Digital Mixed Content Analysis for the Study of Digital Platform Social Data: An Illustration from the Analysis of COVID-19 Risk Perception in the Italian Twittersphere

Gabriella Punziano1, Ciro C. De Falco1, and Domenico Trezza1

Abstract
The explosion of platform social data as digital secondary data, collectable through sophisticated and automatized query systems or algorithms, makes it possible to accumulate huge amounts of dense and miscellaneous data. The challenge for social researchers becomes how to extract meaning and not only trends in a quantitative and in a qualitative manner. Through the application of a digital mixed content analysis design, we present the potentiality of a hybrid digitalized approach to social content applied to a very tricky question: the recognition of risk perception during the first phase of COVID-19 in the Italian Twittersphere. The contribution of our article to mixed methods research consists in the extension of the existing definitions of content analysis as a mixed approach by combining hermeneutic and automated procedures, and by creating a design model with vast application potential, especially when applied to the digital scenario.

Keywords
digital mixed content analysis design, digital platform social data, COVID-19 pandemic, Italy, Twitter

Introduction
As stated by Hamad et al. (2016, p. 2), “in the digital age, social networking sites such as Twitter are increasingly turned to as an information source, as they offer a large amount of digital text and are readily available to multisite apps.” All the content circulating on these social platforms is

1University of Naples Federico II, Napoli, Italy

Corresponding Author:
Gabriella Punziano, Department of Social Sciences, University of Naples Federico II, Vico Monte della Pietà, 1, Naples 80138, Italy.
Email: gabriella.punziano@unina.it
composed not only of text but also images, videos, links, geographical locations, retweets as forms of sharing, and likes and comments as both a mirror of engagement retrieval and of relationships and connection. All these, and many other features as well, constitute the essence of digital platform social data.

The purpose of this article is to show a hybrid, integrated, and mixed experience of analysis on these particularly dense, diversified, and miscellaneous data. This will be accomplished through the construction of a research path that recovers and expands the potential of an approach that has experienced a clear period of oblivion in the years preceding the digital turn, as a paradigmatic change of perspective and analytical object (Lupton, 2014; Marres, 2017). Indeed, our belief is that complex social phenomena that also transit on the net could be investigated with a technique that has found a renewed place on the social research scene, just as Big Data is making its weight felt. The technique we have in mind is Content Analysis. In addition, these complex social phenomena require an epistemological and ontological translation into a multi-comprehensive approach like Mixed Methods research (MM). This means fitting into the debate opened by Hesse-Biber and Johnson (2013), for which

The exponential growth of “big data,” arising from newly emergent user-generated and streaming digital data from networking sites such as Twitter and Facebook, will place pressures on MM researchers to transform traditional modes of collecting and analyzing data generated from these sites. . . . In the coming years, big data methods and analytics may also drive and challenge MM researchers to rethink and innovate and produce new paradigmatic perspectives and research designs and structures. In turn, MM perspectives and praxis can provide models for interpreting and deriving critical insights that may give a more complex understanding of big data that can bring a set of new questions and understanding to the trending data currently extracted from user-generated social networking sites. (Hesse-Biber & Johnson, 2013, p. 107)

This is the reason why new applications, new software, and new algorithms are being developed, allowing the extraction of the knowledge nested into digital data. All the characteristics of content analysis are being recovered with text mining techniques and with a continuous interconnection with network analysis and geographical techniques. This includes both qualitative (Schreier, 2012) and quantitative versions, the latter considered from its dawn (Berelson, 1952) to the present day (Riff et al., 2019). For social researchers, this draws attention to the continuous evolution of the cognitive horizon, which allows access to this new digital frontier of content analysis—a frontier that has led to the breaking down of the boundaries between qualitative and quantitative approaches, as well as among different disciplines, leading to the birth of forced hybridizations. What is happening in social research is essentially a paradigm shift from web content analysis to digital content analysis—similar to the shift that took place from internet studies to digital methods.

It was precisely on the basis of these considerations that we wanted to focus on the analysis of data with an innovative mixed digital model to investigate the recognition of risk perception in the first phase of COVID-19 on Twitter by one of the populations most seriously affected by this catastrophe. We will apply and discuss an innovative model devoted to investigating the multivariate nature of digital platform social data capable of providing different results, both deep and extended: a model that we call digital mixed content analysis design. We will explain all the methodological and analytical assumptions that help to frame the proposed empirical perspective (the section “Theoretical Framework”). Then, the background of the case study will be briefly introduced (the section “The Background of the Case Study to Break Through the Wall of Possibilities Opened by the Application of Mixed Procedures to Digital Content”). The “Methodology” section will explain characteristics of
the design, the construction of the empirical basis, the methods applied, the techniques developed, and the software used, and it will conclude with a brief description of the advantages and limitations of the presented proposal. Following this, the main results of the study (the section “Findings and Results”) will be presented, along with the strengths and weaknesses of automatized procedures, notes on access to the data and packages used, and an assessment of the contribution this study aims to provide to ongoing research on mixed methods. Finally, conclusions will be drawn (the section “Discussion and Conclusions”), leading to a discussion of the theoretical and conceptual implications of the method we used, clarifying its breaking points, the principles of innovation, and the clarity of results derived through this particular mixed procedure.

**Theoretical Framework**

**Content Analysis: Developments and Digital Scenarios**

The term *content analysis* covers nearly every technique used to extract secondary meaning from information. Content analysis allows researchers to recover and examine the nuances of behaviors, perceptions, and trends, by generically applying two types of processes: a conceptual analysis (also known as thematic analysis) and a relational analysis (also known as semantic analysis) (Carley, 1990). Content analysis assumed the status of a research tool in the 1950s after the publication of fundamental texts like those of Lasswell et al. (1949) and Berelson (1952). Multiple definitions have been given for the analysis of content, each of which implies a given perspective. Krippendorff (2018) identifies three types:

1. Definitions that take content to be inherent in a text;
2. Definitions that take content to be a property of the source of a text; and
3. Definitions that take content to emerge in the process of a researcher analyzing a text relative to a context.

In Krippendorff’s general definition, content analysis “is a research technique for making replicable and valid inferences from texts (or other meaningful matter) to the contexts of their use” (Krippendorff, 2018, p. 403).

One of the advantages of content analysis is that it is economical in comparison with other techniques, if the data useful for research are readily available, like in web-based content analysis, particularly since the advent of Web 2.0 advances in information retrieval. Web 2.0 and the data revolution (Kitchin, 2014) have influenced the content analysis approach in three ways.

Firstly, the wide availability of digital content could explain why content analysis is more widely studied and used in scientific research. On the Web of Science, the use of the topic of “content analysis” has grown almost exponentially in the past 30 years (cf. Figure 1).

Secondly, traditional techniques are now being accompanied by nontraditional techniques borrowed from other disciplines. This allows us to face the challenges posed by *old and new* data, bringing about a paradigm shift from web content analysis, in internet studies, to digital content analysis, in digital methods (Herring, 2009).

Finally, the third influence concerns the quantity/quality dichotomy. In Berelson’s (1952) definition, content analysis was defined as a quantitative approach, while the definitions subsequently given by authors such as Krippendorff (2018) and Cole (1988) do not exclude
This technique from the qualitative approach (Hsieh & Shannon, 2005). According to Schreier (2013), if qualitative content analysis is data-driven and focuses on both explicit and latent content, quantitative content analysis, on the other hand, tends to be concept-driven by the hypotheses being tested. It is precisely in the MM, which aims to integrate qualitative and quantitative research within an abductive logic, jointly data- and concept-driven, that content analysis could find fertile ground, especially in its digital version (Snelson, 2016).

There are two further reasons that may have given rise to this revival of content analysis in the digital scenario. First, the crossing of the dichotomy is directly and indirectly supported by new perspectives, such as those of "live sociology" (Back & Puwar, 2012) and of "punk sociology" (Beer, 2014). In practice, this has the advantage of low costs, in terms of both money and time. Secondly, it should be emphasized that, in a phase in which epistemologically naive approaches (such as those that are data-driven) are being asserted—supported not only by the availability of data but also by the possibility of obtaining sophisticated statistical analyses almost automatically—it is important for researchers to affirm their role. Researchers must emphasize not only the importance of theory but also the importance of facing a cognitive problem through complex approaches in order to offer better answers or, as Creswell (1999) put it, to better understand a situation.

**Digital Methods as Evolution in Content Analysis**

Digital methods can be considered as a set of research and strategy approaches using data produced in digital environments to study sociocultural changes (Caliandro & Gandini, 2016; Rogers, 2009). These differ from virtual methods (Hine, 2005), also known as digitized methods (Rogers, 2009), which are paradigms that study reality by adapting social research tools to the web (for example, the online survey). According to Rogers (2009; 2013), using digital methods presupposes epistemological choices. This implies using knowledge about the internet and the context of the web network not only as an ontological
structure but also as a resource method to study people’s behavior and social groups. The new digital environments have increasingly contributed to a blurring of the boundaries with so-called “offline reality.”

The digital environment provides a permanent research context that offers scholars a range of techniques and native tools to measure and interpret social phenomena. Rogers recommends adopting the behavior of a digital native and following the medium to become more familiar with the digital objects available on the net. These may include, by way of example, likes, tags, retweets, shares, and hashtags tools. They may also include native applications that allow one to interact with large streams of net data, such as API (application programming interfaces) strategies and web scraping that help the researchers to follow the thing, the internet, as the fastest growing social scenario.

Digital methods thus represent the natural evolution of the research paths commonly used by web content analysis. Looking at the transposition on the net of the classical tools and methods of web content analysis procedures, the paradigm shift introduced by digital methods requires the scholar to change his or her mind about conceptualized content analysis on the net. Traditional content analysis tools need to be accompanied by innovative strategies for handling a large amount of data, considering the net as both the object and the instrument of analysis. Careful combinations of digital methods and web content analysis can preserve the strengths of traditional content analysis (systematic rigor and contextual sensitivity) while maximizing the large-scale capacity of Big Data and the algorithmic accuracy of computational methods (Kitchin, 2014; Lewis et al., 2013). Thus, the analyzed content can be considered a product of the action in a specific scenario, imbued with the determinants that characterize it. These determinants could also exasperate the use and the process of signification attributed to the affordances typical of the specific platform. The category of significance to classify content in content analysis is now meeting artificial intelligence techniques, and the connection between concepts, topics, and argumentation can now be seen as a powerful means of understanding a chain or named network.

These, as well as many other innovations, change the way of doing content analysis on digital platform social data, but the potential of this digital approach for content analysis does not exhaust its potential only in this paradigmatic shift. In fact, it is in the practice of analysis that many other possibilities open up. One of these is the possibility of fruitfully approaching integrated analysis models typical of MM research.

**The Mixed Approach in Digital Content Analysis**

MM research centers around researchers being able to collect multiple data using different strategies, approaches, and methods. The desired results of this mixture have the characteristic of being more than the simple combining of individual methods in order to generate grander and more integrated research outcomes (Orina et al., 2015). Many fields of research, with their characterizing methods and techniques, have already experienced the potentiality of combining qualitative and quantitative research approaches to pursue the guiding methodological principle of integration. Nowadays, it is not only a question of methodological principle that addresses social researchers, but also the ever-growing relevance of the kind of data used, the information contained therein, the possible multilayers of reality to which they lead, and the undeniable need for integration between these pieces of reality to build ever more complete paths of knowledge.

Using content analysis in the digital era to analyze digital platform social data means being faced with old and new challenges. Digital content analysis researchers must formulate their cognitive questions and make the purposes of their analysis explicit; identify the sources of the
data and contents; and then select them consistently. The analysis procedures that they decide to adopt, whether quantitative or qualitative or both, will depend on the hegemony of the research question (MM perspective), but above all on the hegemony of the medium that conveys the contents taken into analysis (digital methods perspective). Regardless of these considerations, the content analysis process will consist of the coding of raw data according to a classification scheme. This scheme quantitatively claims to extend and generalize the results. Qualitatively, on the other hand, it will attempt to analyze the content more in-depth. However, it is an understatement to think that a cognitive question on complex data, such as digital platform social data, can involve only one of these sides. The MM perspective is not only necessary; it is mandatory.

In this regard, it is sufficient to think that Holsti (1969), Schreier (2012), and Krippendorff (2018) affirmed that qualitative and quantitative content analysis are not discrete classifications, but rather fall along a continuum. This is a notion also used by Teddlie and Tashakkori (2011) to define the new horizon for social research methods in light of the third approach, the mixed one. Stressing the approach along this continuum allows researchers to extract greater opportunities to gain insight into the meaning of data. Bryman (2012) and Hamad et al. (2016) state that, by definition, content analysis is a research approach that can be situated at the intersection of quantitative and qualitative methods, a place where both methods can meet, quantifying and qualifying both the manifest and the latent meanings of the data. Combining this understanding of content analysis with a solid MM design could allow researchers to reach the maximum result from the massive growth of digital texts and multimedia data. This proposal of a digital mixed content analysis model is the key to our argument.

Of course, it is true that, for researchers using data from social media platforms (e.g., Facebook, Twitter, and LinkedIn), there are few guidelines for data collection, analysis, and evaluation. The digital mixed content analysis model that we present in this study should be seen as an applied example of organizing a framework to guide the application of integrated methods (quantitative and qualitative) of content analysis on digital platform social data and to address their varied nature.

The Background of the Case Study to Break Through the Wall of Possibilities Opened by the Application of Mixed Procedures to Digital Content

In December of 2019, an epidemic of atypical pneumonia began to spread in China and was identified as novel coronavirus (COVID-19). In early 2020, COVID-19 cases were increasing in other countries, including Italy. The Italian coronavirus cases surged to thousands within 2 weeks, marking the biggest coronavirus outbreak outside Asia. On March 8 the Italian government announced the lockdown of the 11 most badly affected Italian towns (DPCM, 2020). Within 2 days, the quarantine was extended throughout Italy through the “I Stay at Home” decree (GIPCM, 2020) as COVID-19 cases were detected across the country. Italy was the first country to announce a nationwide lockdown.

In such a critical context, models of crisis and emergency risk communication (Beck, 1992; Renn, 1992) suggest that it is crucial to understand the risk perception of the population, along with the sources of information that they trust to enable effective communication. Although international and national institutional actors attempted to plan communication strategies for the correct information, there was a high risk of fake news, overflow, and bad information, especially on the shared main social networks (Vaezi &
Rumors and misinformation can undermine many public health actions and should be debunked effectively.

In our case, the relevant statement is that information, common sense, and technicalities—as well as misinformation—contributed to polarizing Italian users’ perceptions of the emergency, from excessive fear and concern to a total lack of interest. Therefore, it is interesting to construct the main semantic categories of the perception and representation of the disease, and to understand which dimensions are related to principal information bias and to news overload. In this way, it will also be possible to consider any relationship between the epidemic outbreak and the change in people’s risk perception and feelings in order to try to improve institutional communication and safety-oriented policies.

**Methodology**

**A Variant of Exploratory Sequential Digital Mixed Content Analysis Design**

The research design at the basis of this proposal for a digital mixed content analysis can be identified in a variant of the exploratory sequential model put forward by Creswell and Plano Clark (2017). This model generally starts with an exploratory qualitative phase to adequately inform the second quantitative phase by specifying the research questions and the variables that will guide it. In the study presented here, we envisage the addition of a preliminary quantitative phase (topic modeling) that leads to better structuring the starting data set. The aim is to shed light on the lack of theoretical and/or empirical knowledge of a phenomenon, so much so that in this work the model is adopted for the development of taxonomies of COVID-19 risk perception in the Italian Twittersphere. In this, the qualitative phase is necessary to develop an emerging theory built on thematic classifications, which will then be tested quantitatively through latency and cluster analysis.

The primary research question that inspired the need for us to use a mixed methods model to better fit with digital platform social data is, “How did the spread of coronavirus direct, polarize, and construct the risk perception of Italian Twitter users?” Starting from this, other more specific questions guided the different steps of the research design. The first question, “Which digital content can detect insights on risk perception and how can we select them?” led us to start our research with a more quantitative phase of analysis that can be divided into two more specific steps. This, in turn, allows the funneling of two other phases, one qualitative and one quantitative, developed sequentially.

Proceeding in stages, the two steps that constitute the first quantitative phase consisted first in constructing the empirical basis of the study. This was done through an automatic hashtag extraction with R on Twitter, which allowed us to develop a first stage of sampling that would be devoted to outlining the eligibility characteristics of the contents to be included in our sample.

The second step then consisted of applying an automatic extraction of meaning from the selected set of data with the application of topic modeling. For this, we used T-Lab software that allows the implementation of a second stage of sampling. At this second stage, we were able to extract the most representative set of tweets for each identified topic. Once the empirical basis was built, in order to answer the more general research question, a phase of qualitative analysis was undertaken. This allowed us to examine and provide answers to more specific questions: “What was the rationale that built the social narrative of coronavirus on Twitter?” and “Who were the relevant actors in managing the most pervasive forms of communication on this social network?”
A third step of analysis arose from these questions and was conducted through a hermeneutic thematic analysis with NVivo. This analysis was focused on the hermeneutic interpretation of each set of tweets, classified by the theme that had been identified through the application of topic modeling. The purpose of this step was to detect new information about the way in which the main differences in communication (methods of treatment, communication styles, polarity, feeling, intensity, and direction of the perceptions) could be distinguished, and then to build the classification system of the contents (the constitutive themes) jointly into the creation of new attributes (the narratives that dominate the scene on Twitter). In order to use the wealth of information structured in this way and to answer the question, “What are the constitutive dimensions to differentiate various risk perception profiles?” A further quantitative step developed on T-Lab was necessary. This step was based on a specific objective: dimensions reduction.

In light of this, two further analytical steps were built that would allow, on the one hand, the reduction of the semantic dimensions to a few manageable constructs (with the application of lexical correspondence analysis, or LCA), and, on the other hand, the reduction of the groupings of emerging risk perception profiles (through the application of cluster analysis, or CA). The integration of the results of all these steps was reached with a concurrent phase of analysis that was aimed at building a classification scheme from the integration of the different quantitative and qualitative results (cfr. Figure 2). In particular, this derives from the two LCA axes (or latent dimensions), from the hermeneutic thematic analysis, and from the CA. This leads to the definition of the attributes for delineating the different emerging risk profiles and, finally, the projection of the COVID-19 emerging risk perception profiles of Italians on Twitter, using the scheme that takes shape with these elements.
The Construction of the Empirical Base: Which Digital Content Can Detect Insights on Risk Perception and How Can We Select Them?

Firstly, we conducted an automatic hashtag extraction on Twitter, supported by R² extract tweet packages that interface with APIs (application programming interfaces) to collect data. The data collection involved all the tweets about COVID-19 in Italian from March 5–15, when several important decisions relating to COVID-19 containment were made. The Italian government first announced a partial lockdown on March 4 (DPCM, 2020); then, after extending it nationwide, the government issued the I stay at home decree (GIPCM, 2020). It should be clarified that Twitter is our research context. Since January 2020 (Global Digital Report, 2020), it has been among the most visited (with an average 10 million visits per day) and most popular of websites (with about 339 million active accounts). In Italy, it is less widely used, but still enjoys fair popularity (6% of Italian netizens have a Twitter account, which is comparable to the world average). Despite its popularity, it would be not correct to say that it is representative of the world population—or worse, of the Italian population. This aspect is part of the already widely documented problems regarding research on the web, and specifically with the Big Data of social networks (Keyes & Westreich, 2019). Therefore, it is not possible to generalize the results to the whole population. The results should be limited to the community of Italians present on Twitter in the period highlighted in our research.

The clarification of the scope as being limited to these specific users is not the only challenge facing this team. In fact, this issue hides other pitfalls relating to the APIs’ environment and data extraction procedures. The APIs are a set of procedures that interface with an application in performing a specific task (extracting Twitter posts, for example). Tweets are public, so there are no privacy constraints. However, we used the basic free version of the APIs (v1.1).³ We are aware of the risks that automated extraction and an uncritical approach bring to the building of large databases (Hernandez-Suarez et al., 2018; Leetaru, 2019). For this reason, we have identified four limitations and tried to solve them: the completeness, the timing, the daily number of extractions, and the limits for each call. The standard search API is focused on relevance and not on completeness. This means that some tweets, and some users, may be missing from the search results. However, the exclusion rate is very small, especially considering the large extent of the corpus. Secondly, the search index has a 7-day limit. In other words, no tweets will be found for a date older than 1 week. For this reason, we adopted a real-time extraction. Thirdly, it is not possible to extract more than 18,000 tweets per call. We used the “retryonratelimit” parameter to establish the extraction point from which to restart. The fourth limitation is that it allows you to extract a maximum of 100,000 tweets per day; however, the daily number of our corpus was significantly lower, so the limitations of the standard APIs only marginally affected the data extraction process. Rtweet⁴ is the R package we used to interface with the API and extract the data. The great advantage of Rtweet is that it allows the interface with the API while allowing us to customize the extraction (for example, choosing the number of tweets, excluding retweets, dividing the extraction by days, or for a longer period of time). The function used is the “search_tweets” function, into which we entered the extraction keywords, the language, the time interval, and the option that excludes retweets.

Going back to the extraction keys (cfr. Figure 3), these were based on six hashtags—that is, those that were potential or effective topic trends for the period in question:
• #coronavirusitalia (coronavirus Italy) and #coronavirus identify the main theme and index a more popular and generalist communication (defined as knowledge-oriented);

• #iorestoacasa (I stay at home), #fermiamoloinsieme (let’s stop it together), and #italiazonapotetta (Italy protected zone): communication more interested in problem solving, that is, about measures for risk reduction (called problem solving-oriented).

The final corpus consists of 2,145,048 tweets (including retweets). Of these, the largest share is represented by tweets defined, a priori, as knowledge-oriented (just over 77%). To facilitate a mixed design, we decided to work on a more limited sample of 10,000 tweets (without retweets), respecting the proportions related to:

1. Daily number of tweets; and
2. Hashtag groups.

The daily tweet percentages suggest that, from the first day of extraction up through March 11, there was a progressive increase in COVID-related tweets. The most active days were those from March 8 through 11 (with an average of more than 10% of the daily tweets). The high number of
tweets is plausibly connected to the implementation of important lockdown orders in Italy, first in the North and then throughout the country. March 11 (after Italy’s lockdown) was, in fact, the day with the most tweets extracted (just over 13% of the entire body). However, there was a slightly decreasing trend after that date.

As for the two hashtag themes, it is evident that, for the first 3 days, there was a clear imbalance in favor of the tweets in the knowledge-oriented group (almost 100% of the tweets extracted have hashtags #coronavirus and #coronavirusitalia). By contrast, on the following days (starting from March 8) the distribution was more balanced. In fact, the tweets of the problem solving-oriented group, together with the implementation of the first lockdown rules, began to circulate more frequently, reaching more than 39% on March 9, and then falling again in the last few days (18.3% on March 14 and 14.1% on March 15) (Table 1).

The 10,000-tweet data set was enriched by automatic extraction of other variables: screen name and account verified (useful to define users’ classifications and to check official accounts, such as media, opinion leaders, political organizations, etc.); date; hour (morning, afternoon, evening, or night); source (mobile or fixed position); text width (short, medium, or long); favorite count (recoded as proxy of engagement level); and retweet count (recoded as proxy of sharing level). This first operation of sampling and organization of the database served to outline characteristics of eligibility and identification of the digital content on which to develop the next steps of the proposed research design.

Findings and Results

The Rationale that Built the Social Narrative of Coronavirus on Twitter and the Relevant Actors in Managing the Most Pervasive Forms of Communication

The first objective of our analysis consists in extracting new information about how to distinguish the main differences in the points made by users and the emerging themes that are detectable from the set of analyzed tweets. Also, much relevance is given to tracing differences about the methods of the treatment, communication styles, polarity, feeling, intensity, and direction of the perceptions expressed. The application of topic modeling to develop a subsequently in-depth hermeneutic thematic analysis, using the typical mixed methods structure, allows us to identify the constitutive themes and narratives that dominate the scene on Twitter relevant to Coronavirus.

The topic modeling technique provides an automatic classification with a quantitative aim of the analyzed contents. Still, in our application, we provided the chance to refine the emerging thematic areas by also treating the extracted themes with hermeneutic thematic analysis of the elementary contexts in an MM manner that uses the qualitative perspective to generate mining from quantitative operations. First, topic models are algorithms for discovering the main themes that pervade a large and otherwise unstructured collection of documents. Topic models can organize the collection according to the discovered themes (Blei et al., 2003). The algorithm applied by the software used, T-LAB, is the Latent Dirichlet Allocation (LDA): a bottom-up Bayesian procedure for probabilistic topic modeling (Zeng, 2012). The basic idea is that the documents are represented as random mixtures over latent topics, where a topic is characterized by a distribution over words. Each word in a corpus is assumed to have been generated by a latent topic, drawn from a document-specific distribution over the totality of identified topics (Jelodar et al., 2019). Our 10,000-tweet data set was subjected to a pre-treatment (with normalization, lexicalization, lemmatization, elimination of stop words, and segmentation) conducted with a supervised automated process on the same software.

Subsequently, from the corpus obtained, 20 thematic groups were extracted with this automatic process of topic modeling. For each theme, the list of specific words most probably shared with
Table 1. Tweet Population and Sample (AV and %) by Day and Hashtag Groups.

| Day    | n tweets | %  | #coronavirus (coronavirusitaly) | #iorestoacasa (istayathome) | #italiazonaprotetta (italyprotectedzone) | #fermamoloinsieme (let'sstopittogheter) | #coronavirusitaly | #italyprotectedzone | #let'sstopittogheter | n tweets |
|--------|----------|----|---------------------------------|-----------------------------|------------------------------------------|------------------------------------------|-----------------|-------------------|---------------------|----------|
| March 5 | 61,693   | 2.9| 100                             |                             |                                          |                                          | 0.0%            |                   | 288                 | 288      |
| March 6 | 140,002  | 6.5| 100                             |                             |                                          |                                          | 0.0%            |                   | 653                 | 653      |
| March 7 | 198,527  | 9.3| 99.8                            |                             |                                          |                                          | 0.2%            |                   | 924                 | 926      |
| March 8 | 234,441  | 10.9| 66.8                           |                             |                                          |                                          | 33.2%           |                   | 730                 | 1093     |
| March 9 | 281,869  | 13.1| 60.6                           |                             |                                          |                                          | 39.4%           |                   | 796                 | 1314     |
| March 10| 262,421  | 12.2| 65.1                           |                             |                                          |                                          | 34.9%           |                   | 797                 | 1223     |
| March 11| 284,753  | 13.3| 81.8                           |                             |                                          |                                          | 18.2%           |                   | 1085                | 1327     |
| March 12| 141,442  | 6.6 | 69.1                           |                             |                                          |                                          | 30.9%           |                   | 455                 | 659      |
| March 13| 143,571  | 6.7 | 69.2                           |                             |                                          |                                          | 30.8%           |                   | 463                 | 669      |
| March 14| 206,125  | 9.6 | 81.7                           |                             |                                          |                                          | 18.3%           |                   | 785                 | 961      |
| March 15| 190,204  | 8.9 | 85.9                           |                             |                                          |                                          | 14.1%           |                   | 762                 | 887      |
| Tot.    | 2,145,048| 100| 77.4                           |                             |                                          |                                          | 22.6%           |                   | 7738                | 10,000   |

Source: Our elaboration on R output.
other themes, along with the words shared with less probability, was evaluated, together with the assessment of the elementary contexts extracted. In this case, the most representative set was a maximum of 100 tweets that characterized each specific theme. This has been assumed as the second sampling procedure, representing the funnel neck of the transition from the automatic-quantitative to the manual-qualitative perspective. The software automatically returns the set of 100 tweets for each theme organized according to this thematic division. In order to give the appropriate sense to these groupings, however, the in-depth hermeneutic thematic analysis for the interpretation of the topics with NVivo software was developed on this second selection. This also assigns characterizations on the types of communication, styles, polarity, sensations, and directions of the perceptions expressed by Twitter users. These features can be used as elements for the classification system applied to integrating the different sets of results. This means creating new attributes that correspond to the identification of the constitutive themes and narratives about COVID-19 risk perception that dominate the scene on Twitter.

The 20 thematic groups extracted by topic modeling and analyzed with the hermeneutic approach can be described as follows in Table 2. Naming them, but also characterizing the content with the element of social narratives (main actors, type of communication, style of language, polarity of sentiments, and risk profiles) builds the main differences among the themes.

Other additional information could be detected from our data set to enrich our interpretation. Looking at Figure 4, we can see a clear shift. In the early days, there was a concentration on policy and medical issues, with the speeches of scientists, technicians, experts, institutions, and some politicians and administration members all focused on the health emergency; in this early stage, the emphasis was on the public/collective sphere. Later, this shifted to a concentration on all the consequences of the pandemic for work, the economy, and the healthcare system, with other politicians and ordinary people focused on the socioeconomic emergency; this represented a shift toward emphasizing the centrality of the private/individual sphere. Typical trends of the risk management phases ranged from the initial information phase towards containment and reconstruction measures; but, as a result of the pandemic, these phases were extremely accelerated.

Regarding engagement and sharing levels, the most retweeted and liked posts concerned medical issues and public communication about the health emergency, showing more retweets about technical information. The most widely shared posts, in addition to medical issues and public communications, concerned, on the one hand, reflections and comparisons with the rest of the world, showing a certain openness, and on the other hand, issues involving local and individual everyday things.

Therefore, we also looked at differences in the building of social narratives. This analysis considered the types of actors involved, communication type and style, sentiment polarities, intensity, direction of the perceptions that were expressed, and the kinds of risk profile associated with the combination of particular thematic areas and their specific features. All this information was reorganized and presented in the following concept map. This map shows the initial way that the categorization scheme was reorganized and developed to encode tweets belonging to the 20 thematic areas in which we used our conceptualizations as a system for classifying the analyzed corpus, as shown in Figure 5.

What are the Constitutive Dimensions Differentiating Various Risk Perception Profiles? A Multidimensional Analysis with LCA and CA

Up to now, we have seen how the themes we have reconstructed starting from an analysis of our corpus assume importance and characterization. Now, we will tackle their relationships and move one step further in our research, focusing on a digital mixed content analysis design. Specifically,
| Topic | Description |
|-------|-------------|
| 1 Daily things (19.4%) | This includes tweets related to different topics, which all refer to daily news or events: clarifications, rule-abiding, staying indoors, behaviors to follow, unjustified outings, or closure of businesses—these are only some of the central nodes. This thematic area belongs more to ordinary people. The type of communication is mostly interpersonal; the style of language is simple and poorly researched, and the polarity of sentiments ranges from sadness and negativity to a call for responsibility in everyday actions, bringing the perception profile of the risk that emerges to be cautiously alarmist and focused on contingent activities. |
| 2 Addresses by politicians and other major actors (13.4%) | This area focuses on issues that should not be underestimated, such as: the need to avoid the spread of the virus and all that at this moment can be damaging, such as controversy. To be called into question as main actors are institutional actors such as the council of Ministers, banks, and industrial associations. The type of communication becomes more political; the style is more refined but emotionally moving, seeking high involvement by shifting the profile of the perception of the risk to be defined as active involvement due to the weight of the situation experienced. |
| 3 Institutional and digital communication (1.8%) | This group leverages the speeches of the actors with authority on the institutional and cultural scene and their transposition into digital. This presents communication full of meanings and information, with a precise and yet personal style, where the polarity of sentiment results in a neutral and relaxing feeling that develops risk perception profile aimed at greater awareness and rationality. |
| 4 Daily limitations (2.8%) | In this thematic area, we find tweets referring both to the limits that the quarantine imposes and respect for the situation. It focuses on the need to understand what is happening and to respect the restrictions that the situation imposes, not to be selfish, and therefore to avoid gatherings and public places such as bars. The main actors of the discourse are ordinary people. The communicative style is mostly informal and direct; the polarity of feeling remains neutral and leads to the development of a risk perception profile based on understanding and respect for the rules imposed. |
| 5 National measures (3%) | Here we find the tweets that refer to the actions put into place against the spread of Coronavirus and the effects of the pandemic. This focuses on those responsible for these measures, and then the political class and those responsible for spreading information, including Borrelli, the head of Civil Protection, and members of the government. It is stressed that the purpose of these measures is to protect the citizens’ health. This type of public and institutional communication once again calls for an emotional component and a collective responsibility, using a simple and direct but technical style, in which the risk perception profile becomes fully informed and adequately educated in the current scenario. |

(continued)
| Topic                                      | Description                                                                                                                                                                                                 |
|-------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 6  New and old habits (1.9%)              | This theme collects the content of tweets that center around the rediscovery of old, forgotten practices, such as reading, listening to music, and watching TV. Paper tissues and disinfectants are elements that are part of the new daily life, together with the most diversified curses towards those who do not respect the restrictions. Communication becomes *interpersonal*, even though the addressees brought up for discussion are often *institutional* and political actors, joined with *ordinary people*; the style is decidedly *informal*, at times *recriminatory*, but at the same time *proactive*, dragging a polarity of expression of feeling which, on one hand, is more *negative*, and on the other, more relaxing; this builds a rather ambivalent risk perception profile. |
| 7  Common sense (0.7%)                    | This theme gathers common perplexities to which the answers that are provided are the common sense indications relating to the attitude of “wait for everything to pass,” as far as highlighting the shortcomings of the actors who should be responsible for solving the problems. Communication still remains interpersonal, the style is very simple, the polarity oscillates between the negative and the yielding, and the perception profile of the risk that derives from it is decidedly passive. |
| 8  Informed opinion (1.6%)                | This thematic class gathers indications to informed opinions of both health professionals (doctors, operators, nurses, etc.) and information professionals, as well as to more informed citizens. Central issues are the arrival of the Chinese medical teams for support and the issues related to the development of research on the topic. Public and scientific communication involving explanation is more widely present and has a rational and informative communication style, neutral polarity, and a perception profile of the risk attached to the construction of knowledge in the various spheres that are involved. |
| 9  Losses and dangers (4.2%)              | This theme collects very homogeneous tweets from ordinary people whose concern, first of all, is the danger of dying and the loss of loved ones, especially of the elderly or of parents, recognizing themselves as possible vehicles of contagion and, therefore, potentially guilty. Once again, it is a type of interpersonal communication, with a common, simple, but emotional style, which reveals the painful parts of this epidemic and highlights a component of fear in the polarization of a mostly negative feeling, where the perception profile of the painful parts of this epidemic and highlights a component of fear in the polarization of a mostly negative feeling, where the perception profile of the risk is that of terror for the unknown, for what is not yet known and what could happen. |
| Topic                                                                 | Description                                                                                                                                                                                                 |
|----------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 10 Nationalist sentiment (0.4%)                                      | This thematic group includes the reference to national values and a feeling of closeness to one’s homeland and to fellow citizens in a sort of shared-hope prerogative of politicians. The communication it includes is public and political, but also interpersonal; the very emotional style leverages an irrational component that polarizes sentiments on a positive side, which leads the risk perception profile to become moderately confident in the future. |
| 11 Public communication on health emergency (4.2%)                  | This is a thematic class that refers to China, America, or Europe, but also Italy and the key concepts that describe the health emergency, such as bulletins, progress, infections, victims and recoveries, and the prerogative of experts. These are clear indications for daily updates, conveying institutional communication of a public nature that focuses on the most problematic issues of the emergency, which also involve political management. The style is technical, the polarity deliberately not alarmist nor negative, and the emerging risk perception profile is that of an alerted actor. |
| 12 Economic and health concerns and hopes (0.2%)                    | In this thematic group, on the one hand, concerns about economic impact are revealed and, on the other, health concerns, highlighting concerns as well as hopes in the ability to produce effective containment solutions to a dramatic situation, the prerogative of politicians. The mixed interpersonal and public communication of this group has style that tends to generate awareness and hope together, trying to shift the polarity of sentiment towards the positive side, and to constitute a risk perception profile in which the awareness that it is possible to come out of this situation limiting the damage prevails. |
| 13 Measures taken for working, smart working, and income (12.8%)    | This theme collects tweets about how the quarantine has changed the habits of those who no longer work from the office but from home in smart working mode. However, not all of them are employees, so for the rest of the subjects, the issues relating to the guarantee of income and the measures that will be implemented for this purpose prevails. This mixture of interpersonal and institutional communication is united by a common and determined style at the same time, the polarity is mixed, and the risk perception profile is focused on future impact rather than on the contingent situation limiting the damage prevails. |
| 14 Quarantine, prayers, and recriminations (9.6%)                   | This thematic area is built around the well-known hashtag #iroestoaCasa (I’m staying at home), especially if you are in a municipal area declared as #zonarossa (red zone), where it is necessary to adopt stringent new habits and respect the prohibitions imposed on ordinary people. In love relationship this turns into an ironic “love in the time of Coronavirus,” but every relationship is affected by the pandemic, including the one with God, and everything is to be reinvented. The prevailing communication is interpersonal, the style is deep and intimate, and polarity uses criticism to become proactive, providing a decidedly constructive risk perception profile. |
| Topic                                    | Description                                                                                                                                                                                                 |
|------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 15 Medical issues (10%)                  | The tweets of this group refer to those affected by the virus and the risks of transmission, including the collapse of medical facilities and the importance of making everyone aware of this emergency. It is a specialized communication in its language and technical style, where, rather than the emotional involvement of the general public, it seeks the involvement of all the actors, from the ordinary citizen to the political decision-maker. The neutral polarity pushes this risk perception profile to be the most markedly technical and specialized one. |
| 16 National resilience (1.8%)            | All the tweets that appeal to the national community fall within this thematic core. In addition to terms like “Italians” and “countries,” we find tweets that use the pronoun “we” or launch appeals on the importance of being together, even while respecting social isolation measures. These users tweet hope that the country will overcome this moment soon and that normalcy will return by rewarding the efforts sustained daily with long bursts of applause from balconies, songs, and burning candles. In this thematic area you can see all the desire for rebirth and reconstruction, with an interpersonal communication, an intimate but collective style, an extremely positive polarization that brings the risk perception profile that emerges from this group to be the most positive one and the one which is most projected towards returning to normal. |
| 17 Reflections and comparisons (9.6%)    | In this thematic class we find reflections on the daily events, with events, with the use of the words “today” and speaking of “dead,” reiterating the information spread by the Istituto Superiore di Sanità (National Health Institute), and highlighting comparisons with what is happening on other countries, including Germany, which had an abnormal number of infections and deaths. It has a communication that is still public, a sensitizing style that seeks answers that are not yet completely clear. The polarity is mixed and the risk perception profile that derives from it is that of seeking answers, of a matured interest in wanting to understand and hot participate as a passive Spector in what is happening. |
| 18 Civic sense and information (1.4%)    | The sense of collective responsibility is enclosed in this thematic group, which refers to the need to disinfect with bleach, pay attention to the documents required to leave the house, and comply with rules, prohibitions, limitations, and daily news updates by ordinary people. Communication is interpersonal, the style intentionally aimed at conscious involvement; it is exhortative and sanctioning. The polarity of feeling is neutral, and the risk perception profile that emerges is moratorium but proactive. |
Table 2. (continued)

| Topic | Description |
|-------|-------------|
| 19    | Health service support (0.4%) Many tweets grouped in this theme refer to a main issue: Helping hospitals, supporting the efforts of medical staff, making sure that the sacrifices made by them are not thwarted, and that the fight against the pandemic becomes something where everyone can do their part. Public, political, institutional, and interpersonal communication intersect in this theme, which finds everyone’s sensitivity and a high emotional involvement, due more to the delicacy of the issue than to the style used to communicate it. Embedded in the ordinary people discourses, the polarity of feeling expressed is mixed, and the resulting risk perception profile is concentrated on the processes of solidarity and mutual support. |
| 20    | Epicenter of the pandemic (0.8%) In this thematic category we see references to Lombardy, to Veneto, to the fears and perplexities that evolve in the first red zone in Italy on the ordinary people living there. Sadness, distrust, and the inability of citizens to see in the long term, are united in this class, with institutional communications that are not always clear, and the resulting risk perception profile is that of citizens drifting at the mercy of events. |

![Figure 4](image.png)

Source: Our elaboration on T-Lab output.

Figure 4. Histogram showing day and topics categorized according to the main argument and percentages in the daily distribution. Source: Our elaboration on T-Lab output.

We will develop a multidimensional reduction quantitative analysis, which was implemented for our data set with the application of Lexical Correspondence Analysis (LCA) (Benzécri, 1973; Lebart et al., 1998) and a Cluster Analysis (CA) (Lebart, 1994). These two techniques were used to reduce the space of mining and to extract the relevant theoretical dimensions, in order to construct
our typological framework for the narratives and perceptions of risk that are contained in our large sets of digital platform social data.  

Like all factorial analysis techniques, LCA aims to extract new variables from the original matrix in order to summarize the information. To understand which patterns represent the extracted factors, it is necessary to understand which modalities of the variables/lemmas are enriched by mining. Having identified the concepts that account for the variability reproduced by the factors, it is possible to proceed with the more general interpretation of the data. The summary of the results of the LCA is achieved by performing the CA simultaneously, or on the new extracted variables. This technique regroups homogeneous elements within a set of data. In our case, CA will group tweets characterized by a similar risk perception expressed in the words used.  

The first result obtained with the application of the LCA is the delineation of two main synthetic dimensions of mining, called factors. These factors can be crossed and used to build a new space of mining that is generated by this crossing. Figure 6 shows the crossing of these new dimensions, the meaning of which is built into the attraction and repulsion relationships among the themes, along with the active variables used for this analysis (e.g., types of users who post the tweets, days on which they post, and time slots).  

The first factor is related to the private-public sphere. On the semi-positive axis, tweets are mainly connected to the individual and private sphere (like daily things), and to individual concerns for the future (like economic and health concerns, hopes, and national resilience). On the
semi-negative axis are issues of a public nature (such as the national measures, or the public debate on the reflections and comparisons concerning the pandemic and informed opinion). The location of user types is decisive. The common user addresses the private sphere, while all other users—and, in particular, political groups and official and administrative bodies—address the public one. To better understand the mining of this factor, we can find on the semi-positive axis lemmas such as aperitif, Netflix, home, and boring, which precisely describe individual experience; on the semi-negative axis, meanwhile, we find the terms health, companies, and OMS, which refer to the public sphere.

The second dimension sees issues such as daily limitations, medical issues, measures for working, smart working, and income on the semi-negative axis, while issues like health service support and the institutional, public, and digital communication about the health emergency appear on the semi-positive axis. The issues on the semi-negative axis seem to refer to the many areas affected by the pandemic and, therefore, to the socioeconomic emergency; those on the semi-positive axis seem to focus only on the health emergency. Health and social concerns, therefore, are the semantic poles of the second factor, which is related to the type of emergency. In particular, on the side of the health emergency, we find terms like containment of the Coronavirus, order, civil protection, and measure, while on the side of the socioeconomic emergency, we find terms like responsibility, awareness, and running away.

These reflections lead us to this question: What are the emerging perceptions and the main narratives regarding the Coronavirus emergency in the Italian Twittersphere? We will try to answer this question by combining the evidence with the results of the cluster analysis shown in Figure 7. There are five groups extracted from the cluster, each characterized by a specific perception of the pandemic that derives from the collectively constructed narration by Twitter users in the first 10 days of the national lockdown.
The first cluster is located near the center of the plane, collecting a very high part of the variability of the opinions expressed, but precisely for this reason also having more in common. It is no coincidence that the characteristic type of user is ordinary people who focus their narrative on losses and dangers, new and old habits, common sense, and experiences of national resilience. While being used as a daily expedient to manage the individual quarantine, they also open to the sense of collective experience, for which the motto “physically distant but close in experience and hopes” holds true. Thus, they also recover the guidelines of politicians and other great actors who tend to want to give off an aura of relaxation in the general experience. The name that can be attributed to this group is that of perception in tension between the most intimate and individual dimension and openness to collective experience, between near and distant worries, between happy scenarios and apocalyptic pictures.

In the second group, which is at the crossroads between a dimension tending to collective-public openness and focused on the health emergency, we find the local and national political-administrative class, the official information, and the top users thus defined for their wide following. The tweets here are those with the highest resonance, and they are mostly centered around a popular narrative that is based on information about national measures in response to the health emergency, an institutional and public communication that passes through digital media. It is based on reflections about the national experience and comparisons with the rest of the world, on economic and social concerns and hopes, and on making reference to the civic sense and the need to be informed. This is a complex narrative that touches various key points of this pandemic precisely because it is the prerogative of the most influential users with afternoon messages that coincide with the circulation of daily update bulletins. The emerging type can be defined as holistic perception.

The third group is harder to define. It reveals a dimension of support for the healthcare system, probably perceived as a necessity by those who have closely experienced losses and greater limitations. The reconstructed narrative is based on informed opinions about the emergency, experienced from a healthcare point of view, along with a more individual concern that weighs not insignificantly. The high information content of these tweets is also motivated by the fact that they are mainly from more engaged and shared users who are, therefore, able
to act on the construction of individual perception, starting from the conscious restructuring of the pandemic narration. The result is a rationalist and consciously alarmist perception.

The fourth group is the one in which a strongly self-centered perception prevails and is, in fact, moved to the more private and individual side of the first constructed dimension. Here we find the tweets that lead back to the effects on the private sphere of the pandemic. The type of user close to this group is once again the ordinary person focusing on everyday things like quarantine, prayer, and recrimination, but also on the new smart working experience, or on the call for relief income for those who cannot work remotely but have to face daily expenses. These are mostly tweeted in the evening and at night, revealing a search for greater intimacy even in a digital dimension of communication and interpersonal sharing.

The fifth cluster, mostly focused on more general issues, and not exclusively medical emergency issues, is characterized by average engagement and sharing users, but also by ordinary users who base their narrative rhetoric on national sentiment, the reference to the pandemic epicenter, and even on very technical medical issues and the regulatory limitations that were imposed on daily life. Mainly, these people tweeted in the morning as they processed, metabolized, and condensed the updates released the previous day, with the expectations and new ideas for pandemic management that came with the new day. The result is a proactive soothing perception in risk management.

Discussion and Conclusions

Strengths of the Mixed Strategy and Major Limitations in the Use of Automated Procedures, Access Issues, and Software Used

The analytical and scraping tools we used represent a real strength in achieving the objectives of our research. However, it is necessary to be aware of the limits reached by the digital approach, especially with the use of digital platform social data analyzed with the application of a mixed methods strategy on this particular content. We can identify four of them:

- **Representativeness of results**: As noted, both due to the percentage of Italian Twitter users and the extraction limits of the R package, the results are not generalizable to the entire Italian population.
- **Assumptions of the algorithms**: Most of the techniques of content analysis incorporate some assumptions about the characteristics of the language (e.g., meaning not contextualized or categorized without a perfect match).
- **Algorithms work**: While the functions of the research software can be easily learned, the criteria of the algorithms are not completely accessible (e.g., extraction work, modeling, R package process, etc.). This could affect research and outcomes.
- **Information loss**: A final aspect concerns the loss of information due to the parallel use of different procedures. However, this loss could be considered acceptable, compared to what has been gained in cognitive terms and through the integration rise.

On the other side, the R package has extracted a lot of valuable information for an extensive exploration of the semantic field related to the analyzed phenomenon. And, it should be emphasized, this information was extracted at no cost. This is certainly a strong first point for social researchers, especially when they have no funding.

Another strong point concerns the combination of the ability to synthesize the information through quantitative techniques, on the one hand, with the opportunity of more carefully exploring the data in the qualitative phase, on the other. This represents the core of our digital mixed content analysis...
research design: the power of automated topic modeling as a first step towards a manual application of hermeneutic sense generation with intense human labor. This combined procedure, together with the subsequent dimensions-reduction steps, helps to achieve a dense system of categorization of digital content while not losing the richness of quality details, but instead refining the results achieved ever more precisely. The integration of these outputs has obtained relevant results that go beyond the limitations, producing the classification scheme that will later be discussed.

Integrating the Result into the Classification Scheme: How the Digital Mixed Content Analysis Design Contributes to the General Development of Mixed Methods Methodological Debate

At this point of the study, we are reaching the moment in which integrations take place, allowing us to answer the main question: “How did the spread of Coronavirus direct, polarize, and construct the risk perception of Italian Twitter users?” The different narratives, styles, thematic classes, and emerging perceptions have allowed us to understand that, however relevant the types of actors who post on Twitter are, the pervasive effect of their communication is not so much in what is followed but in the diversity of ways in which they deal with and construct specific narratives on the pandemic. And it must be noted that this does not make one actor more relevant than the other in absolute terms, but is relevant in relation to the thematic sphere and the way in which it deals with and disseminates it to the rest of Twitter users.

Now it is possible to summarize the results developed sequentially with our digital mixed content analysis research design in an integrated general model of classification, shown in Figure 8. All the extracted dimensions of synthesis will be used to create the typological axes of a

Figure 8. Digital mixed content analysis integrated general model of classification for COVID-19 risk perception in the Italian Twittersphere. Source: Our elaboration.
theoretical framework, within which the types of perceptions and emerging narratives will be positioned systematizing the main actors that cover the related rhetoric.

The reconstructed quadrants show a cross between the two axes extracted with the application of the LCA. The first opposes a narrative focused on the health emergency to a narrative of the effects on the emergency seen from a socioeconomic point of view, recovering in addition medical, psychological, and other related, non-negligible pandemic issues. The second axis opposes an orientation toward the private and individual sphere, albeit including the sense of community that bonds individuals isolated physically but with a community of perspective and current experience, with an orientation toward the public and collective sphere, which loses the connotation of community and addresses the society/nation understood as more of an administrative entity than one of intent.

In the intersection of these dimensions, we find a first rationalist narrative focused on the health emergency. This is the preserve of experts from the most diverse sectors, which builds and directs a consciously alarmist perception. The second narrative is intimate and emotional, focused on the more general socioeconomic emergency. This is the prerogative of ordinary people, and it leads to the construction of a self-centered perception. The third narrative is the most constructivist one, focused on the overall socioeconomic emergency, with a strong collective orientation. It leads to the construction of a proactive perception in dealing with the pandemic. The final narrative is the most vocally multidimensional one, focused on the health emergency. It is more the prerogative of politicians and local administrators, which leads to the construction of a holistic perception. All these groups of perception, as well the readers can understand, become the container that encloses the richness of the arguments, the thematic areas, and the elements of the social narratives from the different steps of the analysis presented.

Thus, in the produced and presented results the potential of a mixed approach applied to content analysis on digital platform social data can be clearly understood, even if reflection is required. However, it is assessable here for its power of theoretical synthesis to restore the vastness of the results in extension and in depth—both qualitative and quantitative—that was produced for this study.

**Contribution to the Field of Mixed Methods Research**

Coming back to our methodological aim, we recall that it consists in showing a hybrid, integrated, and mixed experience of content analysis of particularly dense, diversified, and miscellaneous data: digital platform social data. We can certainly say that the application example on the analysis of risk perceptions of COVID-19 in the Italian Twittersphere worked particularly well to highlight the points on which it is still necessary to reflect (e.g., the black box of algorithms, the pitfalls of packages, and the issues of representativity and generalization, just to mention a few). It also highlighted the strengths that make the mixed approach the only one capable of contrasting and overcoming the limits encountered in each step. This approach was uniquely capable of ensuring, in the final return of the integrated result, something unique, dense, and extremely different from what each method could give us on its own. In particular, thinking about our substantive topic and the research question that opens up the realized model, together with the fact that we have tried to answer this question based on the analysis of the content conveyed on this social platform, what can be affirmed decisively is the ability that our model gives us to get to the essence of risk perception profiles. This is thanks to a mixed analysis procedure that has skillfully combined the power of synthesis with the richness of details in a model that can act as a guide in the analysis of complex phenomena that have their natural extension in the net.

However, the main contribution this paper wants to return to the field of MM research is an extension of the existing definitions of content analysis to include not only qualitative or
quantitative approaches but presenting a real mixed methods approach to content analysis. This is an unsurprising contribution, already somewhat introduced in sedimentary form by Bergman (2010) with the concept of hermeneutic content analysis in a mixed methods perspective. But here it is exploded and operationalized by combining hermeneutic and automated procedures, and by creating a design model with vast application potential, especially when applied to the digital scenario.

Regarding research limitations and further developments, it is obviously necessary to reflect on many points in order to validate the proposed framework. However, it is assessable here for its power of theoretical synthesis to restore the vastness of the results in extension and in depth—both qualitative and quantitative—that were produced for this study. In particular, we will show this result as a way of integrating and visualizing the results coming from a variant of exploratory sequential digital mixed content analysis design, capable of accommodating qualitative and quantitative outcomes, and of allowing a certain order in the reasoning and interpretation of the (almost always complex) phenomenon chosen as a case study. We are aware that we are reflecting on a particularly delicate phenomenon whose evolution and impacts are still ongoing. In such a complex reality, our method proposal can be conceived as a starting point that opens up to new reflections and future developments, continuing to refine the results that can be pursued on both the research paths outlined and on the possibility of their increasingly precise integration. This is because the main research limit resides in the ability to balance the idiosyncrasy of qualitative choices in the pursuit of the extreme objectivity of the qualitative side. Although we have tried to manage this feature, it remains a congenital characteristic of the approach to be implemented ontologically, pushing the pragmatic vocation that substantiates the approach and the possibility of presenting a study with the characteristics of the one carried out in these pages. Finally, returning to the reflection of Hesse-Biber and Johnson (2013) reported in the opening of this article, MMR sees in the current scenario the proliferation of those models for interpreting and deriving critical insights that the authors mention. Rather than searching for new models, it is now time for MM to also bring this approach to the social digital turn, pushing the methodological reflection onto the ontological plane.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

Gabriella Punziano  https://orcid.org/0000-0001-8783-2712

Notes

1. This article should be considered as a collective elaboration, especially for the introduction and the main conclusion. However, the sections “The Mixed Approach in Digital Content Analysis,” “A Variant of Exploratory Sequential Digital Mixed Content Analysis Design,” “The Rationale that Built the Social Narrative of Coronavirus on Twitter and the Relevant Actors in Managing the Most Pervasive Forms of Communication,” “Integrating the Result into the Classification Scheme: How the Digital Mixed Content Analysis Design Contributes to the General Development of Mixed Methods Methodological Debate,” and “Contribution to the Field of Mixed Methods Research” are to be attributed to Gabriella Punziano; the
sections "Content Analysis: Developments and Digital Scenarios," "Digital Methods as Evolution in Content Analysis," and "What Are the Constitutive Dimensions Differentiating Various Risk Perception Profiles? A Multidimensional Analysis with LCA and CA" to Ciro Clemente De Falco; and the sections "The Background of the Case Study to Break Through the Wall of Possibilities Opened by the Application of Mixed Procedures to Digital Content," "The Construction of the Empirical Base: Which Digital Content Can Detect Insights on Risk Perception and How Can We Select Them?,” and “Strengths of the Mixed Strategy and Major Limitations in the Use of Automated Procedures, Access Issues, and Software Used” to Domenico Trezza.

2. R is a free software environment for statistical computing and graphics. R provides a wide variety of statistical (linear and nonlinear modeling, classical statistical tests, time-series analysis, classification, clustering) and graphical techniques over packages.

3. For more information, see https://developer.twitter.com/en/docs/twitter-api/v1/tweets/search/api-reference/get-search-tweets.

4. Rtweet provides users a range of functions designed to extract data from Twitter’s REST and streaming APIs.

5. T-Lab, born with the explicit intent of enslaving the computational and automatic analysis of the text, allows us to apply procedures ranging from the definition of the dictionary system to the basis of vocabulary selection for the construction of the corpus under analysis, to the descriptive statistics of occurrences and co-occurrences, the definition of automatic classification systems, and up to the more sophisticated thematic and multidimensional analysis of texts.

6. NVivo is a mainly qualitative software useful for the management, classification, and analysis of textual, audio, and video material. Very interesting is the function that allows you to work with heterogeneous documents, unstructured data, and large files that, in this study, we frame as digital platform social data.

7. This consists in the application of a series of exploratory procedures concatenated to the data set in analysis that first involves the application of a factorial analysis technique called LCA on textual data. This is useful for producing a summary of the information contained in the analyzed texts; a graphic representation of the network of associations between words and between words and texts; and the connection between textual data and contextual data. In a second phase, a size reduction analysis aims to synthesize the emerging risk profiles into a few groups that are maximally homogeneous internally and heterogeneous externally.

8. The articles recovered from: Social Work; Education; Business and Finance; Linguistic; Psychology; Social Sciences; Computer Science; Family Studies; Ethics; Urban Studies; Telecommunications; Humanities; and Regional and Urban Planning.

References

Back, L., & Puwar, N. (2012). A manifesto for live methods: Provocations and capacities. *The Sociological Review, 60*(1_suppl), 6–17. https://doi.org/10.1111/j.1467-954x.2012.02114.x

Beck, P. U. (1992). *Risk society: Towards a new modernity*. Sage.

Beer, D. (2014). *Punk sociology*. Palgrave.

Benzécri, J. P. (1973). *L’analyse des données*. Dunod.

Berelston, B. (1952). *Content analysis in communication research*. Free Press.

Bergman, M. M. (2010). Hermeneutic content analysis: Textual and audiovisual analyses within a mixed methods framework. In A. Tashakkori & C. Teddlie (Eds.), *SAGE handbook of mixed methods in social & behavioral research* (2nd ed., pp. 379–396). Sage.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research, 3*(2003), 993–1022.

Bryman, A. (2012). *Social research methods* (4th ed.). Oxford University Press.
Caliandro, A., & Gandini, A. (2016). *Qualitative research in digital environments: A research toolkit*. Routledge.

Carley, K. (1990). Content analysis. In R. E. Asher (Ed.), *The encyclopedia of language and linguistics* (pp. 725–730). Pergamon.

Cole, F. L. (1988). Content analysis: Process and application. *Clinical Nurse Specialist*, 2(1), 53–57. https://doi.org/10.1097/00002800-19880210-00025

Creswell, J. W. (1999). Mixed-method research: Introduction and application. In G. J. Cizek (Ed.), *Handbook of educational policy* (pp. 455–472). Academic.

Creswell, J. W., & Plano Clark, V. L. (2017). *Designing and conducting mixed methods research* (3rd ed.). Sage.

Decreto del Presidente del Consiglio dei Ministri [DPCM]. (2020, March 8). Ulteriori disposizioni attuative del decreto-legge 23 febbraio 2020, n. 6, recante misure urgenti in materia di contenimento e gestione dell’emergenza epidemiologica da COVID-19, (20°01522) (GU Serie Generale n.59 del 08-03-2020). https://www.gazzettaufficiale.it/eli/id/2020/03/08/20A01522/sg

Global Digital Report. (2020). *Datareportal*. https://datareportal.com/reports/digital-2020-global-digital-overview

Governo Italiano Presidenza del Consiglio dei Ministri [GIPCM], (2020, March 9). Decreto iorestoacasa. https://www.governo.it/it/articolo/firmato-il-dpcm-9-marzo-2020/14276

Hamad, E. O., Savundranayagam, M. Y., Holmes, J. D., Kinsella, E. A., & Johnson, A. M. (2016). Toward a mixed-methods research approach to content analysis in the digital age: The combined content-analysis model and its applications to health care Twitter feeds. *Journal of Medical Internet Research*, 18(3), Article e60. https://doi.org/10.2196/jmir.5391

Hernandez-Suarez, A., Sanchez-Perez, G., Toscano-Medina, K., Martinez-Hernandez, V., Sanchez, V., & Perez-Meana, H. (2018). A web scraping methodology for bypassing twitter API restrictions. arXiv preprint arXiv:1803.09875

Herring, S. C. (2009). Web content analysis: Expanding the paradigm. In J. Hunsinger (ed.) *International Handbook of Internet Research*, Netherlands: Springer, 233–249. https://doi.org/10.1007/978-1-4020-9789-8_14

Hesse-Biber, S., & Johnson, R. B. (2013). Coming at things differently: Future directions of possible engagement with mixed methods research. *Journal of Mixed Methods Research*, 7(2), 103–109. https://doi.org/10.1177/1558689813483987

Hine, C. (2005). *Virtual methods: Issues in social research on the Internet*. Berg Publisher.

Holsti, O. R. (1969). *Content analysis for the social sciences and humanities*. Addison-Wesley.

Hsieh, H., & Shannon, S. E. (2005). Three approaches to qualitative content analysis. *Qualitative Health Research, 15*(9), 1277–1288. https://doi.org/10.1177/1049732305276687

Jelodar, H., Wang, Y., Yuan, C., Feng, X., Jiang, X., Li, Y., & Zhao, L. (2019). Latent Dirichlet allocation (LDA) and topic modeling: Models, applications, a survey. *Multimedia Tools and Applications*, 78(11), 15169–15211. https://doi.org/10.1007/s11042-018-6894-4

Keyes, K. M., & Westreich, D. (2019). UK Biobank, big data, and the consequences of non-representativeness. *Lancet*, 393(10178), 1297. https://doi.org/10.1016/s0140-6736(18)33067-8

Kitchin, R. (2014). *The data revolution: Big data, open data, data infrastructures and their consequences*. Sage.

Krippendorff, K. (2018). *Content analysis: An introduction to its methodology* (4th ed.). Sage.

Lasswell, H. D., Leites, N., Kadner, R., Goldsen, J. M., Grey, A., & Janis, I. L. (1949). *The language of politics. Studies in quantitative semantics*. George Stewart.

Lebart, L. (1994). Complementary use of correspondence analysis and cluster analysis. In M. J. Greenacre & J. Blasius (Eds.), *Correspondence analysis in the social sciences* (pp. 162–178). Academic.

Lebart, L., Salem, A., & Berry, L. (1998). *Exploring textual data*. Springer Science & Business Media.
Leetaru, K. (2019, March 7). How data scientists turned against statistics. Forbes. https://www.forbes.com/sites/kalevleetaru/2019/03/07/how-data-scientists-turned-against-statistics/?sh=3d5366257c8

Lewis, S. C., Zamith, R., & Hermida, A. (2013). Content analysis in an era of big data: A hybrid approach to computational and manual methods. Journal of Broadcasting & Electronic Media, 57(1), 34–52. https://doi.org/10.1080/08838151.2012.761702

Lupton, D. (2014). Digital sociology. Routledge.

Marres, N. (2017). Digital sociology: The reinvention of social research. John Wiley & Sons.

Orina, W. A., Mwangi, G. F., Sitati, R. N., & Nyabola, F. (2015). Content analysis and a critical review of the exploratory design. General Education Journal, 4(2), 32–45.

Renn, O. (1992). Risk communication: Towards a rational discourse with the public. Journal of Hazardous Materials, 29(3), 465–519. https://doi.org/10.1016/0304-3894(92)85047-5

Riff, D., Lacy, S., Fico, F., & Watson, B. (2019). Analyzing media messages: Using quantitative content analysis in research (4th ed.). Routledge.

Rogers, R. (2009). The end of the virtual: Digital methods. Amsterdam University Press.

Rogers, R. (2013). Digital methods. MIT Press.

Schreier, M. (2012). Qualitative content analysis in practice. Sage.

Schreier, M. (2013). Qualitative content analysis. In E. Flick (Ed.), The Sage handbook of qualitative data analysis (pp.170–183). Sage.

Snelson, C. L. (2016). Qualitative and mixed methods social media research: A review of the literature. International Journal of Qualitative Methods, 15(1), 1–15. https://doi.org/10.1177/1609406915624574

Teddlie, C., & Tashakkori, A. (2011). Mixed methods research: Contemporary issues in an emerging field. In N. K. Denzin & Y. S. Lincoln (Eds.), Handbook of qualitative research (4th ed., pp. 285–300). Sage.

Vaezi, A., & Javanmard, S. H. (2020). Infodemic and risk communication in the era of CoV-19. Advanced Biomedical Research, 9(1), 10. https://doi.org/10.4103/abr.abr_47_20

Zeng, J. (2012). A topic modeling toolbox using belief propagation. The Journal of Machine Learning Research, 13(1), 2233–2236.