Sustainable Development Goals as unifying narratives in large UK firms' Twitter discussions

Alessia Patuelli[1,2,*] **and Fabio Saracco**[3,4,1]

[1]Laboratory for the Analysis of CompleX Economic Systems (AXES), IMT School for Advanced Studies Lucca, 55100 Lucca, Italy
[2]Department of Economics and Management, University of Florence, Florence, 50127, Italy
[3]Institute for Applied Computing, National Research Council of Italy, 00185 Rome, Italy
[4]'Enrico Fermi' Research Center (CREF), 00184 Rome, Italy
[*]alessia.patuelli@imtlucca.it

ABSTRACT

Since 2015, the United Nations have called for a global effort to reach Sustainable Development Goals (SDGs). Firms play a vital role in contributing to SDGs. While many empirical approaches were used to map firms' contributions to SDGs, online social networks are an underexplored but promising setting. This paper maps large UK firms' discussions on Twitter, specifically focusing on their SDG-related discussions, with complex network methods from statistical physics. Results show that: 1) SDGs are the topics that tie conversations among major UK firms together; 2) compared to the environmental and economic dimensions, the social dimension is predominant; 3) the attention to different SDGs varies depending on the community and sector firms belong to; 4) the use of retweets on SDGs-related tweets highlights a high stakeholder engagement on global challenges; 5) large UK companies and stakeholders generally behave differently from Italian ones. This paper provides theoretical contributions, combining institutional, stakeholder and legitimacy theories. It also contributes to developing the literature on businesses and SDGs with an interdisciplinary approach. It offers practical implications and a big-data based tool to monitor firms' discussions on SDGs on Twitter. Last, the paper proposes new avenues for further research.

Introduction

Online social networks have changed communication, making it cheaper and faster than before and providing a new channel for businesses to engage and directly interact with their stakeholders. They now represent a crucial means of disseminating firms' Corporate Social Responsibility (CSR) activities and involving stakeholders. Although the term "corporate social responsibility" has been widely defined[1], it refers to a company's relationships and responsibilities to society, regarded as the groups of stakeholders with which it interacts[1–3]. It comprises all firms' activities beyond what is required by law[4]. What CSR means in practice varies on the cultural and historical environment in which a company operates. It may also represent the difficulties that a company is dealing with at the time[3].

Stakeholder engagement has been differently defined and may be viewed under many different theoretical viewpoints while being a separate but related term to CSR. It has been conceptualized as "practices the organization undertakes to involve stakeholders positively in organizational activities"[5] (p. 315). Online social networks provide the tools to measure it, assuming that the users belong to the firm's stakeholders[6]. On Twitter, users have the chance to retweet and like posts, which can be considered an endorsement of the message's substance[7]. Legitimacy theory and stakeholder theory are the two primary approaches[8,9] that explain why companies are active in online social networks. On the one hand, legitimacy theory claims that businesses act following society's expectations and ideals. These are not constant and change across time and space. Although several scholars have related legitimacy theory to CSR, it is not necessarily restricted by CSR or stakeholder expectations. According to this perspective, firms use online social networks to justify their position in society[10]. On the other hand, following stakeholder theory, firms should follow stakeholders' expectations to create long-term value. Consistently with this approach, firms utilize online social networks to communicate with their stakeholders and share their strategies and outcomes[11].

In 2015, the United Nations established 17 Sustainable Development and 169 targets to be reached by a joint effort from all members of society by 2030. The goals balance three dimensions of sustainable development (economic, social, and environmental) and encourage action in areas vital for humanity and the world. The SDGs have put substantial pressure on international economies and firms, and firms are crucial development players in achieving the SDGs[12]. Despite being primarily a societal phenomenon, SDGs have the potential to significantly advance CSR research[13] , with CSR serving as a theoretical framework to examine how and to what degree businesses contribute to the SDGs. In fact, the 17 SDGs can be grouped into three dimensions according to the well-known classification of CSR[15]: the social dimension (1–5, 10, 16, and 17), the economic

(7–9, 11, and 12), and the environmental one (6, 13–15)[16].

Research on firms and SDGs is relatively recent and varied. Broadly, studies are investigating the reasons why and factors that drive firms' SDGs adoption[17], their challenges[18], how SDGs are implemented in the firms' strategies and activities[19–21] and how SDGs achievements are reported[17,18,22] and communicated[23]. In some cases, studies investigated several dimensions at one time[24]. Researchers have used different empirical approaches, which include case studies[18,19,21], the analysis of financial or sustainability reports[22], and of websites[24]. The first findings show that businesses seem to focus on specific SDGs. The most common ones are SDG 3 (good health and well-being), 4 (quality education), 9 (industry, innovation and infrastructure) and 12 (responsible consumption and production). Belonging to a sector with high sustainability impacts (for example, energy, communication technology, utility sectors), as well as being based in specific regions (firms based in South America and Europe show the highest percentages of SDGs reporting) seem to influence the decision to pursue SDGs[25]. As for now, research is overlooking the role of online social networks in understanding firms' contributions to SDGs. To the best of our knowledge, only a few papers explore to what extent firms discuss SDGs themes on online social networks[26,27] and to what extent stakeholders interact on these themes[28–30]. As for now, it seems that SDGs have limited relevance in the online debate[26]. Also, there seems to be a limited involvement of stakeholders on SDGs posts, consistently with similar studies on CSR[10]. However, present studies use limited samples, or focus on specific industries, while a broader perspective is currently missing. In this paper, we aim to answer the following broad research question: *To what extent are businesses discussing SDGs themes on online social networks?*. We base on Twitter, which is extensively used in online societal debates since its structure is particularly suitable for short messages. It is an excellent source for investigation, together with its data availability. In contrast to interviews and survey-based data, this method does not rely on response rates or an individual's desire to respond to get a bigger sample size. We decided to focus on large firms, as they have a high social impact[31,32] and are eager to engage in social and environmental dissemination[11]. Moreover, compared to small and medium enterprises, they often have more stakeholders requiring information[31]. Overall, we believe that large firms are suitable for our study and focus on all large firms in one European country, the United Kingdom. Additionally, we compare our findings on the UK firms with analogous research investigating Italian large firms' Twitter discussion and CSR orientation. Developing from the social and institutional paradigm, we expect that firms belonging to countries with different institutions, cultures, and values show different patterns regarding firms' communication behaviours[33,34].

This research aims to respond to previous calls for an empirical investigation of firms' advancements towards SDGs with big data, a quick and low-cost tool[35,36]. Indeed, online social networks provide a new, underexploited tool to understand firms' challenges, CSR activities and stakeholder engagement[37]. As for now, only one study[28] examines firms' discussion about SDGs and stakeholder engagement on Twitter. However, the sample of firms considered is quite limited. To the best of our knowledge, this is the first study to apply complex network methods to a wide set of Twitter data to address the problem of uncovering firms' discussion and contributions to SDGs.

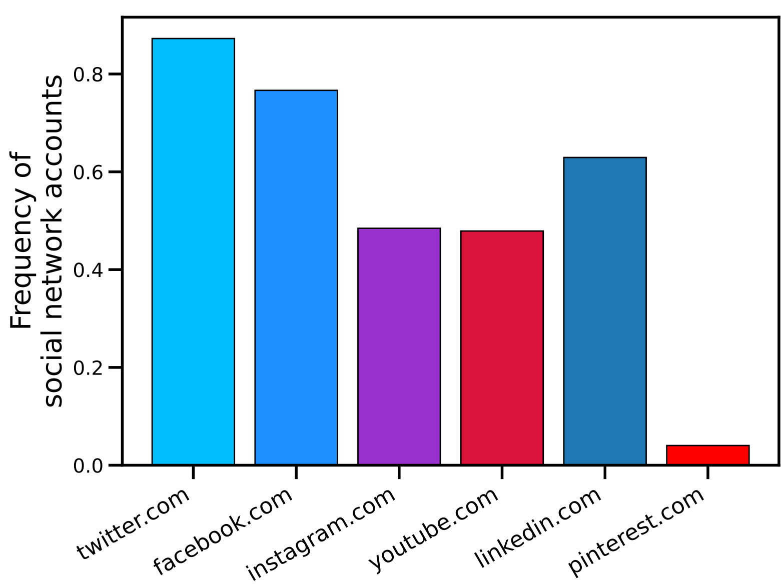

Figure 1. Online social networks adoption among large firms in the UK. Twitter is the most popular online social network for large firms in UK, overcoming the other ones.

Results

Data description

In order to highlight firms' communication about SDGs, we first downloaded from Orbis (Bureau Van Dijk, https://www.bvdinfo.com/) the primary information regarding large companies (i.e. those with more than 250 employees), such as the name, address, number of employees, total assets, NACE code, and the website. Then, we automatically extracted the Twitter account of the related firm from each website, if present. Please find more details about the automatic Twitter account search in the Methods section. We found that Twitter is an excellent tool for this analysis, as it is the most widely adopted online social network by UK large firms. As Fig. 1 shows, nearly 87.3% of the largest UK firms have a Twitter account, overcoming other online social platforms.

Finally, we downloaded the timeline of each Twitter account using the official API (specifically, using the command: `GET /2/users/:id/tweets`) via `tweepy` python wrapper. Doing so allows us to access the most recent ~ 3200 messages. We focused on the period between 2021/02/17 and 2022/02/17. As in Ref.[37], we consider active accounts in the entire period (and not just a portion). While this choice may appear too conservative, it allows concentrating our efforts on subjects that have continuously contributed to creating a common shared narrative. Using this time restriction, we ended up with 3.1M tweets, out of which 596k retweets and 609k replies. As we focus on the interaction between Twitter accounts and hashtags, we excluded 318 accounts out of 6179 of the original dataset since they did not use any hashtag in the considered period.

Before presenting the main results, let us simply measure the recurrence of SDG hashtags in our dataset, i.e. count how many messages contain hashtags related to SDGs, see Fig. 2. SDGs are a crucial topic in firms' communication: each SDG hashtag appears, on average, in 99.01 messages, against the 7.56 of the average hashtag in our dataset.

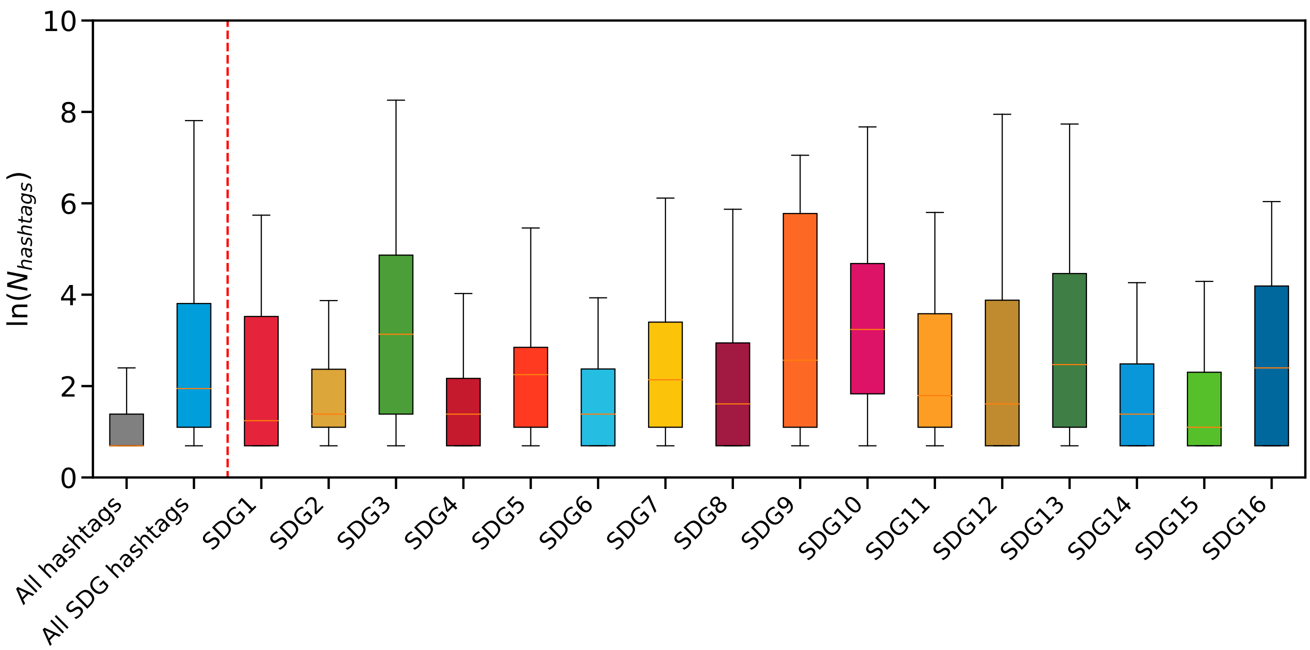

Figure 2. Distribution of the number of messages per (SDG) hashtag. The boxplots compare the distribution of the number of messages in which each hashtag appear for: all hashtags (the gray box on the left); for SDG hashtags (all the boxes beyond the red line; boxes are colored using the official indication from the United Nations, `https://www.un.org/sustainabledevelopment/wp-content/uploads/2019/01/SDG_Guidelines_AUG_2019_Final.pdf`) and; for all the SDGs hashtags (the sky blue box on the left). The boxplots show the distribution of the logarithm of the number messages per hashtags, since the distributions are heavy-tailed. In this sense, boxplots may not be the perfect tool for capturing the distribution properties but can effectively deliver the message about the rough differences among the various distributions. In particular, SDGs hashtags appear more frequently than "standard" hashtags in the communication strategy of large firms, thus representing crucial topics.

The validated network of Twitter accounts

To study how UK companies contribute to the evolution of common narratives, we considered the bipartite network composed of the firms' Twitter accounts and their hashtags. In a bipartite network, nodes are divided into two sets, called *layers*, and links can only connect nodes from different layers. In our application, the two layers include 5,859 accounts and 136,504 different hashtags. To highlight those accounts using the same hashtags, as in Ref.[37], we use the validation projection procedure proposed in Ref.[38]. In a nutshell, any couple of accounts are connected if the number of hashtags they used is statistically significant (i.e. it cannot be explained by their hashtag usage and the popularity of the various hashtags). Please find more details in the Methods section. The result of this validation projection is a monopartite network of Twitter 3,629 accounts and 59,158 links. The relative Largest Connected Component (LCC) is represented in Fig. 3.

Before describing the network and its structure, we highlight a few remarks. First, comparing the filtering of UK data to the same procedure on Italian data as in Ref.[37], the percentages of validated Twitter's account nodes are 61.9% and 19.2%, respectively. These percentages indicate the fraction of accounts whose usage of hashtags differs substantially from a random behaviour and whose communication strategy presents significant similarities with other accounts. In this sense, a low frequency of validated nodes can mean that many accounts focus on the peculiarities of their communication. In contrast, a high frequency of validated nodes can mean that the communication is more homogeneous and strongly related to common narratives.
Second, SDGs are among the subjects contributing the most to developing common narratives. Validated users (those passing the validation procedure described above) contribute with no less than the 85% of the SDGs hashtags of the entire datasets, see Table 1. Otherwise stated, most Twitter accounts using SDGs in their communications pass our filter. This result is remarkable since the validation procedure of Ref.[38] is restrictive, as tested in different contexts: such a strong signal indicates a non-trivial activity on SDG communication.

SDG01	SDG02	SDG03	SDG04	SDG05	SDG06	SDG07	SDG08
85.19%	97.91%	96.15%	92.65%	91.07%	90.54%	93.56%	92.50%

SDG09	SDG10	SDG11	SDG12	SDG13	SDG14	SDG15	SDG16
95.93%	94.89%	88.08%	96.13%	95.28%	92.80%	94.79%	86.55%

Table 1. SDG hashtags used by validated users over their usage in the entire dataset. These high percentages indicate that most active users using SDG hashtags are validated by the filtering procedure. In turn, it implies that SDGs are among the main subjects shaping the various common narratives of large firm's accounts on Twitter.

Description of the communities

To extract more information, we ran the Louvain community detection algorithm [39] on the validated network, highlighting four main communities displayed in the left panel of Fig. 3. Rerunning the same Louvain community detection algorithm inside each community shows a more detailed description, which is represented in the right panel of Fig. 3.
The communities in Fig. 3 mostly revolve around social and environmental themes, showing that CSR themes are indeed fundamental in firms' communication on Twitter, consistently with [37]. Communities are generally coherent with the sectors the firms belong to, as captured by NACE code Rev.2 at 1 digit (see Fig. 4). This coherence reflects the themes discussed: the most addressed CSR themes are the ones closest to the firms' sector, as represented in Fig. 5.
This confirms that CSR changes according to the specific context[3]. Moreover, we show that the social dimension appears more critical than the environmental one. Although this finding contrasts with most previous literature[40], it seems in line with more recent findings[37]. Community Cyan is a sort of exception among the various groups, as it comprises three main sectors (Professional, scientific and technical activities; Information and communication; Manufacturing). Its top hashtags reflect digital innovation, environmental sustainability, social and economic themes (see Table 4). They are coherent with the wide range of SDGs mentioned (SDG10 refers to the social dimension; SDG9 to the economic one; SDG12 and SDG13 to the environmental dimension). The other communities revolve around social themes and hashtags. Community Orange-red is composed of firms from two sectors (Human health and social work activities; Education) and focuses on social themes (see Tables 5 and 6). Coherently, hashtags relate to SDGs from the social dimension, namely SDG3, SDG5 and SDG10. Community Yellow comprises firms from one sector (Education) and mainly discusses social themes (see Table 7). The most mentioned SDGs come from the social dimension: SDG1, SDG3, SDG4, SDG5, SDG10, and SDG16. Similarly, Orchid has companies from only one sector (Human health and social work activities). It is focused on social themes as well (see Table 8). As community Yellow, its SDGs belong to the social dimension: SDG3, SDG5, and SDG10.

Hashtag frequency

In this subsection we will focus on the four major communities, i.e. Cyan, Orange-red, Yellow and Orchid. For each community, we show the most recurring hashtags in the biggest subcommunities (with more than 50 nodes), which reflect the main themes that businesses discuss. The following analysis is based on the results summarised in Tables 4, 5, 6, 7 and 8.

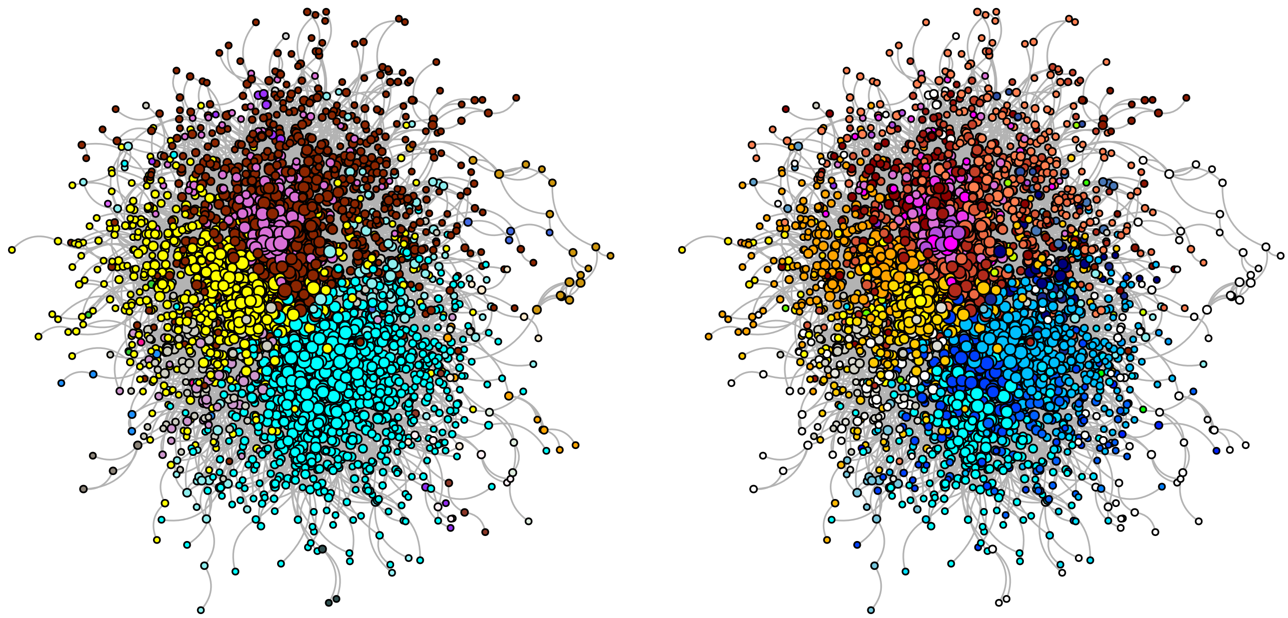

Figure 3. The Largest Connected Component of the validated projected network of users. The dimension of the nodes is proportional to the logarithm of their degree. In the left panel, the different node colors represent the various communities, as detected by Louvain algorithm; in the right panel the colors identify the subcommunities (nodes that do not fall in one of the main subcommunities are plotted in white).

The Cyan subcommunity show some diversity in its themes, which span from digital innovation, and environmental sustainability to social and economic themes. While subcommunity n. 0 is focused on digital innovation themes (e.g., "cloud", "technology", "digital", "digitaltransformation", "innovation", "webinar"), subcommunity n. 2 mostly focuses on environmental sustainability (e.g. "sustainability", "netzero", "climatechange", "earthday"). While mostly focusing on social themes, subcommunity n. 1 also contains some environmental themes (e.g., "blackhistorymonth", "pridemonth", "internationalwomensday" for the social side; "earthday" and "sustainability" for the environmental one). Similarly, subcommunity n. 5 does not have a clear orientation towards one theme: it contains both environmental and economic themes (e.g., "sustainability", "climatechange", as well as "esg", "inflation" and "supplychain"). Differently from the other subcommunities, n. 6 does not have a specific focus (it includes various hashtags, as "covid", "budget", "webinar", "internationalwomensday", "Brexit" and "podcast").

Compared to the Cyan community, the themes in the Orange-red, Yellow and Orchid show more homogeneity. The Orange-red community is highly focused on social themes. Among the 7 subcommunities, all of them show hashtags related to social themes. Six of them have social themes as the prevalent ones within the subcommunity. For example, subcommunity n. 1, which has a high prevalence of hashtags related to the social dimension, has "internationalwomensday", "mentalhealthawarenessweek", "pridemonth", "blackhistorymonth" among the most frequent hashtags. Something similar happens with the subcommunity n. 2, 3, 4 and 5. Conversely, the subcommunity n. 0, while having a few hashtags related to social themes, is more focused on festivities (e.g. "Christmas", "valentinesday", "halloween").

Community Yellow discusses social themes in all the three subcommunities, which are mostly homogeneous. For example, all the three subcommunities mention "mentalhealthawarenessweek" and "internationalwomensday" among their top 10 hashtags, with some differences in other hashtags. Only subcommunity n. 3 comprises themes related to engineering education (including hashtags like "engineering", "education", "construction", "apprenticeship").

Community Orchid is focused on social themes as well. It includes Covid themes among all its subcommunities: in this case, the social dimension is connected to the pandemic. For example, all the four subcommunities associate "covid" and "covidvaccine" with "nhs", "internationalwomensday" and "mentallhealthawarenessweek". Overall, the hashtag "covid" is generally found in many subcommunities , but it is not associated with other related words. It has a higher relevance only in the Orchid community.

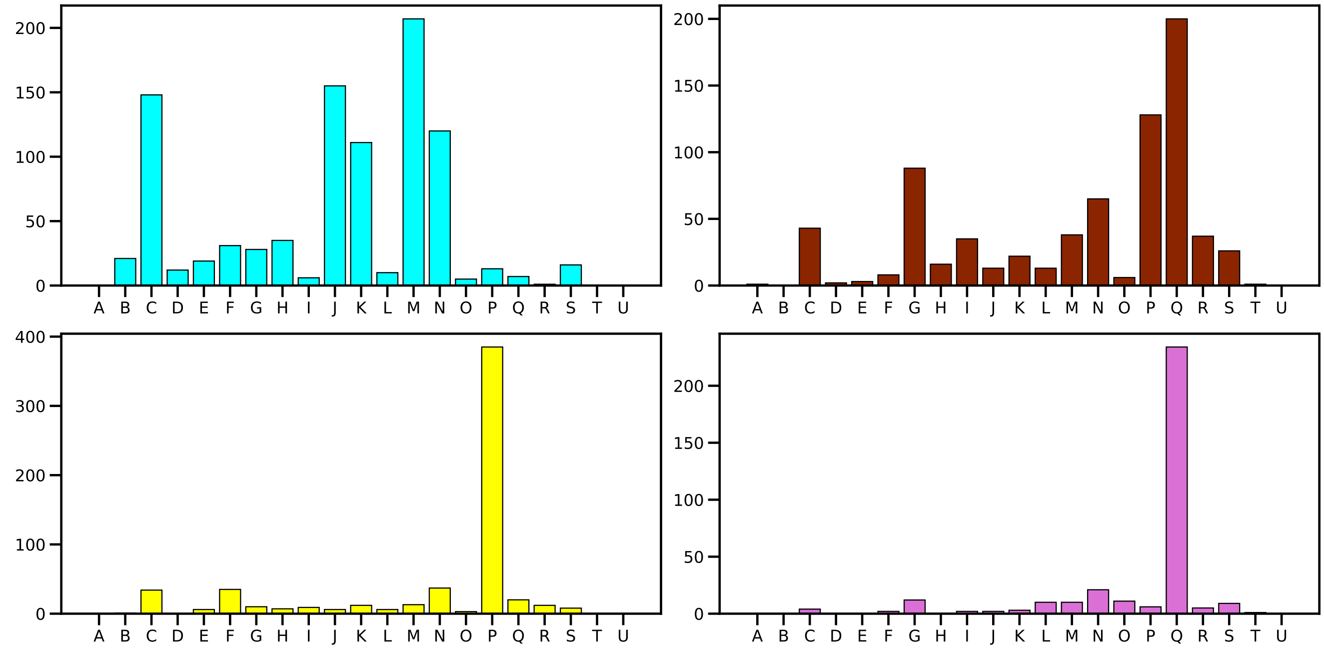

Figure 4. The NACE Rev.2, 1 digit composition of the four greatest community in the validated network of Fig. 3. The colors of the different bar charts are the same as in the left panel of Fig. 3. Please find the identification of the various NACE in Table 3 in the Methods section. To enhance clarity, we anticipate here the most present sectors, which are: C) Manufacturing; G) Wholesale and retail trade; repair of motor vehicles and motorcycles; J) Information and communication; K) Financial and insurance activities; M) Professional, scientific and technical activities; N) Administrative and support service activities; P) Education; Q) Human health and social work activities.

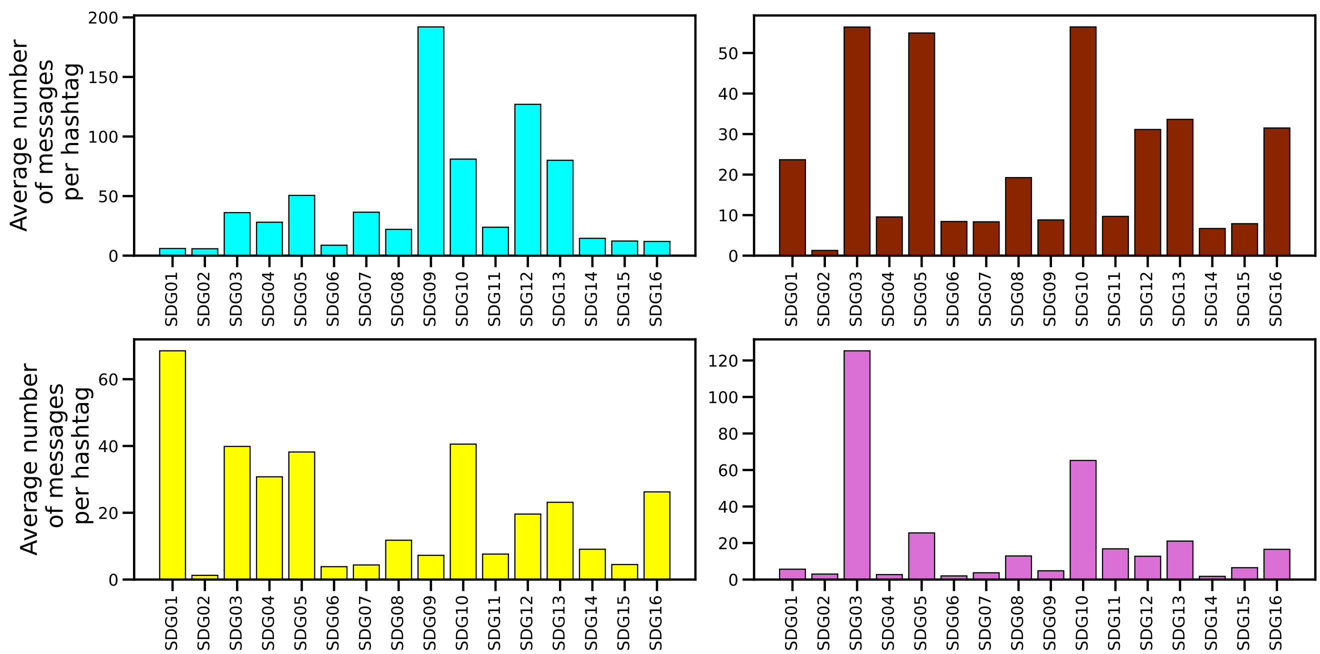

Figure 5. The SDG activity of the 4 greatest community in the validated network of Fig. 3. The colors of the different bar charts are the same as in the left panel of Fig. 3. The identification of the various hashtags with the different SDGs is described in details in the Methods section.

Stakeholder engagement

This subsection focuses on stakeholder engagement with the narratives developed by firms' accounts. First, the dataset shows that the average number of retweets and likes per hashtag is 15.85 and 23.16. This shows that UK firms' stakeholders tend to use more likes than retweets when interacting on Twitter. However, further analyses reveal that this pattern is the opposite when stakeholders interact with companies on SDGs subjects.As Fig. 6 shows, when stakeholders interact with SDGs hashtags, they put a lower number of likes but retweet more than they do with non-SDGs hashtags. The average numbers of likes and retweets per SDG hashtag are 6.71 and 19.83, highlighting a higher engagement on SDGs themes. We also highlight that stakeholder engagement with large companies in the UK is different compared to Italy[37], where the average number of retweets and likes per hashtag were 5.39 and 14.83.

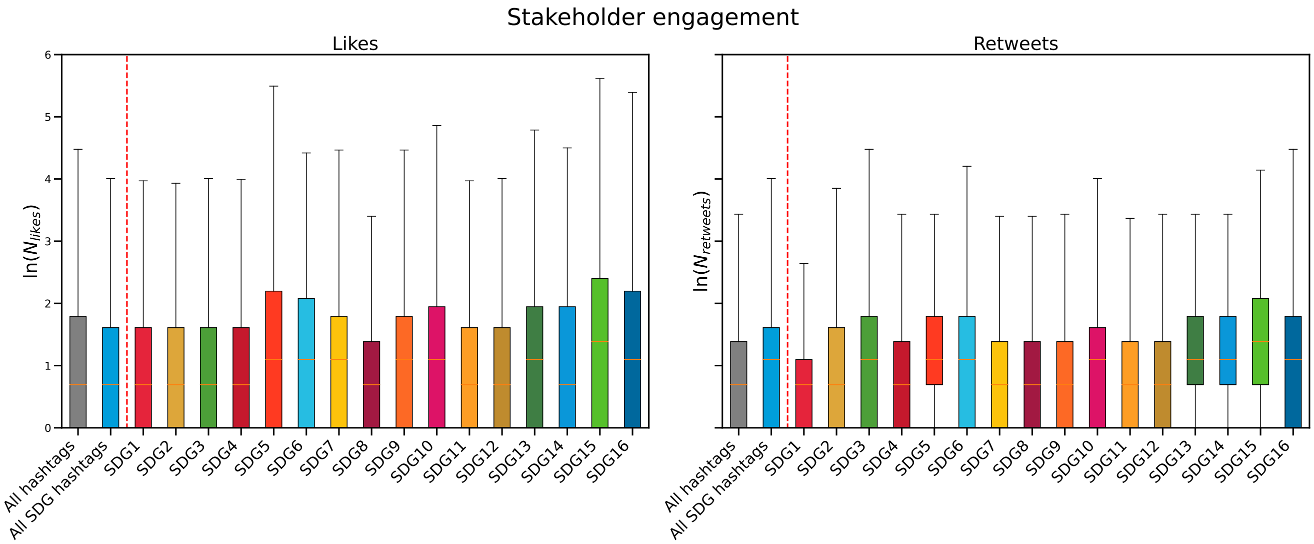

Figure 6. Stakeholder engagement on the SDGs. These boxplots compare the distribution of the number of likes and retweets for all hashtags (the gray box on the left), for each SDG (all the boxes beyond the red line; boxes are colored using the official indication from UN, `https://www.un.org/sustainabledevelopment/wp-content/uploads/2019/01/SDG_Guidelines_AUG_2019_Final.pdf`), and for all the SDGs hashtags (the sky blue box on the left). The boxplots show the distribution of the logarithm of the number of likes and retweets. We used the logarithms because the distributions are heavy-tailed. In this sense, boxplots may not be the perfect tool for capturing the distribution properties but can effectively deliver the message about the rough differences among the various distributions.

Stakeholder engagement on the various SDGs depends on the community and sector. For example, in community Orange-red, SDG5 and SDG16 hashtags (i.e. 'Gender Equality' and 'Peace, Justice, and Strong Institutions') received more retweets, on average, than other SDG hashtags, and more than random hashtags. This community mainly comprises firms in sectors 'P' and 'Q' ('Education' and 'Human health and social work activities'). Analogous considerations can be done for the other communities.

Discussion and conclusions

This paper presents large UK firms' discussions on Twitter, specifically focusing on SDGs. It shows that: 1) SDGs are the themes that unite firms' discussions; 2) the social dimension is prevalent, compared to the environmental and economic ones; 3) the interest in specific SDGs depends on the community and sector a firm belongs to; 4) stakeholders are highly engaged on SDGs themes, using more retweets than likes when interacting with a tweet that contains an SDG-related hashtag; 5) overall, large UK firms and stakeholders show substantially different behaviours compared to the Italian ones. We will discuss these points in the following paragraph.

First, communities of discussion naturally arise from the data. These communities are uniform and based on common narratives. Most importantly, the shared narratives are centred around SDGs themes. Large UK firms use Twitter to participate in broader discussions on widely acknowledged themes (such as "internationalwomensday"). Thus, we believe that our results for the UK support stakeholder theory: large firms use Twitter to engage in discussions on highly socially relevant themes. This finding gives a different perspective compared to previous research, which states that CSR themes are overlooked by firms in their

communications on online social networks[10,37,41,42]. While not contradicting these previous studies, we show that SDGs themes unify the firms' discussions, creating different communities in the UK debate on Twitter. In doing so, we also highlight the importance of integrating different methodologies into business research, uncovering patterns that would not show using traditional methods[43].
Second, the recurrent themes in the communities mainly focus on the social dimension, with discussions on environmental and economic themes that are present but less relevant. Our findings oppose traditional CSR literature, which maintains that environmental themes are the primary dimension[40].
Third, we highlight that the interest in SDGs depends on the community a firm belongs to, and the community mostly depends on the firm's sector. This is consistent with previous works about SDGs, which argue that the interest in SDGs depends on the sector the firm belongs to[25]. It is also consistent with previous findings about large Italian firms discussing CSR themes[37]. It shows that in the UK and Italy large firms' dialogue largely depends on the sectors to which they belong. Only community Cyan discusses the themes shared by large Italian firms[37] (i.e. the digital transformation, environmental sustainability, Covid and the economic dimension). This appears to be substantially different from the behaviour of large Italian firms, as described in[37], where legitimacy theory was used to explain firms' online behaviour.
Fourth, results highlight stakeholder engagement with retweets is higher on SDG-related tweets than on general tweets. As retweets are a more significant endorsement of the author of the post[44–46], the higher number of retweets on SDGs themes highlights a more significant engagement with the global challenges.

Overall, the results highlight a different use of Twitter by UK firms compared to the Italian ones. Thus, we highlight that, consistently with institutional theory[47], different institutional and cultural settings translate into different behaviours, including corporate communications on online social networks. These differences are not limited to localized behaviours. However, they appear to relate to the fundamental reasons why companies interact in online social networks, highlighting that the results support different theories in the two countries.

Our paper brings several contributions. First, we contribute to institutional[47], stakeholder and legitimacy theories[8,9], explaining the UK and Italian firms' different attitudes on Twitter by combining the three approaches. We argue that stakeholder and legitimacy theories coexist and explain firms' behaviours in the two countries. Following the institutional theory, we believe that different institutional settings, values and cultures explain different behaviours and their reasons.
Second, we answer previous calls to map firms' contributions to SDGS[35,36], advancing preliminary studies[28] with interdisciplinary approaches on a wide dataset. Our research highlights that Twitter posts concerning SDGs unify the firms, naturally creating discussion communities. It also shows the prevalence of the social dimension, as opposed to the environmental and economic one, and a higher engagement of stakeholders on these themes compared to general posts.
Third, on the practical side, this research offers a tool for monitoring the discussions about SDGs that firms are developing on Twitter. Such a tool could be helpful for policymakers to map the extent to which firms pay attention to global challenges. It could also be helpful to understand how firms' and users' perceptions of SDGs vary.
This paper opens new paths for future research. First, as SDGs group large firms' online discussions, it would be interesting to investigate how this phenomenon developed over time. As SDGs were set in 2015, it would be interesting to go back in time and check if and how this trend increased in the past years. Also, according to the United Nations, SDGs should be reached by 2030. It would be interesting to develop a study on the whole 15-years period.
Consistent with Ref.[37], this research shows a higher interest in discussing social themes rather than environmental ones. As these results contradict[40] but are limited to two European countries, it would be interesting to check if such findings hold for other contexts. Thus, future research should investigate to what extent firms discuss the social and environmental dimensions on online social networks on a broader scale. Although such an analysis would primarily focus on the communication dimension, we believe it would shed light on the prominent CSR dimension firms focus on.
Last, we show the SDGs firms are most talking about. While we can assume that online communication reflects the firms' strategies and activities[48], we do not have enough data to claim that companies are actually pursuing the SDGs they are discussing in the online social networks. Further research could dig deeper into this issue to unravel how firms' communication about SDGs themes is consistent with their real-world activities on global challenges.

Methods

From websites to Twitter accounts

As a first step, we downloaded companies' websites from Orbis and, automatically accessing them, we got the relative Twitter accounts, when present. In order to test our scraping algorithm, we took a sample of 100 websites and manually extracted the Twitter accounts from them.
Then, if both the human and the scraping algorithm agree on the account, we assign a true positive (TP) to each of them. If they cannot find any Twitter account, we assign a true negative (TN) to both. In other cases (i.e., when one finds an account and the

other does not, or if they disagree on the account found), we manually checked directly from Twitter which method returns the correct answer. Even if the scraping algorithm performances are not astonishing, they overcome the human ones. Please find the performances of the automated tool for getting the Twitter accounts in Table 2.

	Machine	Human
precision=$\frac{TP}{TP+FP}$	**96.7%**	94.9%
accuracy=$\frac{TP+TN}{P+N}$	**83.0%**	81.0%
sensibility=$\frac{TP}{TP+FN}$	**79.5%**	77.8%
specificity=$\frac{TN}{TN+FP}$	**92.6%**	89.3%

Table 2. Scraping algorithm performances vs. human annotation; best performances in bold. Machine performances always overcome human ones. From the data above, the primary human and machine problem seems to be missing the existing Twitter accounts (low sensibility). At the same time, if they find anything, they get the right one in most cases (high precision). The high specificity tells us that both the human and the machine can spot when the Twitter account is absent.

Hashtag data cleaning: edit distance

In order to properly consider the issue of misspelt hashtags or to consider as a single word singular and plural nouns, we used edit distance, as implemented by the `py_stringmatching` python module[49]. In order to obtain the most effective threshold, we randomly picked couples of keywords and selected the first 100 couples with an edit similarity score greater than 0.8. Then, we manually checked when the hashtags effectively represent different words or if they refer to the same concept. For this 100 couples' sample, we calculated the precision and the accuracy for the various values of the threshold (sensibility and specificity are trivial and do not carry any relevant information): the most effective threshold for edit similarity is 0.86.

Hashtag cleaning

A significant part of hashtags refers to acronyms, and comparisons among them may cause false matches between unrelated terms. Thus, we first removed digits from the hashtags, except for 'Euro2020', since it is an event that was central in the UK in the analysed period. Removing digits would have introduced mismatches and errors. Then we turned all hashtags to lower cases and considered their frequencies. In principle, using edit distance for hashtag cleaning, we should have compared all couple of hashtags, thus performing $O(N^2)$ tests.
In order to limit the efforts dedicated to hashtag cleaning to $O(N)$, we implemented the following procedure. First, we selected all hashtags appearing in the dataset more than 50 times, resulting in 922 different hashtags. In this 'benchmark set', we first select all couples of words displaying an edit distance greater than the edit threshold of 0.86. Among those, we choose the less frequent hashtag for every couple and remove it from the benchmark set, resulting in a total of 916 different hashtags. We finally compared all hashtags with the ones in the benchmark set. All hashtags that displayed an edit similarity greater than the threshold with another hashtag in the benchmark set were then substituted with their more frequent partner. After the cleaning, we have 136,504 different hashtags.

Bipartite Configuration Model analysis

After the cleaning, we build a bipartite network in which the two layers represent respectively firms' Twitter accounts and the used hashtags, as in Ref.[37]. The two layers includes respectively 5,859 accounts and 136,504 different hashtags.
In order to have a proper benchmark for our analyses, we leverage on the Bipartite Configuration Model (BiCM,[50]), i.e. the extension to bipartite networks of the entropy-based null-models reviewed in Ref.[51].
In a nutshell, the procedure is based on 3 main steps. First, we define an ensemble of (bipartite) networks, all having the same number of nodes per layer as in the real systems, but displaying all possible edge configurations, from the empty graph to the fully connected one. We then maximise the Shannon entropy associated to the ensemble, constraining some topological quantities of the network [52] (this approach replicate the approach of Jaynes for deriving Statistical Physics from Information theory[53]). In particular in the Bipartite Configuration Model, we constrain the average (over the ensemble) degree sequences for both layers to the values observed in the real system. Finally, in order to obtain the numerical value of the related Lagrangian multipliers, we maximize the Likelihood of the real system, i.e. the probability, according to our null-model, of getting the observed network[54].
Using the present procedure, we are getting a benchmark that is maximally random (due to the entropy maximization), but still tailored on the real system (due to fixing the degree sequences to one observed in the real network). In the following, we will first introduce briefly the formalism, then the Bipartite Configuration Model and, finally, its application for the validation of the co-occurrences.

Formalism

Let us call $\top$ and $\perp$ the two layers of the bipartite network and use Latin and Greek indices to indicate elements in the respective sets; we indicate with $N_\top$ and $N_\perp$, respectively, the dimension of the two layers. The biadjacency matrix $\mathbf{B}$ associated to the bipartite network is a $N_\top \times N_\perp$ matrix whose generic entry $b_{i\alpha}$ is either 1 or 0 if either there is or there is not a link connecting node i with node α. Therefore the degree sequences for both layers read $k_i = \sum_\alpha b_{i\alpha}\, \forall i \in N_\top$ and $h_\alpha = \sum_i b_{i\alpha}\, \forall \alpha \in N_\perp$.

The Bipartite Configuration Model

Let us call $\mathscr{G}_{\mathrm{Bi}}$ the bipartite networks' ensemble containing all possible graphs in which the dimension of the layers are respectively $N_\top$ and $N_\perp$. If $S = -\sum_{G_{\mathrm{Bi}} \in \mathscr{G}_{\mathrm{Bi}}} P(G_{\mathrm{Bi}}) \ln P(G_{\mathrm{Bi}})$ is the Shannon entropy, its maximization, constraining the average degree sequence on both layers, is equivalent to the maximization of S' defined as

$$S' = S + \sum_i \eta_i \Big[k_i^* - \sum_{G_{\mathrm{Bi}} \in \mathscr{G}_{\mathrm{Bi}}} P(G_{\mathrm{Bi}}) k_i(G_{\mathrm{Bi}})\Big] + \sum_\alpha \theta_\alpha \Big[h_\alpha^* - \sum_{G_{\mathrm{Bi}} \in \mathscr{G}_{\mathrm{Bi}}} P(G_{\mathrm{Bi}}) h_\alpha(G_{\mathrm{Bi}})\Big] + \zeta \Big[1 - \sum_{G_{\mathrm{Bi}} \in \mathscr{G}_{\mathrm{Bi}}} P(G_{\mathrm{Bi}})\Big],$$

where quantities with an asterisk $*$ represent the values observed in the real network and η_i, θ_α and ζ are the Lagrangian multipliers associated, respectively, to the degree sequence of layer $\top$, to the degree sequence of layer $\perp$ and to the normalization of the probability $P(G_{\mathrm{Bi}})$. The maximization of the S' returns the functional form of the probability per graph $P(G_{\mathrm{Bi}})$ in terms of the Lagrangian multipliers:

$$P(G_{\mathrm{Bi}}) = \prod_{i,\alpha} \frac{e^{-(\eta_i + \theta_\alpha) b_{i\alpha}(G_{\mathrm{Bi}})}}{1 + e^{-(\eta_i + \theta_\alpha)}}.$$

Therefore, $P(G_{\mathrm{Bi}})$ can be interpreted as the product of independent probability $p_{i\alpha} = \frac{e^{-(\eta_i + \theta_\alpha)}}{1 + e^{-(\eta_i + \theta_\alpha)}}$ of connecting node i with node α. In order to get the numerical values of Lagrangian multipliers η_i and θ_α, we can maximise the Likelihood associated to the observed network: it can be shown (see Ref.[54]) that it is equivalent to setting

$$\begin{cases} k_i^* = \langle k_i \rangle = \sum_\alpha p_{i\alpha} = \sum_\alpha \dfrac{e^{-(\eta_i + \theta_\alpha)}}{1 + e^{-(\eta_i + \theta_\alpha)}}; \\ h_\alpha^* = \langle h_\alpha \rangle = \sum_i p_{i\alpha} = \sum_i \dfrac{e^{-(\eta_i + \theta_\alpha)}}{1 + e^{-(\eta_i + \theta_\alpha)}}. \end{cases}$$

Validated projection of bipartite networks

Using the Configuration model defined in the previous subsection, it is possible to validate the projection of the bipartite network on one of the two layers. This procedure aims at stating the statistical significance of the co-occurrences observed in real systems. Consider, for instance a couple of nodes (i, j) belonging to $\top$ layer: the probability that they both link node $\alpha \in \perp$ is

$$P(V_\alpha^{ij} = b_{i\alpha} b_{j\alpha}) = p_{i\alpha} p_{j\alpha}, \tag{1}$$

where V_α^{ij} is the event "both i and j are linked to α" and $p_{i\alpha}$ is the probability of connecting nodes i and α. Using Eq. 1, we can calculate the probability that the total number of co-occurrences between i and j is exactly n as the sum of the contributions from all possible ways to choose n nodes in $\perp$ layer. If we call A_n this last quantity, the probability of observing $V_{ij} = \sum_\alpha V_\alpha^{ij} = n$ is

$$f_{PB}(V_{ij} = n) = \sum_{A_n} \Bigg[\prod_{\alpha \in A_n} p_{i\alpha} p_{j\alpha} \prod_{\alpha' \notin A_n} (1 - p_{i\alpha'} p_{j\alpha'}) \Bigg]. \tag{2}$$

Since, in principle, every $p_{i\alpha}$ is different, the distribution described by Eq. 2 is a sequence of Bernoulli events, each with different probability and equal to the one expressed in Eq. 1 and takes the name of Poisson-Binomial distribution[55].

Once we have the BiCM distribution for the number of co-occurrences between nodes i and j, we can then calculate the statistical significance of the observed V_{ij}^* via the p-value, i.e.

$$\text{p-value}\big(V_{ij}^*\big) = \sum_{V_{ij} \geq V_{ij}^*} f_{PB}\big(V_{ij}^*\big). \tag{3}$$

Iterating the calculation of Eq. 3 for every couple of nodes belonging to the $\top$ layer results in $\binom{N_\top}{2}$ p-values; to state the statistical significance of each of them, it is necessary to adopt a multiple hypothesis testing correction. In particular, the False Discovery Rate (FDR,[56]) is particularly effective since it permits to control the false positives rate.

The procedure described in this subsection was developed in Ref.[38]. For the actual implementation, we used `bicm` python module, available on `pypi` and described as part of `NEMtropy` package, in Ref.[57].

NACE	descritpion
A	Agriculture, forestry and fishing
B	Mining and quarrying
C	Manufacturing
D	Electricity, gas, steam and air conditioning supply
E	Water supply, sewerage, waste management and remediation activities
F	Construction
G	Wholesale and retail trade; repair of motor vehicles and motorcycles
H	Transportation and storage
I	Accommodation and food service activities
J	Information and communication
K	Financial and insurance activities
L	Real estate activities
M	Professional, scientific and technical activities
N	Administrative and support service activities
O	Public administration and defence; compulsory social security
P	Education
Q	Human health and social work activities
R	Arts, entertainment and recreation
S	Other service activities
T	Activities of households as employers; undifferentiated goods- and services-producing activities of households for own use
U	Activities of extra-territorial organisations and bodies

Table 3. NACE Rev.2, main division The description of the categories was taken from Ref.[58].

Relating hashtags to SDGs

To identify the SDG subjects that UK companies talk about, we used a threefold approach to have a good covering tailored to the available data set. As a list of SDGs related keywords suitable for online social network searches does not exist, we had to create one. Considering the many attempts to map academic articles' contributions to SDGs, we first started from the list of the University of Auckland (available here), which is used in business research[59] and consider the presence of words in our data set. This list mainly refers to keywords used in research papers in Elsevier's Scopus database, while in the present data set we are referring to Twitter's hashtags. In this sense, we gather multiple words in a single keyword, as it is customary for hashtags: for instance, "Child Labor Laws" became "childlaborlaws". Sometimes, the keywords were annotated under more than a single SDG: we disambiguated the multiple identifications manually, focusing on the main target of the various SDGs.

At this level, the identified SDG keywords represented less 0.52%, i.e. quite a limited coverage. Since we were not aware if the limited coverage of SDGs subjects was due to short attention to those arguments or not effective identification of SDGs hashtags, we manually annotated the hashtags among the 300 most frequent ones related to an SDG. The two authors independently performed the identification and agreed on 86.3% of the annotations; when they did not agree on the hashtag categorisation, they discussed each hashtag and finally attributed an SDG when they reached an agreement. Using this approach, we reached the 0.60% of all hashtags used by accounts in the validated projection of Fig. 3.

To further enlarge the SDG covering, we used a network approach. Using the bipartite representation of accounts and hashtags already used to obtain the validated projection on the account layer, we projected the network on the layer of hashtags, using the technique described in the subsection above and introduced in Ref.[38]. We remind the reader that in the validated projection, two nodes are present if they share a significant number of nearest neighbours in their bipartite representation. In this paper, two hashtags are connected in the validated projected network if they were both used by a significant number of different users. In this sense, a link in this network represents a non-trivial measure of similarity in how the various Twitter accounts use hashtags. Some might argue that we are interested in hashtags appearing in the same messages. This point is debatable: an account

interested in subjects related to, for example, SDG3 may use some of them related to different facets of SDG3 in different messages and focusing on hashtags used in the same messages will miss this information. Moreover, we avoid the risk of validating too many close hashtags since the procedure defined in[38] is highly restrictive. For instance, in the hashtag validated projected network, the link density is extremely low, i.e. 0.09%. Nevertheless, even in this case, we had to check the “automatic" annotation manually: in fact, the (validated) link between two hashtags may be due to a different reason than the adherence to the aims of the SDG: for instance, the keyword *#worldengineeringday* is connected to only the hashtag *#inwed*, i.e. the acronym for the International Women in Engineering Day, but it is not necessarily related to Gender Inequality (SDG5), as its neighbour. In a sense, the validated network represents a hint to follow to spot possible SDG hashtags related to the already labelled ones. Moreover, it permits spotting SDG hashtags specific to the current data set. It is the case, for instance, of hashtags of the various campaigns of the National Health Systems (all of them have been classified in SDG3) that are not general or the ones related to the Covid19 vaccination. We remark that in the validated network, the SDG hashtags represent a greater percentage (8%), signalling that there is collective attention of company accounts on the various subjects.
We focused on all hashtags that were not assigned an SDG that have at least an SDG hashtag among their neighbours since we expect that the former hashtags are related to the SDG of their neighbours. In order to be more restrictive, we focus on hashtags whose neighbours that were assigned an SDG represented more than half of their degree. Then they were assigned the most frequent SDG in their neighbours. The association was later manually checked to manage the case of ties in the SDGs in the neighbours, resulting in 146 newly annotated hashtags. The annotated hashtags now represent 0.68% of all hashtags in the data set.

References

1. Carroll, A. B. Corporate social responsibility: Evolution of a definitional construct. *Bus. & society* **38**, 268–295 (1999).
2. Dahlsrud, A. How corporate social responsibility is defined: an analysis of 37 definitions. *Corp. social responsibility environmental management* **15**, 1–13 (2008).
3. Snider, J., Hill, R. P. & Martin, D. Corporate social responsibility in the 21st century: A view from the world’s most successful firms. *J. Bus. ethics* **48**, 175–187 (2003).
4. McWilliams, A. & Siegel, D. Corporate social responsibility: A theory of the firm perspective. *Acad. management review* **26**, 117–127 (2001).
5. Greenwood, M. Stakeholder engagement: Beyond the myth of corporate responsibility. *J. Bus. ethics* **74**, 315–327 (2007).
6. Bonsón, E. & Ratkai, M. A set of metrics to assess stakeholder engagement and social legitimacy on a corporate facebook page. *Online Inf. Rev.* (2013).
7. Saxton, G. D., Gomez, L., Ngoh, Z., Lin, Y. P. & Dietrich, S. Do CSR Messages Resonate? Examining Public Reactions to Firms’ CSR Efforts on Social Media. *J. Bus. Ethics* **155**, 359–377, DOI: 10.1007/s10551-017-3464-z (2019).
8. Brown, N. & Deegan, C. The public disclosure of environmental performance information - a dual test of media agenda setting theory and legitimacy theory. *Account. Bus. Res.* **29**, DOI: 10.1080/00014788.1998.9729564 (1998).
9. Guthrie, J. & Parker, L. D. Corporate social reporting: A rebuttal of legitimacy theory. *Account. Bus. Res.* **19**, DOI: 10.1080/00014788.1989.9728863 (1989).
10. Manetti, G. & Bellucci, M. The use of social media for engaging stakeholders in sustainability reporting. *Accounting, Auditing Accountability J.* **29**, 985–1011, DOI: 10.1108/AAAJ-08-2014-1797 (2016).
11. Giacomini, D., Zola, P., Paredi, D. & Mazzoleni, M. Environmental disclosure and stakeholder engagement via social media: State of the art and potential in public utilities. *Corp. Soc. Responsib. Environ. Manag.* **27**, 1552–1564 (2020).
12. Scheyvens, R., Banks, G. & Hughes, E. The private sector and the sdgs: The need to move beyond ‘business as usual’. *Sustain. Dev.* **24**, 371–382 (2016).
13. Bebbington, J. & Unerman, J. Achieving the united nations sustainable development goals: an enabling role for accounting research. *Accounting, Auditing & Accountability J.* (2018).
14. Vildåsen, S. S. Corporate sustainability in practice: An exploratory study of the sustainable development goals (sdgs). *Bus. Strateg. Dev.* **1**, 256–264, DOI: 10.1002/bsd2.35 (2018). Cited By 8.
15. Elkington, J. Cannibals with forks. *The triple bottom line 21st century* **73** (1997).
16. D’Adamo, I., Gastaldi, M., Imbriani, C. & Morone, P. Assessing regional performance for the sustainable development goals in italy. *Sci. Reports 2021 11:1* **11**, 1–10, DOI: 10.1038/s41598-021-03635-8 (2021).

17. Khaled, R., Ali, H. & Mohamed, E. K. The sustainable development goals and corporate sustainability performance: Mapping, extent and determinants. *J. Clean. Prod.* **311**, DOI: 10.1016/j.jclepro.2021.127599 (2021).

18. Dalton, V. The challenge of engaging with and reporting against the sdgs for smes such as sydney theatre company. *J. Manag. Organ.* **26**, DOI: 10.1017/jmo.2020.23 (2020).

19. Ike, M., Donovan, J. D., Topple, C. & Masli, E. K. The process of selecting and prioritising corporate sustainability issues: Insights for achieving the sustainable development goals. *J. Clean. Prod.* **236**, DOI: 10.1016/j.jclepro.2019.117661 (2019).

20. Abd-Elrahman, A.-E. & Ahmed Kamal, J. Relational capital, service quality and organizational performance in the egyptian telecommunication sector. *Int. J. Emerg. Mark.* DOI: 10.1108/IJOEM-11-2019-0983 (2020). Cited By 2.

21. Tabares, S. Do hybrid organizations contribute to sustainable development goals? evidence from b corps in colombia. *J. Clean. Prod.* **280**, DOI: 10.1016/j.jclepro.2020.124615 (2021).

22. Silva, S. Corporate contributions to the sustainable development goals: An empirical analysis informed by legitimacy theory. *J. Clean. Prod.* **292**, DOI: 10.1016/j.jclepro.2021.125962 (2021).

23. García-Sánchez, I. M., Rodríguez-Ariza, L., Aibar-Guzmán, B. & Aibar-Guzmán, C. Do institutional investors drive corporate transparency regarding business contribution to the sustainable development goals? *Bus. Strateg. Environ.* **29**, DOI: 10.1002/bse.2485 (2020).

24. Lopez, B. Connecting business and sustainable development goals in spain. *Mark. Intell. Plan.* **38**, DOI: 10.1108/MIP-08-2018-0367 (2020).

25. Elalfy, A., Weber, O. & Geobey, S. The sustainable development goals (sdgs): a rising tide lifts all boats? global reporting implications in a post sdgs world. *J. Appl. Account. Res.* **22**, DOI: 10.1108/JAAR-06-2020-0116 (2020).

26. Franco-Riquelme, J. N. & Rubalcaba, L. Innovation and sdgs through social media analysis: Messages from fintech firms. *J. Open Innov. Technol. Mark. Complex. 2021, Vol. 7, Page 165* **7**, 165, DOI: 10.3390/JOITMC7030165 (2021).

27. Grover, P., Kar, A. K. & Ilavarasan, P. V. Impact of corporate social responsibility on reputation—insights from tweets on sustainable development goals by ceos. *Int. J. Inf. Manag.* **48**, 39–52, DOI: 10.1016/J.IJINFOMGT.2019.01.009 (2019).

28. Luca, F. D., Iaia, L., Mehmood, A. & Vrontis, D. Can social media improve stakeholder engagement and communication of sustainable development goals? a cross-country analysis. *Technol. Forecast. Soc. Chang.* **177**, DOI: 10.1016/j.techfore.2022.121525 (2022).

29. Zhou, Y., Li, X., Wang, X. & Yuen, K. F. Intelligent container shipping sustainability disclosure via stakeholder sentiment views on social media. *Mar. Policy* **135**, 104853, DOI: 10.1016/J.MARPOL.2021.104853 (2022).

30. Mehmood, A., Hajdini, J., Iaia, L., Luca, F. D. & Sakka, G. Stakeholder engagement and sdgs: the role of social media in the european context. *EuroMed J. Bus.* **ahead-of-print**, DOI: 10.1108/EMJB-11-2021-0173/FULL/XML (2022).

31. Campopiano, G. & De Massis, A. Corporate Social Responsibility Reporting: A Content Analysis in Family and Non-family Firms. *J. Bus. Ethics* **129**, 511–534, DOI: 10.1007/s10551-014-2174-z (2015).

32. Iaia, L. *et al.* Family businesses, corporate social responsibility, and websites: The strategies of Italian wine firms in talking to stakeholders. *Br. Food J.* **121**, 1442–1466, DOI: 10.1108/BFJ-07-2018-0445 (2019).

33. Hofstede, G. *Culture's consequences: International differences in work-related values*, vol. 5 (sage, 1984).

34. GRAY, S. J. Towards a theory of cultural influence on the development of accounting systems internationally. *Abacus* **24**, DOI: 10.1111/j.1467-6281.1988.tb00200.x (1988).

35. Mio, C., Panfilo, S. & Blundo, B. Sustainable development goals and the strategic role of business: A systematic literature review. *Bus. Strateg. Environ.* **29**, 3220–3245, DOI: 10.1002/bse.2568 (2020). Cited By 18.

36. de Villiers, C., Kuruppu, S. & Dissanayake, D. A (new) role for business – promoting the united nations' sustainable development goals through the internet-of-things and blockchain technology. *J. Bus. Res.* **131**, 598–609, DOI: 10.1016/j.jbusres.2020.11.066 (2021). Cited By 4.

37. Patuelli, A., Caldarelli, G., Lattanzi, N. & Saracco, F. Firms' challenges and social responsibilities during covid-19: A twitter analysis. *PLOS ONE* **16**, 1–30, DOI: 10.1371/journal.pone.0254748 (2021).

38. Saracco, F. *et al.* Inferring monopartite projections of bipartite networks: An entropy-based approach. *New J. Phys.* DOI: 10.1088/1367-2630/aa6b38 (2017). 1607.02481.

39. Blondel, V. D., Guillaume, J.-L., Lambiotte, R. & Lefebvre, E. Fast unfolding of communities in large networks. *J. Stat. Mech. Theory Exp.* **10008**, 6, DOI: 10.1088/1742-5468/2008/10/P10008 (2008). 0803.0476.

40. Pedersen, E. R. Modelling csr: How managers understand the responsibilities of business towards society. *J. Bus. Ethics* **91**, 155–166 (2010).

41. Gomez, L. M. & Vargas-Preciado, L. 140 characters for CSR communication: An exploration of Twitter engagement of Fortune companies. *Dev. Corp. Gov. Responsib.* **9**, 205–221, DOI: 10.1108/S2043-052320160000009009 (2016).

42. Etter, M. Reasons for low levels of interactivity. (Non-) interactive CSR communication in twitter. *Public Relations Rev.* **39**, 606–608, DOI: 10.1016/j.pubrev.2013.06.003 (2013).

43. Choudhury, P. R., Allen, R. & Endres, M. G. Machine learning for pattern discovery in management research. Tech. Rep. (2018).

44. Conover, M., Ratkiewicz, J. & Francisco, M. Political polarization on twitter. *Icwsm* DOI: 10.1021/ja202932e (2011).

45. Conover, M. D., Gonçalves, B., Ratkiewicz, J., Flammini, A. & Menczer, F. Predicting the political alignment of twitter users. In *Proc. - 2011 IEEE Int. Conf. Privacy, Secur. Risk Trust IEEE Int. Conf. Soc. Comput. PASSAT/SocialCom 2011*, DOI: 10.1109/PASSAT/SocialCom.2011.34 (2011).

46. Conover, M. D., Gonçalves, B., Flammini, A. & Menczer, F. Partisan asymmetries in online political activity. *EPJ Data Sci.* DOI: 10.1140/epjds6 (2012). 1205.1010.

47. Hofstede, G. *International Differences in Work-Related Values* (1984).

48. Pilăr, L. *et al.* Twitter analysis of global communication in the field of sustainability. *Sustain. (Switzerland)* **11**, DOI: 10.3390/su11246958 (2019).

49. Doan, A. *Magellan project, py_stringmatching module* ((accessed on 17/11/2020)).

50. Saracco, F., Di Clemente, R., Gabrielli, A. & Squartini, T. Randomizing bipartite networks: the case of the World Trade Web. *Sci. Rep.* **5**, 10595, DOI: 10.1038/srep10595 (2015).

51. Cimini, G. *et al.* The Statistical Physics of Real-World Networks. *Nat. Rev. Phys.* **1**, 58–71, DOI: 10.1038/s42254-018-0002-6 (2018). 1810.05095.

52. Park, J. & Newman, M. E. J. Statistical mechanics of networks. *Phys. Rev. E* **70**, 66117, DOI: 10.1103/PhysRevE.70.066117 (2004).

53. Jaynes, E. Information theory and statistical mechanics. *The Phys. Rev.* **106**, 181–218, DOI: 10.1103/PhysRev.106.620 (1957).

54. Garlaschelli, D. & Loffredo, M. I. Maximum likelihood: Extracting unbiased information from complex networks. *Phys. Rev. E - Stat. Nonlinear, Soft Matter Phys.* **78**, 1–5, DOI: 10.1103/PhysRevE.78.015101 (2008).

55. Hong, Y. On computing the distribution function for the Poisson binomial distribution. *Comput. Stat. Data Anal.* **59**, 41–51, DOI: 10.1016/j.csda.2012.10.006 (2013).

56. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B* **57**, 289–300 (1995).

57. Vallarano, N. *et al.* Fast and scalable likelihood maximization for exponential random graph models (2021). 2101.12625.

58. Eurostat. Nace rev. 2 – statistical classification of economic activities in the european community. *Off. for Off. Publ. Eur. Communities* p. 230 (2008).

59. Sinkovics, N., Vieira, L. M. & van Tulder, R. Working toward the sustainable development goals in earnest – critical international business perspectives on designing and implementing better interventions. *Critical Perspectives on Int. Bus.* **18**, 445–456, DOI: 10.1108/CPOIB-05-2022-0059/FULL/XML (2022).

Author contributions statement

AP: Conceptualization, Data curation, Writing – original draft, Writing – review & editing FS: Data curation, Formal analysis, Methodology, Visualization, Writing – original draft, Writing – review & editing

Data Availability

Twitter ID data can be downloaded from the following link. The list of hashtags associated to the various SDGs can be download from the following link. The firms' ID and financial data that support the findings of this study are available from AIDA (Bureau Van Dijk). Restrictions apply to the availability of these data, which were used under license for this study.

Competing interests

The authors declare no competing interests.

0		2		5		6		1	
hashtag	frequency	hashtag	frequency	hashtag	frequency	hashtag	frequency	hashtag	frequency
cloud	151	cop	270	covid	61	covid	147	earthday	63
technology	140	sustainability	259	sustainability	51	cop	101	blackhistorymonth	53
digital	135	netzero	248	esg	50	budget	97	pridemonth	53
data	135	climatechange	193	cop	48	webinar	94	internationalwomensday	50
digitaltransformation	134	sustainable	187	china	40	esg	94	covid	50
cybersecurity	120	earthday	182	inflation	39	internationalwomensday	91	iwd	49
innovation	111	internationalwomensday	170	supplychain	35	brexit	88	sustainability	46
covid	109	innovation	148	sustainable	34	diversity	80	pride	42
tech	106	energy	144	innovation	34	investment	79	choosetochallenge	38
webinar	100	worldenvironmentday	135	climatechange	32	podcast	78	womenshistorymonth	36

Table 4. **Frequency of top 10 most frequent hashtags in the subcommunities with more than 50 nodes in the Cyan community.**

1		0		6		2	
hashtag	frequency	hashtag	frequency	hashtag	frequency	hashtag	frequency
internationalwomensday	88	christmas	97	carehome	42	mentalhealthawarenessweek	85
mentalhealthawarenessweek	75	valentinesday	88	care	41	internationalwomensday	83
blackhistorymonth	58	halloween	81	christmas	40	covid	63
pridemonth	52	internationalwomensday	80	internationalwomensday	39	mentalhealth	61
iwd	44	mondaymotivation	76	covid	37	worldmentalhealthday	58
earthday	43	fridayfeeling	72	socialcare	37	volunteersweek	56
worldmentalhealthday	43	mothersday	71	dementia	29	christmas	45
choosetochallenge	36	win	69	halloween	28	blackhistorymonth	45
pride	35	easter	67	valentinesday	26	wellbeing	38
diwali	33	bankholiday	67	mothersday	25	worldbookday	38
cop	33			mentalhealthawarenessweek	25		
				internationalnursesday	25		

Table 5. **Frequency of top 10 hashtags in the subcommunities with more than 50 nodes in Orange-red community, 1/2.**

4		4		3	
hashtag	frequency	hashtag	frequency	hashtag	frequency
covid	43	internationalwomensday	85	christmas	34
learningdisability	41	cop	85	volunteers	30
socialcare	35	covid	77	volunteersweek	29
mentalhealthawarenessweek	35	blackhistorymonth	72	charity	28
autism	35	mentalhealthawarenessweek	71	givingtuesday	25
internationalwomensday	32	iwd	66	hospicecareweek	24
learningdisabilities	30	earthday	55	internationalwomensday	23
mentalhealth	28	mentalhealth	50	internationalnursesday	22
learningdisabilityweek	25	unimentalhealthday	50	londonmarathon	22
autistic	25	diwali	50	fundraising	20

Table 6. **Frequency of top 10 hashtags in the subcommunities with more than 50 nodes in Orange-red community, 2/2.**

2		0		3	
hashtag	frequency	hashtag	frequency	hashtag	frequency
apprenticeship	125	worldbookday	147	apprenticeship	92
mentalhealthawarenessweek	103	antibullyingweek	111	mentalhealthawarenessweek	74
naw	99	mentalhealthawarenessweek	109	apprentice	72
internationalwomensday	97	internationalwomensday	85	internationalwomensday	67
apprentice	83	childrensmentalhealthweek	85	naw	67
iwd	64	christmas	80	engineering	60
nationalapprenticeshipweek	61	remembranceday	77	education	53
mentalhealth	57	backtoschool	71	construction	52
choosetochallenge	55	onekindword	70	collegesweek	51
cop	54	saferinternetday	67	careers	51
		science	67		
		wellbeing	67		

Table 7. **Frequency of top 10 hashtags in the subcommunities with more than 50 nodes in Yellow community.**

3		0		2		1	
hashtag	frequency	hashtag	frequency	hashtag	frequency	hashtag	frequency
covid	128	mentalhealthawarenessweek	47	covid	106	covid	55
nhs	123	covid	43	covidvaccine	81	mentalhealthawarenessweek	40
covidvaccine	111	nhs	36	mentalhealthawarenessweek	79	internationalwomensday	38
internationalnursesday	103	mentalhealth	35	volunteersweek	67	internationalnursesday	38
volunteersweek	89	internationalwomensday	34	internationalwomensday	62	nhs	36
ahpsday	86	covidvaccine	34	euro2020	58	covidvaccine	33
blackhistorymonth	84	worldmentalhealthday	33	worldmentalhealthday	57	worldmentalhealthday	32
internationalwomensday	83	volunteersweek	32	nhs	55	timetotalk	26
mentalhealthawarenessweek	81	blackhistorymonth	32	grabajab	54	vaccine	26
nhsbirthday	80	wellbeing	32	everymindmatters	49	mentalhealth	25
				mentalhealth	49	grabajab	25

Table 8. **Frequency of top 10 hashtags in the subcommunities with more than 50 nodes in Orchid community.**