Text-mining forma mentis networks reconstruct public perception of the STEM gender gap in social media

Massimo Stella Corresp. [1]

[1] Complex Science Consulting, Lecce, Italy

Corresponding Author: Massimo Stella
Email address: massimo.stella@inbox.com

Mindset reconstruction maps how individuals structure and perceive knowledge, a map unfolded here by investigating language and its cognitive reflection in the human mind, i.e. the mental lexicon. *Textual forma mentis networks* (TFMN) are glass boxes introduced for extracting, representing and understanding mindsets' structure, in Latin *forma mentis*, from textual data. Combining network science, psycholinguistics and Big Data, TFMNs successfully identified relevant concepts, without supervision, in benchmark texts. Once validated, TFMNs were applied to the case study of the gender gap in science, which was strongly linked to distorted mindsets by recent Big Data Analytics studies. Focusing over social media perception and online discourse, this work analysed 10,000 relevant tweets. "*Gender*" and "*gap*" elicited a mostly positive perception, with a trustful/joyous emotional profile and semantic associates that: celebrated successful female scientists, related gender gap to wage differences, and hoped for a future resolution. The perception of "*woman*" highlighted discussion about sexual harassment and stereotype threat (a form of implicit cognitive bias) relative to women in science "sacrificing personal skills for success". The reconstructed perception of "*man*" highlighted social users' awareness of the myth of male superiority in science. No anger was detected around "*person*", suggesting that gap-focused discourse got less tense around genderless terms. No stereotypical perception of "*scientist*" was identified online, differently from real-world surveys. This analysis thus identified Twitter discourse as promoting a mostly stereotype-free, positive/trustful perception of gender disparity, of relevance for closing the gap. TFMNs enable new ways for monitoring online mindsets emerging from user-generated content, offering detailed data-informed ground for policy making.

Text-mining forma mentis networks reconstruct public perception of the STEM gender gap in social media

Massimo Stella[1]

[1]**Complex Science Consulting, Via Amilcare Foscarini 2, 73100 Lecce, Italy**

Corresponding author:
Massimo Stella[1]

Email address: massimo.stella@inbox.com

ABSTRACT

Mindset reconstruction maps how individuals structure and perceive knowledge, a map unfolded here by investigating language and its cognitive reflection in the human mind, i.e. the mental lexicon. *Textual forma mentis networks* (TFMN) are glass boxes introduced for extracting, representing and understanding mindsets' structure, in Latin *forma mentis*, from textual data. Combining network science, psycholinguistics and Big Data, TFMNs successfully identified relevant concepts, without supervision, in benchmark texts. Once validated, TFMNs were applied to the case study of the gender gap in science, which was strongly linked to distorted mindsets by recent Big Data Analytics studies. Focusing over social media perception and online discourse, this work analysed 10,000 relevant tweets. ”Gender” and ”gap” elicited a mostly positive perception, with a trustful/joyous emotional profile and semantic associates that: celebrated successful female scientists, related gender gap to wage differences, and hoped for a future resolution. The perception of ”woman” highlighted discussion about sexual harassment and stereotype threat (a form of implicit cognitive bias) relative to women in science ”sacrificing personal skills for success”. The reconstructed perception of ”man” highlighted social users' awareness of the myth of male superiority in science. No anger was detected around ”person”, suggesting that gap-focused discourse got less tense around genderless terms. No stereotypical perception of ”scientist” was identified online, differently from real-world surveys. This analysis thus identified Twitter discourse as promoting a mostly stereotype-free, positive/trustful perception of gender disparity, of relevance for closing the gap. TFMNs enable new ways for monitoring online mindsets emerging from user-generated content, offering detailed data-informed ground for policy making.

INTRODUCTION

Perception is in the mind of the beholder. Every experience contributes to building a memory or mental reconstruction of the outer world which, in turn, deeply impacts future behaviour (Malt et al., 2010; Aitchison, 2012; Beasley and Mason, 2015; Seli et al., 2019; Li et al., 2020). Negative perceptions can inhibit efficient learning (Chavatzia, 2017) or drastically alter information processing (Shapiro and Williams, 2012), while positive perceptions can contribute to the acceptance and establishment of social norms (Malt et al., 2010; Welles and González-Bailón, 2020; Waterloo et al., 2018). In this way, reconstructing and understanding the cognitive perceptions of social groups is key to achieving insights about human behaviour and social patterns.

Discovering the perception of a given audience is a challenge that can be broken mainly in two parts: (i) acquiring cognitive data, reflecting how the public perceives a certain phenomenon, and (ii) processing such data in an efficient way, suitable for the extraction of new knowledge (de Arruda et al., 2019). Data can be gathered by mining social media (Stella et al., 2018b; Bovet et al., 2018), which represent an invaluable source of information about the users' experiences and perceptions of specific topics (Welles and González-Bailón, 2020). Recently, social media, Twitter in particular (Jansen et al., 2009), have been increasingly analysed by the scientific community in order to detect complex phenomena such as the emotional dynamics of voting events (Stella et al., 2018b; Bovet et al., 2018), the promotion of self-branding and journalistic content also through social bots (Varol and Uluturk, 2020), the spread of disinformation (Pierri et al., 2020) and the fostering of online hate dissemination (Waqas et al., 2019).

Towards a cognitive approach to information processing

Although Twitter messages provide several types of information, such as visual cues (e.g., pictures and videos) or multi-language cues (e.g., emojis and hashtags), their nature is mainly textual (Welles and González-Bailón, 2020; Jansen et al., 2009). In this way, the problem of quantifying how an audience perceives a given topic can be related to stance detection from their tweets (Mohammad, 2016). Approaching stance detection by human coding becomes quickly intractable when faced with the large volumes of messages exchanged daily on the Twittersphere. This limitation leads to the above need for developing and adopting efficient techniques for knowledge extraction from massive amounts of text as exchanged on social media. Given the extraordinary possibility for humans to communicate their mental constructs through language (Aitchison, 2012), a key way of detecting perceptions is through communication. Both in computer science and linguistics, the problem of detecting positive or negative perceptions from language is known as *stance detection* (Mohammad, 2016). Rather than focusing on language in itself, this work shifts the attention to the cognitive reflection of language in the human mind, in the so-called *mental lexicon* (Malt et al., 2010; Aitchison, 2012; Dóczi, 2019; De Deyne et al., 2019). Such lexicon includes semantic memory, a well-studied repository of conceptual meanings and word features (Kenett et al., 2017; Dóczi, 2019), an also other memory supports storing syntactic/phonological/orthographic and even affective knowledge, together with other aspects of language (Dóczi, 2019). Whereas syntax, phonology and orthography express word-level conceptual knowledge (Aitchison, 2012), affective knowledge links concepts to the emotions they elicit (Malt et al., 2010) and therefore it represents an important component of any stance/perception.

Harnessing the complex cognitive structure of the mental lexicon means accessing how conceptual knowledge is structured and emotionally perceived by individuals as the outcome of their previous experiences and current attitudes (Aitchison, 2012; Stella et al., 2019; Stella, 2020). In other words, accessing the conceptual and affective representation of knowledge in the mental lexicon means reading minds(ets), with strong repercussions for information processing. Discovering how an audience perceives a given topic, i.e. performing mindset reconstruction, can provide crucial knowledge for understanding and intervening upon specific trends (Welles and González-Bailón, 2020; Amancio et al., 2012; Stella et al., 2019). An example is represented by the finding that mindsets vehiculating positive emotions reach larger audiences on social media whereas negative emotional content can spread at faster rates (Ferrara and Yang, 2015). Another example for the relevance of mindset reconstruction is uncovering and acting upon traces of science anxiety in student populations in order to improve their learning experiences (Stella, 2020; Stella and Zaytseva, 2020) or detecting sexual harassment through large-scale web surveys (Karami et al., 2020).

A key area where mindset reconstruction is particularly promising is understanding the sources and dynamics of the *gender gap in science* (Hogue and Lord, 2007; Moss-Racusin et al., 2012).

As outlined by recent Big Data Analytics studies considering decades of scientific careers (Huang et al., 2020; Odic and Wojcik, 2019; Chavatzia, 2017), gender disparities in science cannot be explained by intrinsic differences in attitudes to science between genders but rather have to be traced in the establishment of implicit gender biases promoted by news media and social media representations of science (Shapiro and Williams, 2012; Moss-Racusin et al., 2012; Madsen and Andrade, 2018; Steinke, 2017). In this way, the reconstruction of mindsets vehiculated by information systems becomes a key point for understanding, acting upon and closing the gender gap in science. Given the recency of the above mentioned Big Data Analytics studies and the methodological issues in reconstructing mindsets with black-box machine learning techniques (often neglecting contextual information (Nasar et al., 2019)), reconstructing online perceptions of the gender gap in social media is a challenge not fully explored yet.

Research aim

This work introduces textual forma mentis networks (TFMNs) as quantitative tools for reconstructing the mindset of online users engaging in social discourse with the research aim of investigating in detail the mindset emerging from user-generated content about the gender gap in science. In other words, this work uses textual forma mentis networks with the aim of reconstructing how online users discussed and perceived messages revolving around the STEM gender gap. In order to mirror the general perception of this gap, a window without special events about women in science was chosen. TFMNs quantified the general stance from 10,000 tweets publicly available on Twitter, produced between October 8 2019 and October 22 2019, containing the words or hashtags “science” or “stem” and “women” or “gendergap”. These tweets encapsulated information about how the online authors perceived women in science, with relevant impact for education research and the computational social sciences.

Before providing additional details about the adopted methodology, it is necessary to compare TFMNs against past approaches, also relating forma mentis networks within the relevant literature about the gender gap in science.

Literature review on relevant past approaches

It has to be underlined that only recently automatic text-mining studies started investigating gender gap in online discourses (Teso et al., 2018; Chavatzia, 2017). However, these approaches mainly focused on detecting differences in language use *among* different genders. Differently from other information processing investigations aiming at identifying emotions on social media in relation to phenomena like hateful speeches (Waqas et al., 2019) or disinformation spreading (Pierri et al., 2020), gender-focused investigations of social media did not explore large-scale mappings of the online perception of the gender gap in science as embedded in the messages exchanged between social users of any gender. Within an information management setting, the work closest to an investigation of the overall mindset about gender biases was the study by Karami and colleagues (Karami et al., 2020). The authors investigated online self-reports of sexual harassment experiences and through a topic analysis they highlighted evidence for sexual harassment in academia mainly targeting women and involving coercion, gender discrimination and retaliation. Building upon the knowledge extraction approach of Karami et al. (Karami et al., 2020), this study shifts its attention from explicit sexual harassment to the larger topic of gender biases in science, which includes harassment itself but also implicit biases (Shapiro and Williams, 2012), gender pay gaps (Courey and Heywood, 2018) and stereotypical perceptions about leadership (Pennington et al., 2016; Ely et al., 2011). Furthermore, rather than focusing on self-reports, this study aims at tackling a different information system, namely Twitter, where social users can engage in social discourse and reach large audiences (Welles and González-Bailón, 2020). This reconstruction of social media attitudes about the gender gap in science is the main gap that the current work aims to fulfil by using a network-powered approach, which has several similarities and novelties in comparison to past network frameworks.

Within the literature of network-powered analyses of large volumes of text (de Arruda et al., 2019), previous works successfully used word co-occurrences in text (e.g. “like” and “stem” occurring one after the other in text, cf. (Cancho and Solé, 2001)) in order to characterise language content through average statistical markers (Amancio et al., 2012; Amancio, 2015) or time-evolving dynamics (Akimushkin et al., 2017). These networks were considerably powerful at a global level and very successful in tasks like author identification (Amancio, 2015; Akimushkin et al., 2017; de Arruda et al., 2019). However, the validity of co-occurrence networks as representations of the mental lexicon at the microscopic level of individual conceptual associations has been recently reconsidered (Ninio, 2014; Rizvi, 2018). In fact, co-occurrences can lead to spurious conceptual links, whose influence vanishes in global statistical approaches and that do not represent syntactic similarities in text (Ninio, 2014). In order to overcome this limitation and achieve a more faithful microscopic representation of the mental lexicon, TFMNs directly harness the full ensemble of syntactic associations of a sentence (e.g. “like” being a verb referring to the object “stem”, cf. (i Cancho et al., 2004)) and enrich them by considering also semantic overlap between words (e.g. “appreciate” and “like” being synonyms across different sentences, cfr. (Miller, 1998; Amancio et al., 2012)).

Textual forma mentis networks automatically extract these conceptual associations from text without requiring human supervision and are therefore suitable for processing large volumes of text. The resulting network structure is informative of the cognitive layout of conceptual associations emerging from a given textual corpus and hence represents how text authors organised, structured and associated microscopically their knowledge around topics and concepts. This makes TFMNs “glass boxes” (Nasar et al., 2019), where the knowledge structure of a certain stance can be accessed and directly read, differently from previous “black box” machine learning approaches which accurately reproduced the positive or negative nature of a stance without providing information on its semantic content (Mohammad, 2016; Teso et al., 2018; Rudkowsky et al., 2018; Nasar et al., 2019). In addition to conceptual associations, TFMNs are endowed also with sentiment labels, indicating the sentiment (Warriner et al., 2013) and the basic emotions (Ekman and Davidson, 1994; Mohammad and Turney, 2013) elicited by a given concept in a population of individuals involved in behavioural studies. Sentiment scores of positive/negative affect are also called word valence in psycholinguistics (Warriner et al., 2013; Recchia and Louwerse, 2015) and represent how positively or negatively a given concept was perceived in a behavioural study. Word valence and emotional profiles add more emotional contextual information about the stance reconstructed by conceptual associations (Stella, 2020). Although large-scale datasets about sentiment and emotions have only been recently made available to the scientific community by cognitive studies (Warriner et al., 2013), they have quickly become predominant in predicting a wide variety of human behaviour (Li et al., 2020) and information processing patterns such as consensus formation in social networks (Konstantinidis et al., 2017) or information sharing on microblogs (Ferrara and Yang, 2015). The combination of network patterns and sentiment data is important, as considering only the frequency of sentiment labelled content in short media has been reported to lack interpretative contextual power for estimating how people really feel about a given topic (Beasley and Mason, 2015).

Different types of forma mentis networks

As outlined above, text-based forma mentis networks represent multiplex lexical networks (Stella et al., 2018a) of concepts interconnected through syntactic and semantic associations and enriched with sentiment and emotional labels. Previous works showed how lexical networks of concepts including multiple layers of associations (i.e. "multiplex") were better than single-layer complex networks at predicting a variety of cognitive processes involved in information acquisition (Stella et al., 2018a; Stella, 2019) and search (Castro and Stella, 2019; Siew et al., 2019). However, these approaches did not explore the influence that emotions and sentiment can have over information processing (Warriner et al., 2013; Dóczi, 2019; Rudkowsky et al., 2018; Li et al., 2020). The introduction of behavioural forma mentis networks (BFMNs) (Stella et al., 2019; Stella and Zaytseva, 2020) tackled such gap by combining conceptual associations with sentiment scores and providing access to knowledge structure and positive/negative perceptions of concepts in a given mindset. However, BFMNs' mindset reconstruction comes at the cost of involving individuals in a cognitive experiment (cf. (Stella et al., 2019)). This limitation translates into the possibility of reconstructing only the mindsets of those who participated in the experiment, preventing access or monitoring of remote systems like social media. Textual forma mentis networks do not require behavioural experiments but can rather be built starting from any written text. As a consequence, TFMNs are suitable for the investigation of speech over online platforms, whereas BFMNs are unstuitable for such task. However, as outlined above, both the representations exploit the theoretical framework of mental lexicon representations (Dóczi, 2019) and language processing in cognitive network science (Siew et al., 2019). Thanks to such foundations, TFMNs are applied here for reconstructing the mindset of online discourse about the gender gap in science, as motivated in the next subsection.

Recent literature identifies a link between the gender gap and distorted mindsets

Overwhelming evidence indicates that the gender gap in science is a complex phenomenon deeply affecting society, economics and science advancement (Ely et al., 2011; Madsen and Andrade, 2018; Hogue and Lord, 2007; Pietri et al., 2018). The gender gap in the scientific, technological, engineering and mathematical (STEM) disciplines is a disparity of how different genders enter in and progress through a career in science (Shapiro and Williams, 2012). Between 2014 and 2016, UNESCO estimated only around 30% of all female students in higher education enrolled in STEM-related fields of study at University level (Chavatzia, 2017). This gender gap in STEM education and participation is almost absent at the level of primary education but then becomes particularly apparent during upper secondary education, coincidentally with subject selection, and it gets worse at higher levels of education (Ely et al., 2011; Chavatzia, 2017; Hogue and Lord, 2007). Globally, only 28% of all the world's researchers are women. This disparity is deeply embedded in educational systems, in particular in terms of attitudes and perceptions towards science (Shapiro and Williams, 2012; Hogue and Lord, 2007; Huang et al., 2020). The PISA 2015 report *Excellence in Education* reported that across 35 OECD countries, only 22% of 15-years old girls intend to pursue a career in STEM, less than half the proportion (48%) of STEM driven 15-years old boys (cf. (Chavatzia, 2017)). In the last few years, a great attention has been devoted to explaining such gap by using either external, e.g. system embedded discrimination (Chavatzia, 2017; Ely et al., 2011), or internal factors, e.g. implicit stereotypes (Shapiro and Williams, 2012; Pietri et al., 2018). Studies trying to explain gender gap through gender-based learning achievements reported a complex landscape, with boys and girls being more or less proficient than each other according to the task being measured (cf. (Chavatzia, 2017)). This complex mosaic of multiple findings makes it complicated to state that any gender is more or less advanced in any given STEM subject: the gender gap has to be rooted in more subtle forms of discrimination, stereotypes and perceptions (Moss-Racusin et al., 2012; Lane et al., 2012). Despite this fragmentation, the data unanimously indicate that women are paid less and leave STEM careers way more than their male colleagues either at University level or at subsequent stages of professional growth in science (Courey and Heywood, 2018).

Even professional impact in STEM is influenced by a strong gender gap. The recent longitudinal study by Huang and colleagues (Huang et al., 2020) considered professional impact of women in STEM through bibliographic Big Data spanning over 60 years of publications. The authors reported evidence for gender differences in the cumulative productivity of research output but not in the annual rate of publication or career-wise impact. Through a quantitative, longitudinal analysis, Huang and colleagues related such gender gap to dropout rates and differences in length of publishing careers between men and women. The authors could not explain such disproportion only in terms of intrinsic gender-based proficiency in STEM, thus providing additional large-scale evidence for the presence of strong contextual factors affecting everyone's experience of the STEM gender gap. Analogous differences through Big Data approaches were recently found also by Odic and Wojcik in psychology, a field where 3 in 4 students are women whereas 3 in 4 academic professionals are male (Odic and Wojcik, 2019).

The need to expose distorted mindsets about the gender gap in information systems

The predominance of strong academic biases in contrast with educational patterns suggests the presence of hidden roots to the STEM gender gap, as indicated by several independent studies on the topic (Pennington et al., 2016; Shapiro and Williams, 2012; Pietri et al., 2018). Hence, understanding the overall experience and subsequent perception of this gender gap in a large audience can provide key elements for better detecting the presence of potentially subtle yet strong *gender-based stereotypes*, as perceived and communicated by the main actors of STEM. Given that most of these stereotypes can act at a subconscious level and take place without the explicit awareness of those perpetrating them (Pennington et al., 2016; Madsen and Andrade, 2018; Lane et al., 2012), detecting the presence of such distorted perceptions remains a difficult challenge.

Monitoring and detecting the diffusion of stereotypical endorsements in social media represents an important way of better understanding and countering gender discrimination in science (Karami et al., 2020), especially in the current society dominated by virtual social media (Welles and González-Bailón, 2020; Jansen et al., 2009).

Manuscript organisation

This manuscript is organised in several subsections. The Methods section contains quantitative details about the implemented methodology and analysed data. The Results section is split in two subsections. The first part is a quantitative benchmark, reporting on the effectiveness for TFMNs in finding semantically relevant concepts and topic features in short texts annotated by authors. The second part explores the reconstructed mindset of online users towards the gender gap by considering the stance towards key domains and aspects of the gender gap, e.g. concepts like "woman", "man", "person", "scientist", and related aspects, e.g. "gender", "gap" and "stem". The detected conceptual associations and emotional patterns are investigated and related with previous studies within the Discussion section.

METHODS

This section introduces the following elements: (i) the dataset about the online perception of the gender gap in STEM as retrieved from Twitter, (ii) the methodology behind the construction of text-based forma mentis networks, and (iii) the cognitive data used for detecting conceptual overlap in meaning between words and their valence. This section also reports on the benchmark data used for testing TFMNs, namely the text "Complexity Explained" (https://complexityexplained.github.io/ - Last Accessed: 16/03/2020) and the behavioural forma mentis network of international STEM researchers analysed in (Stella et al., 2019) and (Stella, 2020).

Twitter dataset

The main dataset used in this investigation was a collection of 10384 tweets publicly available on Twitter and produced between October 8 2019 and October 22 2019. Tweets were gathered through the *ServiceConnect[]* function for Twitter crawling implemented in Mathematica 11. Crawling was performed in accordance with Twitter's policies via the account of Complex Science Consulting, which received authorisation by Twitter for research-focused text mining. Tweets were included in the dataset if they contained either the hashtags (or the words) "#science" or "#stem" and "#women" or "#gendergap". Twitter's policy prevents the redistribution of these tweets outside of the Twitter platform. Nonetheless, for the sake of scientific reproducibility, the IDs of these tweets were attached to this manuscript as Supplementary Information.

Each tweet contained a short text. Pictures and emoticons were discarded. Hashtag characters were removed. Tweets with less than three words were not included in the analysis but constituted less than 1% of the whole dataset.

Network construction

Text-based forma mentis networks were built in three different stages:

1. Extraction of syntactic relationships between concepts/words from a sentence;
2. Addition of semantic relationships (synonyms) between concepts as indicated by an external dataset;
3. Addition of valence labels ("positive", "negative", "neutral") to each single concept as indicated by another external dataset.

The above three steps were repeated for all the sentences in a given tweet. At the end, all connections between the extracted and labelled concepts formed a multiplex lexical network (Stella et al., 2018a) where nodes represented words/concepts and were connected across multiple linguistic layers, namely: (i) a semantic layer indicating meaning overlap between words (e.g., "famous" and "notable" sharing the same meaning), and (ii) a syntactic layer indicating dependencies in meanings as encapsulated within a sentence. In general, syntactic structure defines how entities (e.g. nouns) are specified in a given sentence through verbs, determiners, prepositions, adjectives and adverbs (i Cancho et al., 2004). Although more or less convoluted syntactic structures can be encoded in sentences, the simplest form of the syntactical dependencies is a subject being specified as an object. For instance, in the sentence "love is weakness", "love" is specified as "weakness" through the verb "is". The verb "is" does not encapsulate any intrinsic meaning but it can be replaced by a link between "love" and "weakness". Syntactic dependencies can be more general, for instance in the sentence "the cat sat on the chair" the nominal subject "cat" is linked to the prepositional object "chair" through the verb "to sit", which retains some meaning. The determiner "the" and the preposition "on" do not retain meaning by themselves but connect the other parts of speech and are thus essential for extracting syntactic dependencies. The specification of such dependencies was implemented through the *TextStructure[]* command in Mathematica 11, which produces all syntactic relationships between the parts of speech of a given sentence. From the resulting network of directed dependencies, prepositions and auxiliary verbs were removed as nodes and replaced by syntactic links. In order to avoid the inflection of lemmas with the same meaning, i.e. having "weak" and "weakness" in the same network, word stemming was performed at the network level.

After the extraction of syntactic dependencies, the resulting syntactic network identified a set of connected, stemmed concepts. Syntactic structure can be informative about conceptual links in the mental lexicon and be predictive of language learning and processing (i Cancho et al., 2004; Stella et al., 2018a). However, syntactic dependencies neglect the possibility of using and exchanging synonyms in the same syntactic structure providing the same meaning (e.g. saying "he has a quiet character" conveys the same meaning of saying "he has a quiet nature"). In order to account also for meaning overlap, the syntactic structure was enriched with synonym relationships from WordNet 3.0 (Miller, 1998), as implemented in the curated repository *WordData[]* available in Mathematica 11. Meaning overlap between concepts is also representative of semantic memory patterns in the mental lexicon, with previous works showing how synonym networks can predict language learning and facilitate conceptual navigation (Stella et al., 2018a; Siew et al., 2019).

After the construction of the network layers of syntactic relationships and of synonyms, every concept was connected either by syntactic or semantic links. From the affective mega-study of Warriner and colleagues (Warriner et al., 2013), each concept was endowed with a valence label, e.g. "positive", "neutral" or "negative". These labels represented the average valence attributed to each concept by a population of individuals involved in a large-scale behavioural study rating more than 13,900 English words (Warriner et al., 2013). Words were classified as positive, neutral or negative, respectively, according

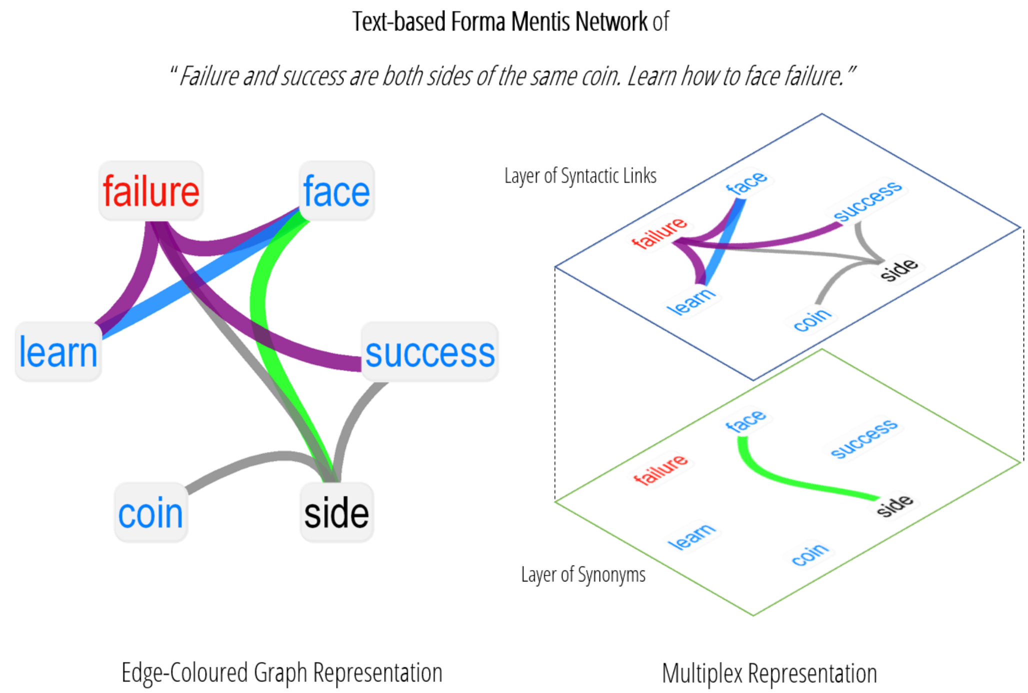

Figure 1. Example of a text-based forma mentis network. A TFMN can be represented either as an edge-coloured graph (left) or as a multiplex network (right). Positive (negative) words are highlighted in cyan (red). Syntactic links between positive (negative) words are highlighted in cyan (red) too. Syntactic links between positive and negative concepts are in purple. All semantic links of meaning overlap are highlighted in green.

to their location in the upper quartile (higher valence), interquartile range (neutral valence), or lower quartile (lower valence) in the distribution of valence scores for over 9000 different word stems. These valence scores were obtained from averaging over words with the same stemmed root.

Figure 1 provides an example of a text-based forma mentis network. A TFMN can be represented either as an edge-coloured graph or as a multiplex network (Stella et al., 2018a). These two representations are equivalent and their only purpose is to distinguish between syntactic and semantic links.

Additional cognitive data

In order to test the power of text-based forma mentis network in identifying relevant concepts in text, additional cognitive data was used as benchmark. Semantic similarity and relevance of scientific concepts, as extrapolated from short scientific texts, was investigated and compared against free association data from the Complex Forma Mentis project (Stella et al., 2019). The benchmark text adopted here was the booklet "Complexity Explained", co-authored by several researchers in complexity science and composed of 7 short paragraphs describing one specific concept each (cf. https://complexityexplained.github.io/, Last Accessed: 16/03/2020). The analysed paragraphs were about: "interactions", "emergence", "dynamics", "self-organisation", "adaptation", "interdisciplinarity" and "methods". The results of such benchmark are reported at the beginning of the Results section.

Emotional profiling

Compared to previous approaches using behavioural forma mentis networks (Stella, 2020; Stella and Zaytseva, 2020; Stella et al., 2019), the current analysis introduces also an emotional profiling of stances as encapsulated in the reconstructed mindsets. Emotional profiling is defined in terms of considering how many associates of a concept, in the TFMN, elicit a given emotion. Concepts linked to a negation (e.g. "not") were transformed into their antonyms, so that both original concepts and their negated meanings were considered for reconstructing an emotional profile. Considered emotions included:

- **Anger**, a negative emotion representing reactions of irritation and rage towards an external threat;

- **Disgust**, a negative emotion indicating aversion and closure;
- **Fear**, a negative emotion indicating a need to actively avoid and prevent potential threats;
- **Trust**, a positive emotion indicating openness towards the outer world;
- **Joy**, a positive emotion of excitement and satisfaction;
- **Sadness**, a neutral emotion, neither positive nor negative, indicating states of sorrow, thoughtfulness and inhibition;
- **Surprise**, a neutral emotion relative to being upset or startled by an unexpected event;
- **Anticipation**, a neutral emotion indicating one's projection into future events, including desire or anxiety.

The above constitute basic building blocks of a wide spectrum of emotional states (Ekman and Davidson, 1994). For a more detailed description of emotions, please see (Ekman and Davidson, 1994) and (Mohammad and Turney, 2013). The mega-study by Mohammad and colleagues is also known as the Word-Emotion Association Lexicon, indicating how human participants associated individual concepts to emotional states. In other words, the resulting dataset indicated which emotions were elicited or *evoked* in human subjects reading individual concepts. The study used crowd-sourcing on Mechanical Turk for achieving a large-scale mapping of English words (including 14,000 unique lemmas) and was used also in other successful investigations about affect patterns (cf. (Mohammad and Turney, 2013)).

Cognitive measures of semantic similarity on language networks

This work built networks of conceptual associations between concepts. Several independent studies have reported that on such networks, metrics like network distance are powerful proxies for quantifying semantic relatedness (Kenett et al., 2017; Stella and Zaytseva, 2020). Network distance d_{ij} is defined as the minimum number of links (here conceptual associations) connecting any two words i and j on a network structure (Siew et al., 2019). On multiplex lexical networks, links of any colour/layer can be used (Stella et al., 2018a; Stella, 2019). Measures based on network distance such as closeness centrality have been found to identify also concepts of relevance from a cognitive perspective when detecting key words for word learning (Stella et al., 2018a; Stella, 2019), language processing (Kenett et al., 2017; Siew et al., 2019; Castro and Stella, 2019) and knowledge exploration (Stella and Zaytseva, 2020; Akimushkin et al., 2017; Amancio, 2015). Closeness centrality (Siew et al., 2019) $c(i)$ is attributed to node i by checking how distant it is to its connected neighbours, in formulas:

$$c(i) = \frac{N}{\sum_{j=1}^{N} d_{ij}}. \tag{1}$$

Notice that the above formula applies to fully connected components of a network but does not enable direct comparison of components including different numbers of nodes (e.g. a component including only 5 nodes versus a component with 1000 nodes).

Forma mentis networks included also structural features of the mental lexicon of individuals combined with affective patterns (Warriner et al., 2013). This unique combination allowed for Stella and colleagues (Stella et al., 2019) to introduce the network metric of *valence auras*, i.e. the mode of valence labels in a given neighbourhood. In (Stella et al., 2019) a word was defined as having a positive (negative) valence aura if mostly linked to positive (negative) words. This measure was used also in the current analysis, although it has to be underlined that in here these affect measures represented how a large-scale population, independent from the one producing the texts, perceived individual concepts. Since valence auras still depended on the connectivity of conceptual associations as assembled by text authors, this metric quantified how groups of these authors structured and perceived their knowledge. In the current analysis, valence auras rather than single-word valence labels were key to reconstructing positive/negative perceptions of individual concepts by checking their conceptual associates in the analysed online discourse.

Table 1. Top-ranked concepts in the TFMNs obtained from the 7 paragraphs of the booklet *Complexity Explained*. The ranking is based on closeness centrality. Every paragraph revolves around one topic, clearly established by the authors and reported here too.

Rank/Topic	*Interactions*	*Emergence*	*Dynamics*	*Self-organisation*	*Adaptation*	*Interdisciplinarity*	*Methods*
1	component	system	system	pattern	adapt	system	compute
2	interact	property	change	emerge	system	understand	model
3	whole	part	state	may	able	science	method
4	study	component	behaviour	organ	become	complex	mathematics
5	system	whole	point	become	function	use	lead
6	make	sum	show	interact	damage	variety	require
7	difficult	phenomenon	variable	system	evolve	manage	involve
8	part	complex	dynamic	produce	go	domain	analysis
9	new	exhibit	tend	property	may	ecology	forecast
10	consist	deduce	depend	lead	robust	biology	rule

RESULTS

The main results of this manuscript are twofold. On the one hand, the current analysis provides a cross-validation of the information revealed by TFMNs through a benchmark on short texts, revolving around specific topics. Comparison with validated behavioural cognitive data indicates that the structure of TFMNs captures semantically relevant information about text like topics or conceptual relevance. On the other hand, the rest of the Results section focuses around reconstructing and analysing a data-driven picture of the online perception of women in science and the gender gap as reported by online users on Twitter.

Benchmark of text-based forma mentis networks on short texts

This subsection reports results of the benchmarking analysis of "Complexity Explained" (see Methods) through TFMNs. For each of the seven paragraphs of the booklet a TFMN was built. Closeness centrality was used for identifying the most central concepts in every TFMN and produce rankings of the most relevant words in each text. Overall, the resulting forma mentis network contained a median of 49 concepts and were fully connected. Table 1 reports the 10 most central words in each network and their underlying topic.

Table 1 indicates that, beyond an overall agreement between key concepts and topics, as identified by TFMNs, the network topology of syntactic/semantic associations identified the most distinctive conceptual features of topics. For instance, while the ranking for the topic "Interactions" reported mainly concepts relative to structure (e.g., *component, part, whole, consist*), the ranking of "Dynamics" identified concepts related to the evolution over time of a system (e.g., *change, state, behaviour, dynamic*). Words expressing resilience and robustness to attacks, like "damage", "function", "evolve", "adapt" and "robust", were found to be central in the textual forma mentis network of "Adaptation". TFMNs detected "domain" as being a relevant concept in the "Interdisciplinarity" paragraph. This was expected in a text describing complexity science as an umbrella for different research areas. The "Methods" paragraph revolved, as expected, around quantitative concepts, e.g. *forecast, compute, model, method, analysis, rule*. In the same paragraph, mathematics was found to be highly relevant, in agreement with the overall necessity of a mathematical language for investigating complex systems (i Cancho et al., 2004; Siew et al., 2019).

Although the above qualitative analysis indicated an overall agreement between topics and identified key concepts, a more quantitative approach was further pursued. With the aim of assessing whether the identified concepts were more or less semantically related to the topics designed by text authors, free associations and semantic network distance were adopted. In networks of free associations, nodes/words are linked if they elicit a quick recall of each other in a behavioural task (De Deyne et al., 2019). Previous studies have shown how semantic network distance on networks of free associations are a good proxy of semantic relatedness, superior also to semantic latent analysis (Kenett et al., 2017). Although several datasets for free associations are available in the literature, this benchmark considered two: the large-scale Small World of Words gathered by De Deyne and colleagues (De Deyne et al., 2019) and the small-scale Complex Forma Mentis project gathered by Stella and colleagues (Stella et al., 2019). Many of the scientific terms present in Complexity Explained were absent in the Small World of Words but present in the STEM free associations provided by Stella et al. (Stella et al., 2019). Therefore, this analysis focused on the second dataset of free associations.

The distance between every relevant word and its reference topic on the network of free associations by (Stella et al., 2019) was computed for all topics within the networked mindset of 59 complexity researchers. These topics were "interaction", "emergence", "dynamics", "self-organisation", "interdisciplinary" and "methods" and the relevant concepts were the ones in Table 1. The resulting set of empirical semantic network distances was then compared against a reference null model with randomised TFMNs having the same number of links and nodes of the original networks but with randomly reshuffled links (i.e. configuration models (Stella et al., 2018a)). The reshuffling disrupted semantic relationships between network structure and meaning (Stella et al., 2018a). A location test between the empirical and the randomised network distances (over 50 network realisations) indicated a statistically significant difference between the clustering of relevant words around each topic and the null model (Mann-Whitney test, $U = 17827$, p-value: $0.0267 < 0.05$) at a significance level of 0.05. In the networked free associations representing the scientific knowledge of complexity researchers, the words identified as relevant for a topic by TFMNs tended to be closer to their own topic (median network distance: 3.1) than on the randomised networks (median network distance: 3.7). Since semantic network distance on networks of free associations indicates semantic relatedness (Kenett et al., 2017), these results confirmed that closeness centrality on TFMNs was capable of identifying concepts relevant to the specific topic underlying a given text.

Given the successful outcome of such benchmark, TFMNs can therefore be applied to extracting relevant features of other short texts. The following section reports on the results of TFMNs when used for analysing over 10000 tweets about gender gap in science.

Analysis of the perception of gender gap in STEM with cognitive networks

The textual forma mentis network obtained from processing the selected tweets included a largest connected component including 3005 stemmed concepts and 28004 connections (24693 in the syntactic layer and 3311 in the synonyms layer). The 10 most central concepts in terms of closeness centrality were: "stem", "science", "we", "do", "you", "learn", "make", "get", "work", "need". As expected from the benchmark, these words capture the general context of science of the investigated online discourse. Concepts like "woman" and "man" ranked 103rd and 246th, respectively, mirroring the relevance of women in science for the selected online discourse. "Scientist" ranked higher, at the 54th position, further confirming the scientific scope of the dataset.

0.0.1 Conceptual associations with and around a concept identify word clusters

The TFMN was more clustered than random expectation (using a single-layer clustering coefficient where all links are aggregated together, cf. (Siew et al., 2019)). Concepts linked to a common neighbour tended to get connected with each other too. The empirical network displayed a mean clustering coefficient of 0.327 (0.166 ± 0.007 for reference configuration models (Stella et al., 2018a)). Hence, in the TFMN concepts clustering around a given word (e.g. "STEM") shared syntactic/semantic links too, a tendency that can provide a richer structure about the conceptual organisation of knowledge around specific concepts/topics in terms of (more) conceptual links. Consequently, investigating clustered networked neighbourhoods of words can provide contextual information that would be lost by considering either words in isolation or only the list of associates to a given concept. For this reason, the investigation focused on word clusters in order to better understand the online perception of the STEM gender gap.

0.0.2 Valence auras and global network metrics highlight an overall positive online stance towards STEM and gender gap

The 3005 stemmed concepts extracted from relevant tweets and connected in the forma mentis network were rated as positive (1045), negative (430) and neutral (1943) according to the procedure described in the Methods. Positive concepts were found to have a higher median degree than negative concepts (Mann-Whitney test, numerosity $n_1 = 1045$, median degree $k_1 = 9$, $n_2 = 430$, $k_2 = 4$, $U = 3 \cdot 10^6$, $p < 10^{-6}$). Positive concepts were also found to be more central in the reconstructed mindset in terms of requiring fewer syntactic/synonymy associations in order to reach any other connected concept, i.e. centrality as expressed by multiplex closeness (Stella et al., 2018a), (Mann-Whitney test, numerosity $n_1 = 1045$, median degree $k_1 = 0.3601$, $n_2 = 430$, $k_2 = 0.3307$, $U = 3 \cdot 10^6$, $p < 10^{-6}$). These comparisons indicate that in the analysed corpus of twitter language focusing on women and STEM, positive concepts were more predominant, more well connected and more central than negative concepts. The richer network structure of conceptual associations of positive concepts translated into a generally positive attitude of twitter users towards the STEM gender gap.

Figure 2 (top) reports the emotional auras (Stella et al., 2019; Stella, 2020) of hub concepts in the reconstructed mindset of social media users. While single-word labels were determined from the sentiment of a general population, i.e. participants in an affect mega-study (Warriner et al., 2013), auras emerged from the specific syntactic/synonymy relationships traced in the analysed data and therefore characterised the reconstructed mindset of the population of interest. Concepts like "stem", "gender" and "gap" are neutral in commonly spoken language but were associated mostly with positive concepts in the language of the online discourse, i.e. surrounded by a positive emotional aura. This indicates an overall positive perception of these topics that would go *undetected* when investigating "stem", "gender" and "gap" in isolation. Differently from other investigations with forma mentis networks (Stella et al., 2019; Stella, 2020), the sampled online audience reported a strongly positive perception/aura of "student". Also "bias", a negative concept, was surrounded by a positive emotional aura, indicating an overall mixed stance trying to figure out positive aspects of a bias (e.g., how to overcome it). Further analysis is required in order to better understand the above perceptions.

Figure 2 (bottom) reports the mindset structure around "bias" and other concepts like "unique", "hurt" and "journal". The forma mentis neighbourhood of "bias" included associations to positive concepts like "face", "win", "clear" and to negative concepts like "fight", "cost" and "stereotype", all revolving around contrasting and overcoming biases. Therefore, the reconstructed TFMN indicates how social media users discussed about "bias" in terms of a negative entity to be contrasted, overcome and won, thus explaining the above mixed perception. Other words associated to bias gravitated around the gender pay gap and economic implications (Courey and Heywood, 2018), e.g. "earn", "cost", "value". Links with "unconscious" and "fact" indicate that social media users were aware of hidden, unconscious gender biases affecting the gender gap (Pietri et al., 2018).

Figure 2 (bottom) also reports how "unique" was associated in the online discourse around the gender gap. The mixed perception of "unique" included negative associations to concepts eliciting loneliness, e.g. "oppress", "fail", "alone" and "pretend", which are unexpected when considering the positive perception of "unique" itself. Hence, the TFMN indicates that when associated in messages focusing on women in STEM, the meaning of "unique" shifted from positive to mixed and included negative connotations of social exclusion and sense of failure. This is an example of *contextual valence shifting* (Polanyi and Zaenen, 2006), a phenomenon in which a concept changes its valence according to its semantic context. TFMNs represent a valuable quantitative tool for identifying valence shifting through semantic associates and emotional profiling.

The multiplex structure of TFMNs can influence an emotional aura. Synonymy links (green) do not come from text but rather from a pre-determined lexicon of synonyms (i.e. WordNet (Miller, 1998)) that has general validity over common language. Instead, syntactic (blue/red/gray) associations come from the specifically analysed language. In this way, TFMNs can identify also missing conceptual associations or the predominance of synonymy over syntactic links.

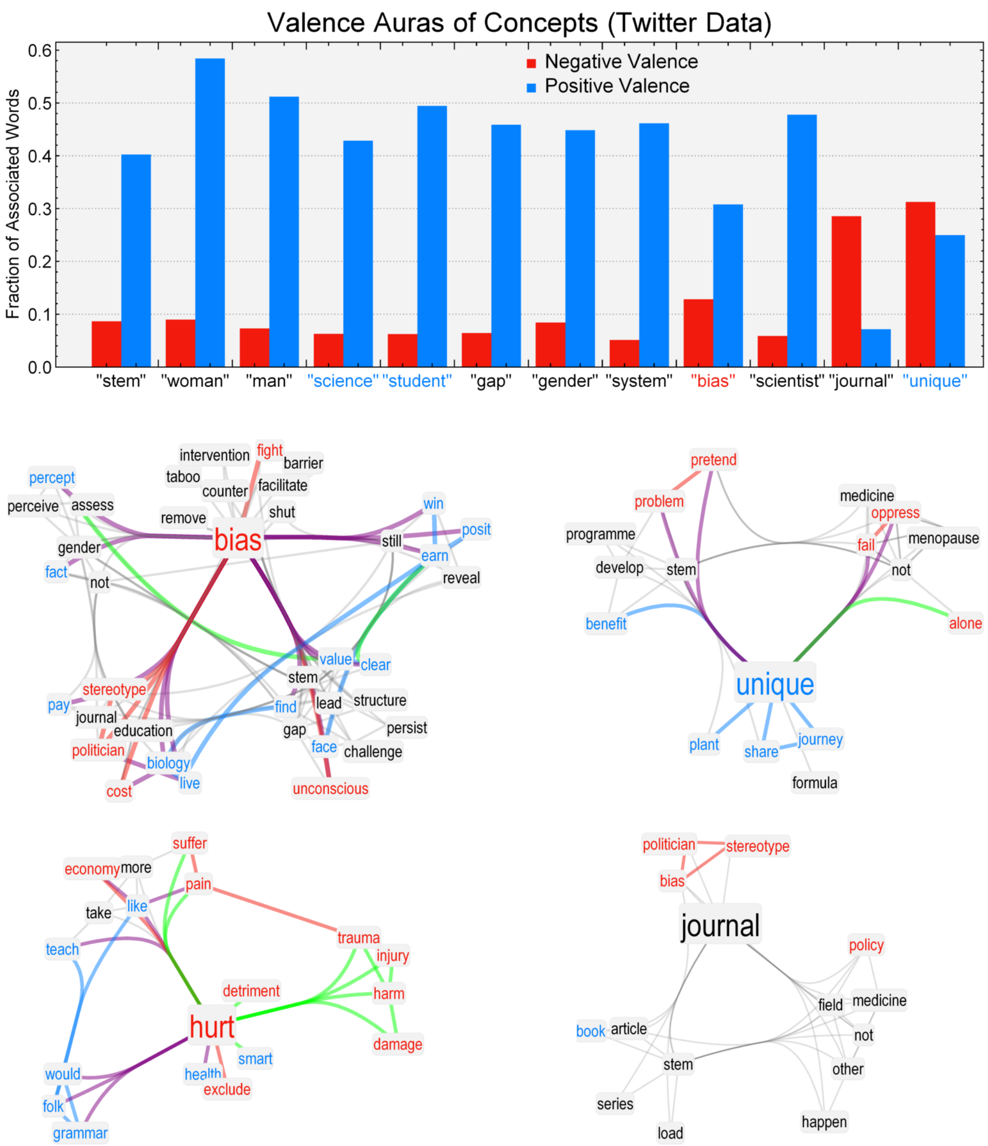

Figure 2. **Bottom:** Valence auras in the global forma mentis network of all 10384 analysed tweets. Positive (negative) words are highlighted in cyan (red). A fraction of 0.4 indicates that 40% of the neighbours of a word (e.g. STEM) are labelled as positive. **Top:** Textual forma mentis network for words linked to the target topics like "bias" (top left), "unique" (top right), "hurt" (bottom left) and "journal" (bottom right).

The negative aura of "hurt" in Figure 2 was mostly due to general synonyms whereas specific syntactic associations included positive concepts like "teach", "like" or "health", thus indicating a lack of hurtful/painful conceptual associations in the investigated online discourse over women in STEM. Network structure was informative also about the negative aura attributed to "journal", which was linked to "politician", "stereotype", "bias" and "policy". These syntactic links indicated a social awareness about journals and news media potentially perpetrating gender stereotypes, a role that was investigated in previous studies (Steinke, 2017).

The online social perception of "woman", "man" and "person" reconstructed by TFMN

Although it is expected for the online discourse to feature also automatic accounts and social bots, previous research identified human users as being more predominant in driving and diffusing message exchanges on social platforms (Stella et al., 2018b). Comparing how people identify themselves in a discourse where they are the main drivers can be informative about social roles and users' self-perception (Varol and

Uluturk, 2020).

Figure 3 reports and compares the forma mentis network around “woman” (top left), “man” (top right) and “person” (middle right). All these three concepts, commonly perceived as positive entities, were surrounded by positive emotional auras in agreement with the overall positive features reported above of the whole TFMN.

The reconstructed social perception of “woman” within the online discourse of gender gap in STEM identified mainly semantic clusters expressing three aspects:

1. leadership and professional recognition in STEM (e.g. “enterpreneur”, “award”, “recognize”, “leader”, “power”, “career”, “success”, “deserve”, “science”, “valid”);

2. learning and education (e.g. “teacher”, “inspire”, “learn”);

3. social welfare (e.g., “finance”, “advocate”).

All these aspects were dominated by positive, concrete concepts eliciting a sense of achievement and professional establishment. The resulting picture was an overall positive sentiment/perception of the online social discourse about the figures of women in science. However, within this positive landscape, a closer look at the local community of concepts clustering around “woman” identified also negative associations, further highlighted in Figure 3 (middle left).

The analysed text associated women also to “indebted”, “kill”, “lone” and “cry”, highlighting a multifaceted perception of women in STEM and including negative traits that might originate from *stereotype threat* (Shapiro and Williams, 2012; Pennington et al., 2016). In cognitive psychology, stereotypes can advantageously form a belief about or characterise key traits of groups of people through little cognitive effort. Stereotypical perceptions come at the cost of being not grounded in empirical knowledge, providing inexpensive but potentially erroneous information that can: deeply affect perception and cognitive processing (Pennington et al., 2016), induce anxiety, and reduce performance under pressure, e.g. causing a sense of threatening. For instance, STEM stereotypes like “girls are not good at maths” have been reported to affect even the self-perception of female STEM students, causing unconscious biases which reduced girls’ performance and retaining of STEM subjects (Shapiro and Williams, 2012; Steinke, 2017; Chavatzia, 2017). The TFMN reported in Figure 3 highlighted the stereotype of the “oppressive, lonely, big-shot woman in STEM”, relative to women in STEM achieving successful careers only at the cost of sacrificing empathy and other positive personality traits (Ely et al., 2011; Madsen and Andrade, 2018). Recent studies have overwhelmingly exposed such stereotype, cf. (Steinke, 2017), and the permanence of these negative associations in online social discourse represents evidence of how such stereotype is still deep-rooted in daily communication. Notice that this might have also positive repercussions, as improving the awareness of the effects of stereotype threat by simply talking about it can boost self-perception and promote the implementation of effective countermeasures fighting stereotypes (Madsen and Andrade, 2018).

The reconstructed online perception of women was overwhelmingly positive and it indicated how women’s success and leadership in STEM deserve recognition (Madsen and Andrade, 2018). This perception has to be compared against the one of males. Figure 3 (top left) considers the forma mentis network around “man”, which included mostly positive concepts like “science”, “passion” and “smile”. However, syntactic links associated “man” also to negative concepts such as “rape”, “issue” and “racist”, indicating the presence in the online discourse of forms of accusations or condemnations of the role played by men in social issues like sexual harassment, rape and racism (Karami et al., 2020). In the forma mentis around “man”, special attention has to be devoted also to the negative particle “not”, which is syntactically linked to “god”, “think”, “make” and “consider”. The negation of all these concepts, as indicated by the syntactic links, provided a multi-faceted perception where men in STEM were related to a superiority complex, a stereotypical self-perception of superiority in science-related achievements, for instance in mathematics assessments (Leyva, 2017), promoted by preliminary and incomplete studies (cf. (Leyva, 2017) and (Chavatzia, 2017)). This quantitative result, embedded in language and highlighted by forma mentis networks, indicates that fighting the gender gap in STEM means also changing men’s stereotypical roles in science.

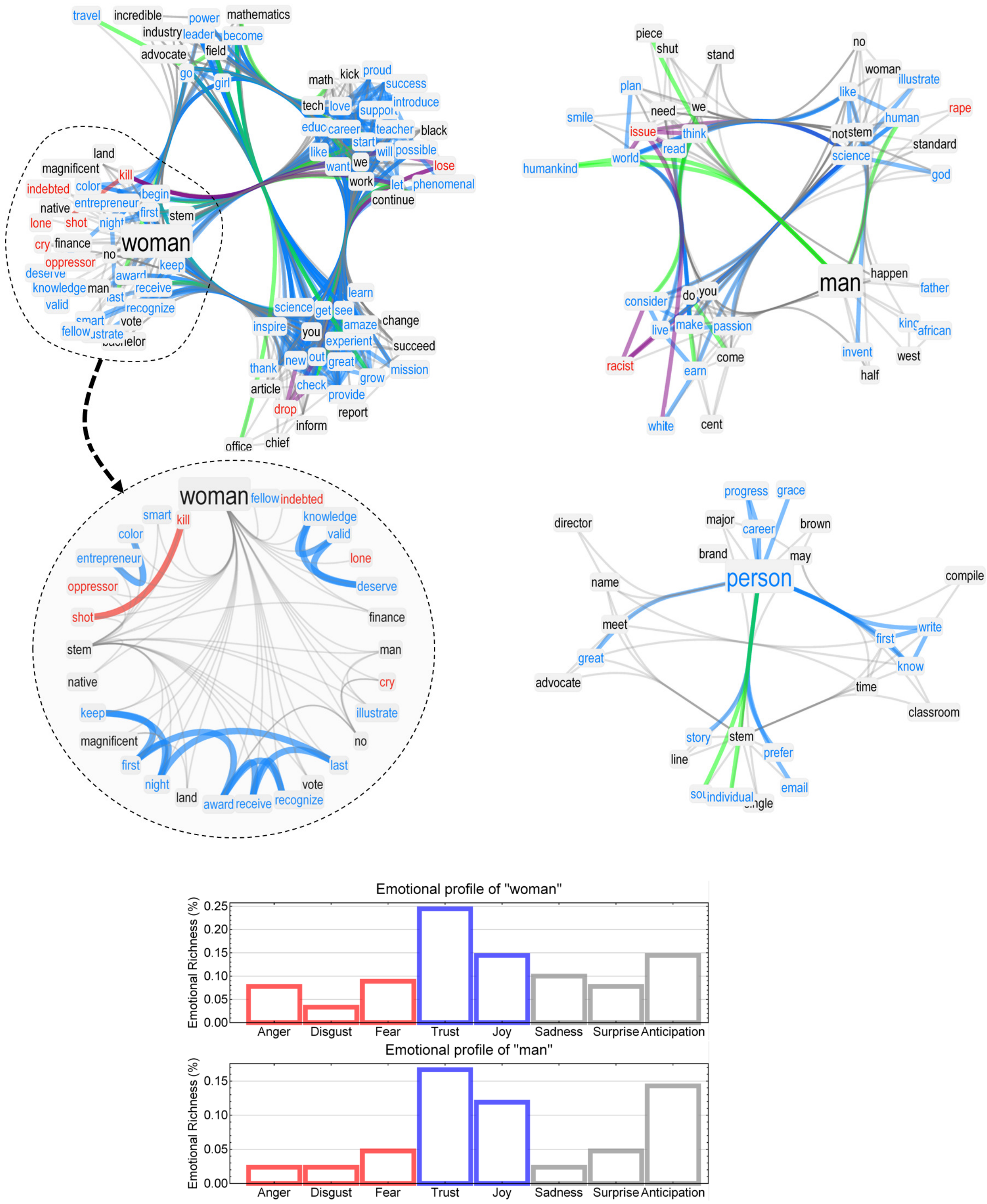

Figure 3. Top: Textual forma mentis network for words linked to the target topics like "woman" (top) and "man" (bottom). On every row, networks on the left include all neighbours adjacent to the target topic and clustered in communities as obtained from a Louvain algorithm (for more details on community detection in Twitter data see (Konstantinidis et al., 2017)). Networks on the right include all words in the same community as the target topic (top) and all words (bottom) but use hierarchical edge bundling for highlighting within-community clusters. In every visualisation, positive (negative) words and links are highlighted in cyan (red). Links between positive and negative concepts are reported in purple. Semantic links between synonyms are in green. **Bottom:** Emotional profile/richness for the neighbourhood of "woman" and "man", indicating what emotions elicit the associates of these concepts.

In terms of emotional profiles, both “woman” and “man” included associates evoking mainly trust and joy. The TFMN around “woman” was slightly richer in anger-eliciting words, like “kill” or “shot”, than the TFMN around “man”. In Figure 3, the TFMN around “person” was devoid of any negative concept or sign of stereotype threat. Word clusters related “person” to a multifaceted perception about career progression, mindfulness and education. The associates of “person” evoked mostly trust (emotional profile not shown for brevity), further confirming the positive perception of “person” itself. Given the gender-less dimension of “person” in English, it is interesting to underline how the above negative perceptions are more tightly connected to gender-rich words for human beings like “man” and “woman” and not to genderless words like “person”. Since words in language do not only describe reality but rather contribute to forging it (Malt et al., 2010), the above result suggests that reducing the incidence of the gender gap might be possible also by speaking more and more in terms of “persons in STEM” rather than in terms of contrasting “men versus women in STEM”, always while respecting individual differences.

The online social perception of gender gap and scientist stereotypes as reconstructed by TFMN

This subsection investigates “gender” and “gap” and their online perceptions. Figure 4 considers the TFMN around “gender” and “gap”, which were both considered as neutral concepts in language. The online discourse over the social platform associated these two concepts with each other, as expected, and related both of them mainly to positive words. Hence, the overall online emotional aura of “gender gap” was mostly positive but it also included some negative associates. The community structure for the TFMN of “gender” identified three perceptual dimensions in the way online users talked about gender:

- a positive dimension of respect and consideration (with associates like “balance”, “close”, “consider” and “respect”);
- a research dimension in understanding the science of gender (with associates like “research”, “support”, “future”, “highlight” and “explain”);
- a mostly negative dimension of gender imbalance and biases (with associates like “doubt”, “imbalance” and “attack”).

Associations of “gender” with concepts like “stereotype”, “unfounded”, “break” and “tackle” (see Fig. 4, top left) indicated an attitude of opposition, among online users, against gender stereotypes in the considered online communications.

Analogously to “gender”, also “gap” included negative associations to “bias” and “attack” and positive links to concepts like “overcome”, “close” and “consider” (see Fig. 4). The conceptual cluster around “gap” highlighted in Fig. 4 (middle left) contained associations mostly focusing on:

1. the expectation to overcome and reduce the gender gap in the future (e.g.,“reverse”, “break”, “crack”, “bridge” and “reduce”);
2. the economic impact of the gender gap (e.g., “pay”, “income”, “wage”).

These two dimensions indicate a positive intention for the online discourse to tackle and reduce gender gaps, while displaying awareness about the gender pay gap (Courey and Heywood, 2018) and also an emotion of anticipation or projection of such challenge in the future (see emotional profiling in Fig. 4, bottom). This provides additional quantitative indication for the importance of achieving equal wages in STEM in order to reduce the gender gap, in agreement with previous studies (Courey and Heywood, 2018). Figure 4 contains also negative associations between “gap”, “bias” and “unconscious”, indicating a remarkable awareness of online users about the gender gap being rooted in unconscious bias and gender preconceptions that are difficult to detect and act upon (Karami et al., 2020; Steinke, 2017; Moss-Racusin et al., 2012; Pietri et al., 2018).

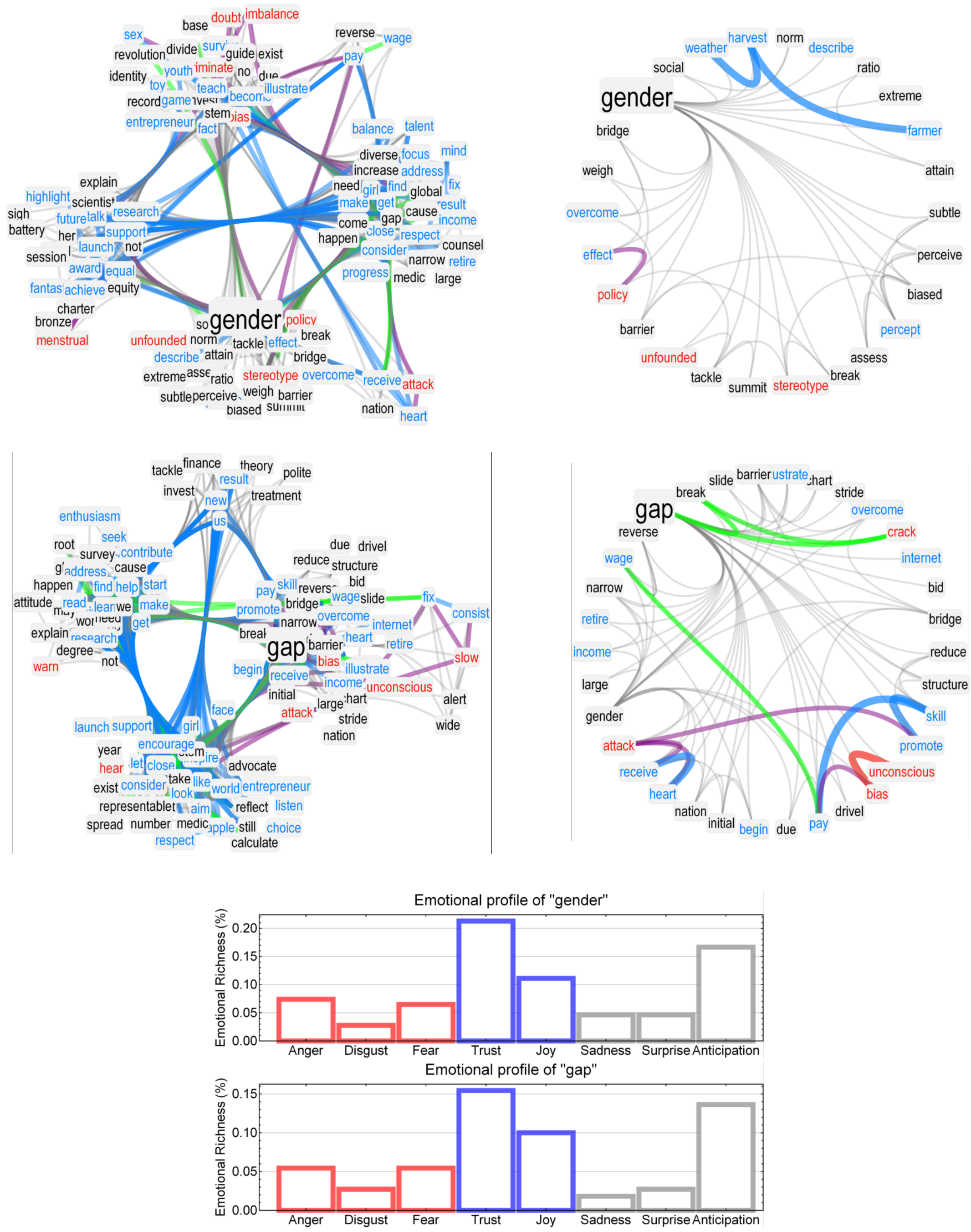

Figure 4. Top: Textual forma mentis network for words linked to the target topics like "gender" and "gap". On every row, networks on the left include all neighbours adjacent to the target topic and clustered in communities as obtained from a Louvain algorithm. Networks on the right include all words in the same community as the target topic but use hierarchical edge bundling for highlighting within-community clusters. In every visualisation, positive (negative) words and links are highlighted in cyan (red). Links between positive and negative concepts are reported in purple. Semantic links between synonyms are in green. **Bottom:** Emotional profile/richness for the neighbourhood of "woman" and "man", indicating what emotions elicit the associates of these concepts.

As already mentioned above, a prominent mechanism of unconscious bias is stereotype threat, where unconscious perceptions cause anxiety and negative latent emotions that influence performance, e.g. girls aware of the preconception that "girls are not good in science" end up performing worse than males in STEM tasks (Moss-Racusin et al., 2012; Shapiro and Williams, 2012; Madsen and Andrade, 2018; Chavatzia, 2017). Given the awareness of online users about unconscious biases, it becomes relevant to explore the networked TFMN mindset in search of signs of stereotype threat. Because of the STEM scope of the dataset, the focus here was devoted to detecting stereotypical perceptions in the figure of "scientist", whose neighbourhood is reported in Fig. 5. The TFMN quantified positive semantic associations surrounding "scientist". Clusters of more tightly associations identified a strong perception of scientists in relation to innovation (e.g., "tech", "society", "entrepreneur", "innovate", "discover", "leadership"). A closer look (cf. Fig. 5, left) revealed also enthusiasm-eliciting concepts (e.g., "great", "notable", "famous", "celebrant", "dream", "congratulate") and career-related jargon (e.g., "showcase", "assess", "work", "peer", "academia"). Hence, scientists were perceived as successful professionals by online users, analogously to what other social groups like high-school students did in other studies (cfr. (Stella, 2020)). Even within this success-centred perception, online users linked scientists to concepts like "suffer", "pain" and "gain", displaying awareness about the cost of success and the need for hard work.

Overall, the mental construct of "scientist" represented by the TFMN identified a well-rounded, balanced and positive online perception of scientists in terms of successful, hard-working professionals, devoid of any significant patterns of stereotype threat associating scientists only to a male-gender sphere, differently from stereotypical perceptions found in other datasets and groups via the Implicit Association Test (Lane et al., 2012; Steinke, 2017).

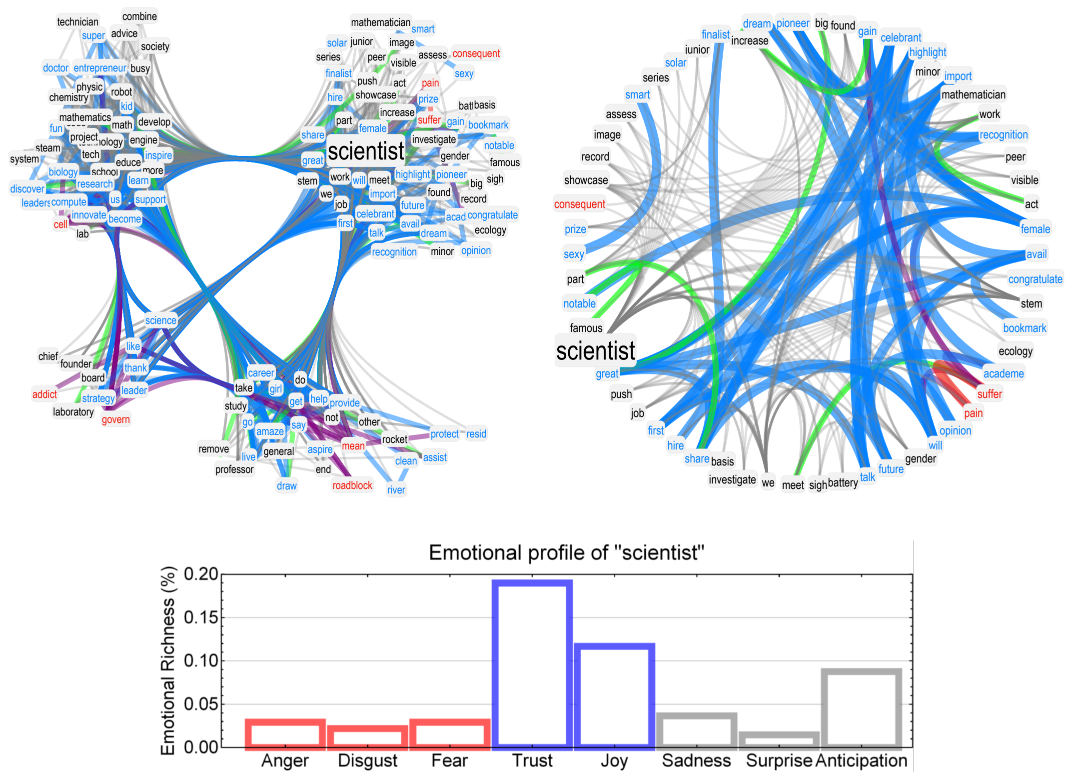

Figure 5. Top: Textual forma mentis network for "scientist". The network on the left includes all words associated to "scientist", whereas plot on the right zooms in the community of words more tightly connected to "scientist". Positive (negative) words and links are highlighted in cyan (red). Links between positive and negative concepts are reported in purple. Synonymy relationships are in green. **Bottom:** Emotional profile/richness for the neighbourhood of "scientist", indicating what emotions elicit the associates of "scientist" itself in the analysed corpus.

DISCUSSION

As outlined in recent Big Data Analytics studies (Huang et al., 2020; Odic and Wojcik, 2019; Chavatzia, 2017), differences in gender discrimination are strongly influenced by distorted mindsets, with deep, often hidden, repercussions over pay gaps (Courey and Heywood, 2018) and sexual harassment (Karami et al., 2020). Combining these findings with the ever-increasing influence of social media over real life (Jansen et al., 2009; Waqas et al., 2019; Stella et al., 2018b; Nasar et al., 2019) highlights an urgent necessity for using information processing in order to understand "if" and "how" specific massive online social platforms promote information on distorted mindsets. The tackling of such research question is essential for countering gender biases with data-informed approaches (Huang et al., 2020). This work tackled this question by extracting, reconstructing and understanding with textual forma mentis networks (TFMNs) how online users perceived and discussed the topics of "women in STEM" and the "gender gap" on Twitter.

As a knowledge extraction technique (Nasar et al., 2019), TFMNs indicated an overall positive attitude of the online discourse towards tackling and reducing gender inequality. Going beyond the positive/negative polarity of standard sentiment analysis (Mohammad, 2016; Stella et al., 2018b), this work attributed TFMNs with a richer emotional profile, using emotional data from (Mohammad and Turney, 2013) and providing a quantitative way for detecting *contextual connotation shifting* (Polanyi and Zaenen, 2006), i.e. a concept being perceived in different ways according to its semantic context and associates. The reconstructed forma mentis networks outlined "man", "woman" and "scientist" as being associated with concepts eliciting predominantly trust, joy and anticipation and, with way less intensity, also anger and fear. Trust is a feeling of confidence, security and positive endorsement (Ekman and Davidson, 1994), its predominance in the retrieved TFMNs suggests a willingness for online users to provide and exchange endorsements to each others' messages while debating the topic of "women in science". This result is in contrast with a previous study reporting Twitter as being more prone to host general-level negative rather than positive emotional content (Waterloo et al., 2018). This disparity might be explained by considering that Twitter interactions focus mainly over weak social ties, including mostly acquaintances and casual contacts rather than strongly personal relationships (Waterloo et al., 2018). Since the expression of topic-inspired negative emotions along weak social ties is considered being less acceptable (Ferrara and Yang, 2015), more positive emotional profiles would be expected from topic-specific Twitter public debate, and not from general level content sharing like the one in (Waterloo et al., 2018). As a future research direction, it would be interesting to detect whether the mostly trustful and positive perception of gender biases would persist also on other social media platforms with stronger social ties such as Facebook or WhatsApp (Waterloo et al., 2018).

Another emotional state prominently featured in all the reconstructed stances towards "woman", "gender", "gap", "man" and "scientist" was anticipation, a neutral emotional state including either pleasure or anxiety towards future events (Ekman and Davidson, 1994). The predominance of anticipation and joy with respect to other emotions like sadness, fear, disgust or anger, suggests the prevalence of positive expectations in the analysed text, examples being the excitement-related concepts associated with "scientist" or the successful dimension attributed to women in STEM. These patterns indicate the occurrence of messages celebrating women in science and their success, contrasting the gender gap with stories of excitement and professional achievement. Extensive research (cfr. (Pietri et al., 2018; Chavatzia, 2017; Madsen and Andrade, 2018)) indicates that promoting professional achievements of women in science has strong beneficial effects in favouring women's representation in STEM, as it enables girls in identifying relatable and inspiring stories of success in STEM going beyond discrimination.

The reconstructed TFMNs reported evidence for online users being aware about unconscious gender biases (Shapiro and Williams, 2012). These biases occur when an individual consciously rejects gender stereotypes but is still influenced by and makes unconscious evaluations based on such stereotypes themselves, see also (Madsen and Andrade, 2018; Ely et al., 2011). At the individual level, unconscious and passive discrimination based on gender stereotypes can have smaller repercussions in comparison to actively promoting gender inequality. However, at the group level, many unconscious small biases can interact in a complex systems fashion (Hogue and Lord, 2007) and lead to the emergence of "powerful yet often invisible barriers" of gender discrimination (Ely et al., 2011). These barriers undermine considerably women representation and professional growth in a variety of fields including also STEM (Shapiro and Williams, 2012; Lane et al., 2012; Madsen and Andrade, 2018; Huang et al., 2020; Odic and Wojcik, 2019). Forma mentis networks captured signals of unconscious gender bias in online discourse around "women in science". These conceptual links can be beneficial in promoting awareness about the above invisible barriers, further suggesting the extreme importance of fighting implicit biases in STEM careers for closing the gender gap in STEM (Pietri et al., 2018).

Another prominent topic emerging from the TFMNs is the gender pay gap, a mismatch between the salaries of individuals of different genders performing the same job (Courey and Heywood, 2018). The online perception reconstructed by forma mentis networks indicates that pay gaps are closely semantically related to both “gender” and “gap”, thus indicating that closing the gender pay gap is key for fighting gender biases in STEM, in agreement with previous relevant studies (Ely et al., 2011; Courey and Heywood, 2018).

Although partial evidence for a stereotypical, angry-eliciting and fearful perception of women in STEM as “lone survivors” was present in the neighbourhood of “woman”, the overall perception of such concept was positive and elicited mostly trustful and joyous concepts, celebrating women’s success in STEM. The reconstructed role of “scientist” did not include stereotypical conceptual associations, indicating a lack of stereotype threat phenomena (Shapiro and Williams, 2012) in the considered online discourse. In cognitive psychology, stereotype threats represent unconscious mechanisms that affect performance of a given group in relation to the stereotypical expectations commonly shared about that group (e.g. “girls are bad at maths” is a stereotype harming girls’ performances in maths, cfr. (Pennington et al., 2016; Chavatzia, 2017)). Forma mentis networks mostly related the online perception of scientists to success and career progression and highlighted a lack of stereotypical perceptions, differently from the detection of scientist-centred stereotypes found in previous approaches with cognitive network science (Stella, 2020), via the Implicit Association Test (Lane et al., 2012). Such virtuous finding might be the consequence of the relatively high participation of STEM professionals over the Twittersphere, which can disrupt stereotypical perceptions. As future research, it would be interesting to apply TFMNs for detecting potential patterns of stereotype threat in other social platforms like Facebook or WhatsApp, that are more exposed to negative content (Waterloo et al., 2018) and also less prone to hosting posts from STEM experts.

From a methodological perspective, TFMNs rely on the recent theory of cognitive networks (Siew et al., 2019) and include syntactic and semantic conceptual associations which are informative of the structure of knowledge perceived by text authors (Stella et al., 2018a; Stella, 2019). Differently from other successful models of knowledge representation such as concept maps (Dóczi, 2019) and knowledge graphs (Amancio, 2015; Akimushkin et al., 2017; de Arruda et al., 2019), forma mentis networks contain also emotional information, outlining sentiment and emotional patterns in the way individuals assembled their stance in a text. This contextual information is essential for the interpretability of a detected stance (Nasar et al., 2019).

The framework of textual forma mentis networks reported here has some important limitations. Differently from the behavioural forma mentis networks introduced in previous studies (Stella and Zaytseva, 2020; Stella et al., 2019; Stella, 2020), in TFMNs the valence of concepts represents population-level averages extracted from mega-studies in psycholinguistics (Warriner et al., 2013). This representation assumes an overall shared perception of the emotional content of concepts that might be preserved at the global level, i.e. on an online platform where large numbers of users of multiple backgrounds interact with each other. However, this assumption might be violated within specific populations. For instance, Stella and colleagues (Stella et al., 2019) showed that a population made entirely of high-school students perceived “maths” as a negative concept whereas a population of international researchers perceived the same concept as positive. The overall perception of “maths” reported in the language norms by Warriner et al. (2013) was neutral, as reported also in the TFMNs presented here (Warriner et al., 2013). It is important to keep this assumption in mind when applying TFMNs to mindset reconstruction in specific and non-heterogenous populations. Nonetheless, even in these populations TFMNs can be informative about conceptual associations as expressed by the semantic/syntactic multiplex network structure, since the emotional attitude towards a concept can be reconstructed from the attitude and meaning of its associates (Polanyi and Zaenen, 2006). As a future research direction, a potential approach for achieving population-specific valence labels would be valence extraction from text, a technique that exploits word embedding and machine learning (Mohammad, 2016; Rudkowsky et al., 2018) and as such works well for longer texts like books or essays but is still relatively less accurate for shorter texts like tweets or posts, which are less rich in semantic information (Polanyi and Zaenen, 2006).

Another assumption underlying TFMNs is treating modifiers of meaning (e.g. negations) as nodes equivalent to concepts. Modifiers can alter the meaning expressed by individual concepts, providing a contextual richness that is not captured by the so-called “bag of words” models (Polanyi and Zaenen, 2006; Rudkowsky et al., 2018), where a text is represented by an unstructured list of its concepts. Textual forma mentis networks provide syntactic and semantic contextual background to concepts so that the investigation of modifiers cannot be independent on the analysis of conceptual associates. For instance, in the current approach, negations were considered in the emotional profiling of conceptual neighbourhoods in the following way: antonyms of words linked to negations were added to emotional counts, providing information about opposite meanings in addition to the concepts originally available in the TFMN (e.g., if “appreciation” was connected to “not” then its antonym “disgust” was considered in the emotional profiling too). Other ways of grading the intensity of statements or negations could be tested and implemented within future work.

Another future research direction would be the implementation of TFMNs in synergy with social network analysis, much alike what recent approaches suggested (Rodrigues and Pietrocola, 2020; Stella et al., 2018b), in order to better capture online social behaviour (Varol and Uluturk, 2020) or better understand misinformation spread (Pierri et al., 2020).

CONCLUSIONS

Textual forma mentis networks (TFMNs) provide contextual conceptual and emotional information from text, providing a rich picture of how text authors perceived and associated multiple topics. When applied to the social media discussion of women in science, TFMNs identified a mostly positive, trustful and anticipation-rich discourse but also highlighted online awareness about relevant issues like unconscious gender biases and the gender pay gap. The application of TFMNs to the analysis of online discourses opens new ways for quantitatively assessing the role played by social media in promoting/hampering gender biases and distorted mindsets. Textual forma mentis networks can open new ways of accessing social media, understanding the content of online communication and providing new bridges for linking social dynamics with cognitive and emotional information spread.

ACKNOWLEDGEMENTS

The author acknowledges Prof. Jim Jansen for valuable feedback on this work.

REFERENCES

Aitchison, J. (2012). *Words in the mind: An introduction to the mental lexicon*. John Wiley & Sons, USA.

Akimushkin, C., Amancio, D. R., and Oliveira Jr, O. N. (2017). Text authorship identified using the dynamics of word co-occurrence networks. *PloS one*, 12(1).

Amancio, D. R. (2015). Probing the topological properties of complex networks modeling short written texts. *PloS one*, 10(2).

Amancio, D. R., Oliveira Jr, O. N., and Costa, L. d. F. (2012). Structure–semantics interplay in complex networks and its effects on the predictability of similarity in texts. *Physica A: Statistical Mechanics and its Applications*, 391(18):4406–4419.

Beasley, A. and Mason, W. (2015). Emotional states vs. emotional words in social media. In *Proceedings of the ACM Web Science Conference*, pages 1–10.

Bovet, A., Morone, F., and Makse, H. A. (2018). Validation of Twitter opinion trends with national polling aggregates: Hillary Clinton vs Donald Trump. *Scientific Reports*, 8(1):1–16.

Cancho, R. F. I. and Solé, R. V. (2001). The small world of human language. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 268(1482):2261–2265.

Castro, N. and Stella, M. (2019). The multiplex structure of the mental lexicon influences picture naming in people with aphasia. *Journal of Complex Networks*, 7(6):913–931.

Chavatzia, T. (2017). Cracking the code: Girls’ and women’s education in science, technology, engineering and mathematics. *UNESCO*, (48):1–16.

Courey, G. and Heywood, J. S. (2018). Gender Wage Gap Trends Among Information Science Workers. *Social Science Quarterly*, 99(5):1805–1820.

de Arruda, H. F., Marinho, V. Q., Costa, L. d. F., and Amancio, D. R. (2019). Paragraph-based representation of texts: A complex networks approach. *Information Processing & Management*, 56(3):479–494.

De Deyne, S., Navarro, D. J., Perfors, A., Brysbaert, M., and Storms, G. (2019). The “small world of words” english word association norms for over 12,000 cue words. *Behavior Research Methods*, 51(3):987–1006.

Dóczi, B. (2019). An overview of conceptual models and theories of lexical representation in the mental lexicon. *The Routledge Handbook of Vocabulary Studies*.
Ekman, P. E. and Davidson, R. J. (1994). *The nature of emotion: Fundamental questions.* Oxford University Press, USA.
Ely, R. J., Ibarra, H., and Kolb, D. M. (2011). Taking gender into account: Theory and design for women's leadership development programs. *Academy of Management Learning & Education*, 10(3):474–493.
Ferrara, E. and Yang, Z. (2015). Quantifying the effect of sentiment on information diffusion in social media. *PeerJ Computer Science*, 1:e26.
Hogue, M. and Lord, R. G. (2007). A multilevel, complexity theory approach to understanding gender bias in leadership. *The Leadership Quarterly*, 18(4):370–390.
Huang, J., Gates, A. J., Sinatra, R., and Barabási, A.-L. (2020). Historical comparison of gender inequality in scientific careers across countries and disciplines. *Proceedings of the National Academy of Sciences*.
i Cancho, R. F., Solé, R. V., and Köhler, R. (2004). Patterns in syntactic dependency networks. *Physical Review E*, 69(5):051915.
Jansen, B. J., Zhang, M., Sobel, K., and Chowdury, A. (2009). Twitter power: Tweets as electronic word of mouth. *Journal of the American Society for Information Science and Technology*, 60(11):2169–2188.
Karami, A., White, C. N., Ford, K., Swan, S., and Spinel, M. Y. (2020). Unwanted advances in higher education: Uncovering sexual harassment experiences in academia with text mining. *Information Processing & Management*, 57(2):102167.
Kenett, Y. N., Levi, E., Anaki, D., and Faust, M. (2017). The semantic distance task: Quantifying semantic distance with semantic network path length. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 43(9):1470.
Konstantinidis, K., Papadopoulos, S., and Kompatsiaris, Y. (2017). Exploring twitter communication dynamics with evolving community analysis. *PeerJ Computer Science*, 3:e107.
Lane, K. A., Goh, J. X., and Driver-Linn, E. (2012). Implicit science stereotypes mediate the relationship between gender and academic participation. *Sex Roles*, 66(3-4):220–234.
Leyva, L. A. (2017). Unpacking the male superiority myth and masculinization of mathematics at the intersections: A review of research on gender in mathematics education. *Journal for Research in Mathematics Education*, 48(4):397–433.
Li, H., Baucom, B., and Georgiou, P. (2020). Linking emotions to behaviors through deep transfer learning. *PeerJ Computer Science*, 6:e246.
Madsen, S. R. and Andrade, M. S. (2018). Unconscious gender bias: Implications for women's leadership development. *Journal of Leadership Studies*, 12(1):62–67.
Malt, B., Wolff, P., and Wolff, P. M. (2010). *Words and the mind: How words capture human experience*. Oxford University Press, USA.
Miller, G. A. (1998). *WordNet: An electronic lexical database*. MIT Press, USA.
Mohammad, S. M. (2016). Sentiment analysis: Detecting valence, emotions, and other affectual states from text. In *Emotion measurement*, pages 201–237. Elsevier.
Mohammad, S. M. and Turney, P. D. (2013). Crowdsourcing a word–emotion association lexicon. *Computational Intelligence*, 29(3):436–465.
Moss-Racusin, C. A., Dovidio, J. F., Brescoll, V. L., Graham, M. J., and Handelsman, J. (2012). Science faculty's subtle gender biases favor male students. *Proceedings of the National Academy of Sciences*, 109(41):16474–16479.
Nasar, Z., Jaffry, S. W., and Malik, M. K. (2019). Textual keyword extraction and summarization: State-of-the-art. *Information Processing & Management*, 56(6):102088.
Ninio, A. (2014). Syntactic networks, do they contribute valid information on syntactic development in children? *Physics of life reviews*, 11:632–634.
Odic, D. and Wojcik, E. (2019). The publication gender gap in psychology. *American Psychologist*, 75.
Pennington, C. R., Heim, D., Levy, A. R., and Larkin, D. T. (2016). Twenty years of stereotype threat research: A review of psychological mediators. *PloS one*, 11(1).
Pierri, F., Artoni, A., and Ceri, S. (2020). Investigating Italian disinformation spreading on Twitter in the context of 2019 European elections. *PloS one*, 15(1):e0227821.
Pietri, E. S., Johnson, I. R., Ozgumus, E., and Young, A. I. (2018). Maybe she is relatable: Increasing women's awareness of gender bias encourages their identification with women scientists. *Psychology of Women Quarterly*, 42(2):192–219.

Polanyi, L. and Zaenen, A. (2006). Contextual valence shifters. In *Computing attitude and affect in text: Theory and applications*. Springer.
Recchia, G. and Louwerse, M. M. (2015). Reproducing affective norms with lexical co-occurrence statistics: Predicting valence, arousal, and dominance. *The Quarterly Journal of Experimental Psychology*, 68(8):1584–1598.
Rizvi, P. (2018). Authorship attribution for early modern plays using function word adjacency networks: A critical view. *A Quarterly Journal of Short Articles, Notes and Reviews*, pages 1–4.
Rodrigues, E. and Pietrocola, M. (2020). Between social and semantic networks: A case study on classroom complexity. *Education Sciences*, 10(2):30.
Rudkowsky, E., Haselmayer, M., Wastian, M., Jenny, M., Emrich, Š., and Sedlmair, M. (2018). More than bags of words: Sentiment analysis with word embeddings. *Communication Methods and Measures*, 12(2-3):140–157.
Seli, P., Beaty, R. E., Marty-Dugas, J., and Smilek, D. (2019). Depression, anxiety, and stress and the distinction between intentional and unintentional mind wandering. *Psychology of Consciousness: Theory, Research, and Practice*, 6(2):163.
Shapiro, J. R. and Williams, A. M. (2012). The role of stereotype threats in undermining girls' and women's performance and interest in stem fields. *Sex Roles*, 66(3-4):175–183.
Siew, C. S., Wulff, D. U., Beckage, N. M., and Kenett, Y. N. (2019). Cognitive network science: A review of research on cognition through the lens of network representations, processes, and dynamics. *Complexity*, 2019.
Steinke, J. (2017). Adolescent girls' stem identity formation and media images of stem professionals: Considering the influence of contextual cues. *Frontiers in Psychology*, 8:716.
Stella, M. (2019). Modelling early word acquisition through multiplex lexical networks and machine learning. *Big Data and Cognitive Computing*, 3(1):10.
Stella, M. (2020). Forma mentis networks reconstruct how italian high schoolers and international stem experts perceive teachers, students, scientists, and school. *Education Sciences*, 10(1):17.
Stella, M., Beckage, N. M., Brede, M., and De Domenico, M. (2018a). Multiplex model of mental lexicon reveals explosive learning in humans. *Scientific reports*, 8(1):1–11.
Stella, M., De Nigris, S., Aloric, A., and Siew, C. S. (2019). Forma mentis networks quantify crucial differences in stem perception between students and experts. *PloS one*, 14(10).
Stella, M., Ferrara, E., and De Domenico, M. (2018b). Bots increase exposure to negative and inflammatory content in online social systems. *Proceedings of the National Academy of Sciences*, 115(49):12435–12440.
Stella, M. and Zaytseva, A. (2020). Forma mentis networks map how nursing and engineering students enhance their mindsets about innovation and health during professional growth. *PeerJ Computer Science*, 6:e255.
Teso, E., Olmedilla, M., Martínez-Torres, M., and Toral, S. (2018). Application of text mining techniques to the analysis of discourse in ewom communications from a gender perspective. *Technological Forecasting and Social Change*, 129:131–142.
Varol, O. and Uluturk, I. (2020). Journalists on Twitter: Self-branding, audiences, and involvement of bots. *Journal of Computational Social Science*, 3:83–101.
Waqas, A., Salminen, J., Jung, S.-g., Almerekhi, H., and Jansen, B. J. (2019). Mapping online hate: A scientometric analysis on research trends and hotspots in research on online hate. *PloS one*, 14(9).
Warriner, A. B., Kuperman, V., and Brysbaert, M. (2013). Norms of valence, arousal, and dominance for 13,915 english lemmas. *Behavior Research Methods*, 45(4):1191–1207.
Waterloo, S. F., Baumgartner, S. E., Peter, J., and Valkenburg, P. M. (2018). Norms of online expressions of emotion: Comparing Facebook, Twitter, Instagram, and WhatsApp. *New media & Society*, 20(5):1813–1831.
Welles, B. F. and González-Bailón, S. (2020). *The Oxford Handbook of Networked Communication*. Oxford University Press, USA.