*

Topic modeling for analysing the Russian propaganda in the conflict with Ukraine 245
*Maria Gabriella Grassia, Marina Marino, Rocco Mazza, Michelangelo Misuraca, Agostino Stavolo*

The relationship between religiosity, religious coping, and anxieties about the future: a multidimensional analysis on the Evangelical churches of Naples 251
*Maria Gabriella Grassia, Marina Marino, Rocco Mazza, Agostino Stavolo*

An application of the Agency for Digital Italy guidelines and CSA Star self-assessment: A Docustar case study 257
*Pierluigi Calabrese, Paola Lunalbi, Vincenzo Ribaudo, Saverio Crisafulli, Antonio Ruoto, Vito Santarcangelo, Diego Sinitò, Carlo Bonelli, Giuseppe Stella*

Remote working in Italy: Just a pandemic accident or a lesson for the future? 263
*Luigi Bollani, Simone Di Zio, Luigi Fabbris*

Repression of the future-oriented disposition of Italians by a never-ending pandemic 269
*Simone Di Zio, Luigi Fabbris*

Monitoring and evaluation of gender equality policies 275
*Giuliana Coccia, Emanuela Scavalli*

An experimental annotation task to investigate annotators' subjectivity in a misogyny dataset 281
*Alice Tontodimamma, Stefano Anzani, Marco Antonio Stranisci, Valerio Basile, Elisa Ignazzi, Lara Fontanella*

Potential risk of gambling products and online gambling among European adolescents 287
*Elisa Benedetti, Gabriele Lombardi, Rodolfo Cotichini, Sonia Cerrai, Marco Scalese, Sabrina Molinaro*

An Open Data platform for decision making in local public administration 293
*Giuseppe Sindoni, Matteo Massenzio*

Educational mismatch and productivity: evidence from LEED data on Italian firms 299
*Laura Bisio, Matteo Lucchese*

A paradata-driven statistical approach to improve fieldwork monitoring: the case of the Non-Profit Institutions census 305
*Gabriella Fazzi, Manuela Murgia, Alessandra Nuccitelli, Francesca Rossetti, Valentino Parisi, Roberta Piergiovanni, Luigi Arlotta, Maura Giacummo*

Gender INequality Indicator for Academia (GINIA) 311
*Margherita Silan, Giovanna Boccuzzo*

Given *N* Forecasting Models, What To Do? 317
*Fabrizio Culotta*

# Topic modeling for analysing the Russian propaganda in the conflict with Ukraine

Maria Gabriella Grassia, Marina Marino, Rocco Mazza, Michelangelo Misuraca, Agostino Stavolo

## 1. Introduction

The conflict between Ukraine and Russia is changing Europe, which is facing a crisis destined to reshape the internal and external relations of the continent, shifting international balances. As the war in Ukraine continues, Russian propaganda about the conflict evolves. Modern propaganda can be seen as an attempt to influence opinion through the communication of ideas and values of a specific persuasive purpose (Abd Kadir et al. 2014).

A political organisation would want to convince people to concur with the message presented and accept it as their own beliefs, rejecting other point of view. It has been argued that practically all governments use forms of propaganda to bolster their support from other nations and citizenry (Pratkanis and Aronson, 1991).

For this reason, we present the preliminary results of an analysis conducted on the content of online newspapers used as propaganda tools by the Russian government. The selected newspapers create and amplify the narrative of the conflict, conveying information filtered by the Kremlin to advance Putin's campaign on the war. The goal of the work, therefore, is to understand the communication strategies that the Russian press used to motivate and justify the conflict in Ukraine and what types of information are disseminated by the selected newspapers. In this regard, through a Symmetric Non-Negative Matrix reduction factorization technique (symNMF), we extracted the main themes found in Russian newspaper articles to identify the topics used for propaganda.

## 2. Non-negative matrix factorization

Non-negative matrix factorization (NMF) is a dimension reduction method to uncover latent low-dimensional structures in high-dimensional data (Kim and Park, 2008). NMF is an unsupervised approach in that the low-rank factor matrices are constrained to have only nonnegative elements (Kuang et al. 2015). So, the basis vectors of the matrix are represented as a linear combination of vectors with positive coefficients.

Nonnegativity improves interpretations of the information extracted from a given data matrix, allowing a better understanding of the results obtained from the analysis process. This is in contrast to dimensionality reduction techniques that rely on the singular value decomposition (SVD) method, such as principal component analysis (PCA). One of the major problems with PCA is that the basis vectors have positive and negative components, and the data are represented as a linear combination of these vectors with positive and negative coefficients (Pauca et al. 2004). This is because the principal components are orthogonal, implying the presence of some negative values. Factors obtained from the NMF, on the other hand, are positive vectors and better approximate the data, but are not necessarily orthogonal (Casalino et al. 2016).

Given a $\mathbf{X}$ matrix of size $m \times n$, the decomposition of $\mathbf{X}$ into a matrix $\mathbf{W}$ of size $m \times k$ (called the base matrix) and a matrix $\mathbf{H}$ of size $k \times n$ (called the encoding matrix), such that their product approximates the matrix $\mathbf{X}$:

$$X \approx WH \tag{1.1}$$

Maria Gabriella Grassia, University of Naples Federico II, Italy, mariagabriella.grassia@unina.it, 0000-0002-7128-7323
Marina Marino, University of Naples Federico II, Italy, marina.marino@unina.it, 0000-0002-0742-5912
Rocco Mazza, University of Naples Federico II, Italy, rocco.mazza@unina.it, 0000-0002-4901-5225
Michelangelo Misuraca, University of Calabria, Italy, michelangelo.misuraca@unical.it, 0000-0002-8794-966X
Agostino Stavolo, University of Naples Federico II, Italy, agostino.stavolo@unina.it, 0000-0001-5890-2195

Referee List (DOI 10.36253/fup_referee_list)
FUP Best Practice in Scholarly Publishing (DOI 10.36253/fup_best_practice)

Maria Gabriella Grassia, Marina Marino, Rocco Mazza, Michelangelo Misuraca, Agostino Stavolo, *Topic modeling for analysing the Russian propaganda in the conflict with Ukraine*, © Author(s), CC BY 4.0, DOI 10.36253/979-12-215-0106-3.43, in Enrico di Bella, Luigi Fabbris, Corrado Lagazio (edited by), *ASA 2022 Data-Driven Decision Making. Book of short papers*, pp. 245-250, 2023, published by Firenze University Press and Genova University Press, ISBN 979-12-215-0106-3, DOI 10.36253/979-12-215-0106-3

where **W** e **H** are both non-negative matrices. The product **WH** is an approximate factorization of rank at most *k*. Generally, the *k* rank of the two matrices **W** and **H** is assumed to satisfy that $k \ll min\{m,n\}$ (Gaujoux and Seoighe, 2010). The value of the parameter *k* identifies the numbers of factors to be used to explain data (Casalino et al. 2016). Matrix multiplication can be implemented as computing the column vectors of **X** as linear combinations of the column vectors in **W** using coefficients given by columns of **H**. Each column of **X** can be computed as follows:

$$x_i = Wh_i \tag{1.2}$$

where $x_i$ is the column vector of the product matrix **X** and $h_i$ is the vector of the matrix **H.**

Suppose we have *n* data points represented as the columns in $X = [X_1, \ldots, X_n]$, and try to group them into k clusters. When **W** and **H** are subject to nonnegativity, it is possible to interpret the dimension reduction in (1.2) as clustering results: the columns of the first factor **W** provide the basis of latent k-dimensional space, and the columns of the second factor provide **H** the representation of $x_1, \ldots, x_n$ in the latent space. So, the cluster assignment for each data point is made by choosing the largest item in the corresponding column of **H** (Kuang et al. 2015).

The matrices **W** and **H** are found by solving an optimization problem defined with the Frobenius norm (a distance measure between two given matrices), the Kullback-Leibler (KL) divergence (a distance measure between two probability distributions), or other divergences.

The usual approach to NMF is to approximate **X** by calculating **W** and **H** to minimize the Frobenius norm of the **X - WH** difference, such that (Pauca et al. 2004):

$$\sum_{i=1}^{n}\sum_{j=1}^{m}(X_{ij} - (WH)_{ij})^2 = ||X - WH||_F^2 \tag{1.3}$$

The formulation in (1.3) has been applied to many clustering tasks in which the *n* data points are available in **X** and are used as an input. The relationship between the data points is represented as a graph, where each node corresponds to a data point and a similarity matrix $A_{n\times n}$ contains the similarity values between each pair of nodes (Moutier et al. 2021). The NMF is not a general clustering method that performs well in every circumstance, where the limitation can be attributed to its assumption on the cluster structure (Kuang et al. 2015). As we know, the goal is to approximate the original data matrix using a linear combination of basis vectors. When the underlying *k* clusters have nonlinear structure, NMF cannot find any *k* basis vectors that represent the clusters respectively.

So, it is used the SymNMF, the symmetric variant of the NMF, that handles symmetric matrices **A** as input. This method is based on a similarity measure between data points and factorizes a symmetric matrix containing pairwise similarity values into the product of a nonnegative matrix and its transpose (Jia et al. 2021). The factorization of **A** will generate a cluster assignment matrix that is nonnegative and captures the cluster structure inherent in the graph representation. Given an $n \times n$ symmetric matrix **A** and a reduced rank *k*, SNMF seeks to find the best factorization so that:

$$A \approx HH^T \tag{1.4}$$

where **H** can be viewed as the cluster indicator and $\mathbf{H^T}$ the transpose matrix.

Compared with NMF, SymNMF concerns only the factorized similarity matrix **A** and doesn't consider whether the structure of the data is linear or non-linear. It can be regarded as a graph clustering method, and it is more effective for nonlinearly separable data than NMF (Kuang et al.2015). It has demonstrated to be a powerful method for data clustering (Jia et al. 2021), for learning topics in text mining (Yan et al.2013).

Also, SymNMF is related to spectral clustering, SC, and both share the same loss function only with different constraints (Ng et al. 2001) and it can directly generate the clustering indicator without post-processing, while SC needs extra post-processing, like K-means, to finalize clustering.

## 3. Methodology

The proposed work shows preliminary results. Specifically, the analysis is carried out from March 2021, when the Russian military moved weapons and equipment into Crimea, to the end of March 2022, the day of the first negotiations in Istanbul. The selection of the newspapers is based on past study: the report "Pillars of Russia's Disinformation and Propaganda Ecosystem" produced by U.S. Department of State. According to the report, the journals cover various geographies, and they have their own target audiences. These newspapers are influenced by the Russian government and institutions, thus highlighting a Kremlin-driven information and regime interpretations given to the facts of war. The papers chosen are as follows:

- Strategic Culture Foundation: it is an online journal directed by Russia's Foreign Intelligence Service (SVR) and closely affiliated with the Russian Ministry of Foreign Affairs.
- Global research: a Canadian website that has become deeply enmeshed in Russia's broader propaganda ecosystem.
- News Front: it is a Crimea-based disinformation with the goal of providing an "alternative source of information" for Western audiences.
- South Front: it is an online information site registered in Russia that focuses on military and security issues.
- Katehon: a journal that plays the role of a provider of material aimed largely at a European audience, with content devoted to the "creation and defence of a secure, democratic and just international system."
- Geopolitics: a platform for Russian ultra-nationalists to spread disinformation and propaganda targeting Western and other audiences.

We extracted 3,396 newspaper articles, and two of them were withdrawn because they were not written in English; so, we had 3,394 articles. As we know, textual data are unstructured, so it's necessary to perform some phases of pre-processing for having structured data. There are different steps:

1. Normalized the text, so convert all the letters of the texts into a lower case;
2. Tokenized the documents, obtaining a set of distinct strings (tokens) separated by spaces or punctuation marks;
3. Removed special characters, punctuations, and numbers from the dataset. Also, hashtags, symbols and stopwords are eliminated;
4. Defined a grammatical tagging, which is the process of marking a word in a text as corresponding to a particular part of speech. In this case, we considered the nouns, verbs, and adjectives.

The pre-processing phase returned a database composed by 40.360 tokens, 5010 types and 3394 documents. In this way, the term-document matrix indicates the number of occurrences of each term in the document. The dimension of term-document matrix is 5010x3394.

## 4. Preliminary results

In the final stage of the pre-treatment process we applied the documents and words matrix vector space model. Each word is considered a vector where each element $a_i$ represents the weight of that element within the individual document. In NMF, the term-document matrix is too sparse to estimate reliable arguments, so more stable and dense data are used.

According to Yan et al. (2013), for reducing the sparsity of term-document matrix, we created a co-occurrence matrix **W** composed of the vectors $\mathbf{w_i}$., whose elements $a_i$ represent the number of times each word pair $<\mathbf{w_i},\mathbf{w_j}>$ co-occurs within the same document. For each pair of vectors, we calculated the cosine measure, a similarity index that measures the similarity between two vectors of an inner product space. In this way we created the similarity matrix **S**. On this matrix was applied the SymNMF for identifying the main topics. According to this, we found five topics:

| Topic 1 | Topic 2 | Topic 3 | Topic 4 | Topic 5 |
|---|---|---|---|---|
| vital | kill | basis | disagreement | manifestation |
| threat | launch | arrangement | development | modern |
| state | kiev | attempt | dispute | identity |
| ultimate | military | background | divide | individual |
| turn | mercenary | asset | disagreement | materialism |
| step | march | attention | economic | philosopher |
| urgent | nearby | auspice | economy | manifest |
| threaten | humanitarian | attack | drive | individualism |
| statement | operation | associate | dream | mankind |
| tolerate | munition | benefit | digital | liberty |

**Tab.1** – Topics extracted from SymNMF

Tab. 1 shows the terms associated with the topics extracted from the SymNMF. It is possible to define the topics that the Russian media used as motivations for the Ukrainian conflict.

The first topic excerpt refers to the threat posed by Ukraine. These newspapers present the conflict as a potential problem for relations with Europe and the West. There is an increasing urgency to seek common ground with the nation, declaring that it is responsible for all the casualties that are occurring. According to the Russian government, Ukraine is exaggerating the issue by not thinking of the collective good and making decisions that only sour the relationship with the Kremlin.

In “topic 2” reference is made to the war dimension. The terms given allow identification of the main arguments the Russians used to justify the invasion. The war is presented as a “humanitarian operation” that Putin undertook to liberate Kiev and Ukraine. Russian propaganda aims to present the Ukrainian people, not as victims but as perpetrators of their crimes and murders: the term "mercenary" refers to a narrative that Ukrainian soldiers murdered Danish mercenaries. This serves to dispel the idea of Ukrainians as a subjugated people. All words referring to the dimension of war emphasize the belligerent spirit of the population, which wants to get rid of the Russian “nearby” enemy even with the use of atomic bombs and munitions received by the U.S.

Related to this, there is “Topic 3” that highlights diplomatic-international relations. Especially, the topic appears to be declined in a general way but such that the dimension described can be interpreted. According to Russian media, President Putin has repeatedly proposed talks and negotiations and set out Russia's conditions, the first is that no NATO base be installed in Ukraine. The topic of international agreements turns out to be central as on the one hand the media talk about the Russian government's willingness to mediate with Ukraine, and on the other hand, they emphasize how the same nation wants to join NATO and improve the relationship with President Biden.

“Topic 4” identifies the economic motivations of the invasion. There is a common view that the Ukrainian government initiated the conflict to increase its economic and geopolitical power to expand into neighbouring countries and counter the Russian nation. In particular, the newspapers report a series of events related to the Ukrainian economy: the emigration of citizens to Poland to

improve wage conditions, the loss of public money, and the resulting debts. The latter aspect is central as it predicts Ukrainian dependence on America. By this, it is narrated that European powers are also obliged to give money to Ukraine to pay for previous situations: for example, the Germans pay a debt to Ukraine for the occupation of Crimea.

The last topic refers to a philosophical-identarian dimension, in which Western and liberal values are criticized and exaggerated. The thoughts of numerous philosophers who present liberalism in a negative light are reported, stating that it "should be opposed to fascism" but it imposes itself on Western civilizations, taking on the same characteristics. For this reason, the U.S. is presented as a nation that does not want to assert other world powers by imposing its own economic and social vision. In this regard, the Russian media propose a vision of its nation as one that engages in the pluralism of ideas and goes to represent an alternative of freedom to the Western world. Numerous articles are referencing how the U.S. wants to change tradition and classical roots (e.g., Dante's works have been called politically incorrect and have undergone liberal cleansing) by going on to criticize the philosophers and thinkers of the time. On the other hand, Russia is presented as the guardian of "true and authentic" European values and not of "globalization" and "liberalism".

## 5. Conclusions

As discussed earlier, the work is in the preliminary stage and the aim is to identify the themes that the Russian government used to narrate the conflict. For future developments, we are expanding the list of Russian information sources in order to conduct a more comprehensive analysis.

Russian Federation invests its propaganda channels and its intelligence services to conduct activities to support their information system, and it leverages outlets on news sites or research institutions to spread these narratives.

So, the Kremlin use these tactics as part of its approach to using information as a weapon. In this regard, the Russian government has issued a series of measures, ordering all media outlets to report on the invasion of Ukraine only through official state sources, blocking numerous sites for spreading unfounded news and threats of high treason. Russia's willingness to employ this approach provides it with three advantages. First, it allows for the introduction of numerous variations of the narratives, to fine tune their information narratives to suit different target. Second, it provides plausible deniability for Kremlin officials when they peddle different information, allowing them to deflect criticism while still introducing damaging information. Lastly, it creates a media multiplier effect that boost their reach and resonance.

## References


Abd Kadir, S., Abu Hasan, A. S. (2014). A content analysis of propaganda in Harakah newspaper. *Journal of Media and Information Warfare* (JMIW)

Casalino, G., Del Buono, N., Mencar, C. (2016). Nonnegative matrix factorizations for intelligent data analysis. *Non-negative Matrix Factorization Techniques* (pp. 49-74). Springer, Berlin, Heidelberg.

Jia, Y., Liu, H., Hou, J., Kwong, S., Zhang, Q. (2021). Self-supervised symmetric nonnegative matrix factorization. *IEEE Transactions on Circuits and Systems for Video Technology.*

Kim, J., Park H., (2008) Sparse nonnegative matrix factorization for clustering, Georgia Inst. of Technology, Tech. Rep.

Kuang, D., Choo, J., Park, H. (2015). Nonnegative matrix factorization for interactive topic modeling and document clustering, in *Partitional Clustering Algorithms* (pp. 215-243). Springer, Cham.

Moutier, F., Vandaele, A., Gillis, N. (2021). Off-diagonal symmetric nonnegative matrix factorization. *Numerical Algorithms*, 88(2), 939-963.

Ng, A., Jordan, M., Weiss, Y. (2001). On spectral clustering: Analysis and an algorithm. *Advances in neural information processing systems*, *14*.

Pauca, V. P., Shahnaz, F., Berry, M. W., Plemmons, R. J. (2004). Text mining using non-negative matrix factorizations, in *Proceedings of the 2004 SIAM International Conference on Data Mining* (pp. 452-456). Society for Industrial and Applied Mathematics.

Pratkanis, A., Aronson, E. (1991). What is news? In *Age of Propaganda: The Everyday Use and Abuse of Persuasion*, (pp. 32 – 39). New York: W. H. Freeman.

U.S. Department of State, GEC Special Report (2020). Pillars of Russia's Disinformation and Propaganda Ecosystem, https://www.state.gov/wp-content/uploads/2020/08/Pillars-of-Russia%E2%80%99s-Disinformation-and-Propaganda-Ecosystem_08-04-20.pdf

Yan, X., Guo, J., Liu, S., Cheng, X., Wang, Y. (2013). Learning topics in short texts by non-negative matrix factorization on term correlation matrix, in *Proceedings of the 2013 SIAM International Conference on Data Mining* (pp. 749-757), Society for Industrial and Applied Mathematics.

This volume collects the contributions presented at the conference "Data-driven Decision Making" organized by the Italian Association for Applied Statistics, held in Genoa from 12 to 14 September 2022. The papers cover a broad range of topics, with a common thread: the use of statistical methods to support decision-making both in public administrations and in private companies.

Enrico di Bella is Associate Professor of Social Statistics at the Department of Political and International Sciences of the University of Genoa.

Luigi Fabbris is President of the Italian Association for Applied Statistics and former Professor of Social Statistics at the University of Padua.

Corrado Lagazio is Professor of Statistics at the Department of Economics of the University of Genoa.

ISSN 2704-601X (print)
ISSN 2704-5846 (online)
ISBN 979-12-215-0106-3 (PDF)
ISBN 979-12-215-0107-0 (XML)
DOI 10.36253/979-12-215-0106-3
www.fupress.com