



# A Twitter-Based Study of the European Internet of Things

Funda Ustek-Spilda<sup>1</sup> · Davide Vega<sup>2</sup> · Matteo Magnani<sup>2</sup> · Luca Rossi<sup>3</sup> · Irina Shklovski<sup>4</sup> ·  
Sebastian Lehuede<sup>5</sup> · Alison Powell<sup>5</sup>

© The Author(s) 2020

## Abstract

We present a methodology integrating social media data, data from qualitative research and network analysis. Qualitative insights gained from ethnographic fieldwork are used to collect and annotate social network data, and social media data is used as part of the ethnography to identify relevant actors and topics. The methodology is presented in the context of an analysis of the Internet of Things in the European context.

**Keywords** Internet of Things · Twitter · Ethnography · Data science · Methodology · Triangulation

## 1 Introduction

An increasing number of devices and services intended for everyday uses are designed with networked and sensing capabilities. Smart devices are populating smart buildings, which in turn comprise smart cities, navigated by smart transport. Smartness has become synonymous with the idea of the Internet of Things (IoT) — sensing technologies that collect and transmit data for algorithmic processing and machine learning. The proliferation of IoT represents unprecedented potential for integration of connected experience (via sensors and networked data) into everyday life — thus providing a way to frame the notions of offline and online as neither blurring or in opposition, but as a coherent daily experience that seamlessly blends the digital with the non-digital and online with the offline. Yet the imaginaries of IoT tend to oscillate between optimistic forecasting of its potential and doomsday pronouncements of surveillance and control. On the one hand, ubiquitous connectivity is much celebrated where all technologies are seamlessly connected to one another which increases the efficiency and productivity of services, expanding and enhancing human experience.

Assistance to the elderly, smart home devices enabling greater efficiency and control of energy consumption or smart city and industry applications for greater productivity, better use of limited resources and a more inclusive distribution of services are all but some examples of this great potential of IoT. On the other hand, anxieties about the potential consequences of pervasive connectivity are abundant, with scenarios à la “Black Mirror” with ultimate surveillance and control often framing the debates about IoT, along with concerns about security, privacy and trust.

The VIRT-EU project<sup>1</sup> had set out to intervene in the design of IoT technologies in Europe by considering the ethical underpinnings of technology development processes, thus potentially helping us to benefit from these technologies while avoiding their pitfalls. A starting point of the project was that the term “IoT” does not refer to a specific technology, but to a type of digitization, where physical objects are made digital and connected to the Internet. When these products are commercialised they are not always branded as IoT, and therefore much of what is commercially available as IoT solutions are not always referred to as IoT. This is one of the main reasons why IoT is still considered as a combination of hardware, software and networks, even though it can as well be software or middleware alone. In this sense, autonomous cars, smart phones, smart home devices or smart watches can all be considered IoT; so can sensor detection software apps or hardware that connects a variety of sensors. These digital technologies are then connected to vast platforms where data are stored, processed and put to use. Given the diffuse nature of IoT as a

This work was partially supported by the European Union’s Horizon 2020 research and innovation programme under grant agreements No. 727040 (Virt-EU).

✉ Matteo Magnani  
matteo.magnani@it.uu.se

Extended author information available on the last page of the article.

<sup>1</sup><https://virteuproject.eu>

concept, it behooves us to ask, what might we mean by IoT. The perils and promises of IoT do not exist in a vacuum, but are situated within the communities of designers, developers and entrepreneurs that imagine it as well as media reporting, policies and regulations that govern it (Ustek-Spilda et al. 2019). In this article we take up IoT in Europe as our case study, as technology developers in Europe are governed by EU-level laws and regulations (e.g. GDPR), even when specific social milieus of technology development might differ across member states.

## 1.1 Contributions

This article makes two main contributions. First, we describe what IoT looks like in the European context from a social media point of view. We do so through a mixed-method study design. Through this study we ask (i) what kinds of actors are central to IoT-related discussions in Europe, (ii) what kinds of geographical clusters emerge, (iii) what are the specific topics that emerge within IoT and finally, (iv) what are the matters of concern discussed by the technology developers, designers and entrepreneurs of IoT (from now on called actors) on Twitter.

Second, in order to answer these questions, we present an innovative methodology we developed that integrates social media data, data from qualitative research and network analysis. More precisely we demonstrate how qualitative insights gained from ethnographic fieldwork can be used to identify key users of IoT in Europe and to gather network data. At the same time, Twitter data is used as part of the ethnography to identify actors to interview/observe and topics under discussion. We validate the proposed method and we suggest it as a possible strategy to map large heterogeneous topics on social media about social and technological issues that are gaining attention.

IoT, as we will elaborate further later in the text, is a complex topic that encompasses myriad technological

solutions and application domains. The proposed combination of social media data and ethnographic observations is not enough to explore all its shades. Nevertheless we claim that it is possible to provide an overview of the phenomenon at a European level and that this constitutes a substantial improvement over mapping strategies based on single-method approaches.

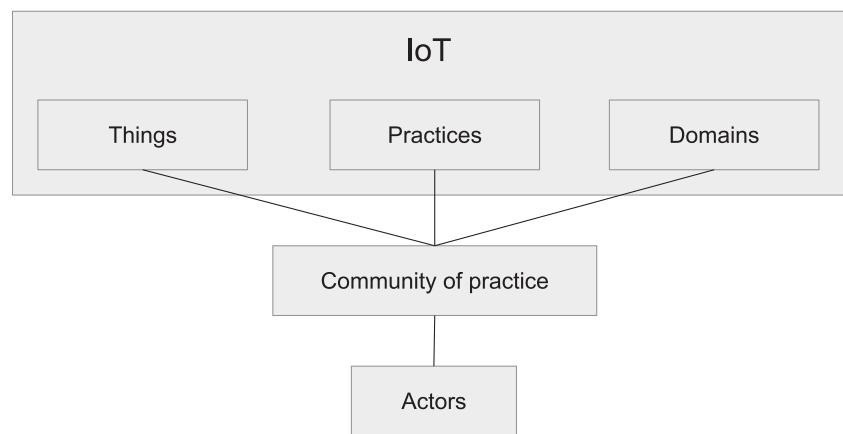
## 1.2 Outline

The remainder of the article is organized as follows. First, we present the theoretical framework that guided our analysis. Then, in Section 3 we describe the data collection methods we used, for reaching what we call the “consolidated dataset” where we merged data obtained from qualitative fieldwork and quantitative data retrieved from Twitter. In Section 4 we present the results of our empirical analysis, and in Section 5 we summarize the main results of the article and also discuss some limitations of alternative analysis methods based on Twitter data.

## 2 Theoretical Framework

In the upper part of Fig. 1 we indicated the field under study, that is, IoT. Based on the fieldwork we have carried out in IoT spaces in Europe and preliminary analysis of online discussions on IoT, we show that the field emerges as a socio-technical assemblage (Kitchin 2017) of different things, issues and approaches. More specifically, we identified IoT as an assemblage of 1) things (connected devices and technologies), 2) practices (i.e. practices of building connections between things and technologies), 3) application contexts (e.g. smart cities, e-health, wearables and so on) and 4) IoT practitioners (i.e. developers, designers, entrepreneurs, etc.). Such a positioning of IoT builds on the assumption that technologies are not neutral, and external

**Fig. 1** Theoretical framework



to the socio (e.g. human) and technical (e.g. things, technologies, network) systems they are part of Bijker et al. (1987) and Latour (2005). Hence, the significance of the results presented in this article is based on the notion that IoT is enacted by the actors and communities of practice that consider themselves as part of the IoT ecosystem and the relations and practices that sustain this assemblage.

Looking at IoT through the framework of method assemblages opens up important methodological possibilities to study it. Here, we pay particular attention to “matters of concern” (Latour 2004; de La Bellacasa 2011) identified and articulated by IoT actors. In contrast to “matters of fact”, objective realities that exist out there, Latour proposes “matters of concern” paying attention to the *gatherings* of things, issues, conditions, ingredients, nonhumans as well as humans (Latour, 2004, 246). In other words, matters of concern draws our attention to the relationships that hold matters of fact together, and the assemblages formed as a result based on agreements, disagreements and the associated consequences of the two. As an example, issues of sustainability or privacy emerge as matters of concern, only if they are discussed as such by actors in the IoT ecosystem and specific ‘things’, ‘concerns’ and risks are attached to these discussions. Similarly, ethics emerges as a matter of concern, only when IoT actors identify particular ethical challenges in their technology development and articulate how these challenges impact their thinking, design and development of IoT technologies.

However, concerns cannot be fully understood if they are taken individually, and the relationships between them and their human and non-human actors are ignored. Through these relationships, they come to be sorted, re-defined and ordered. Powell (2018) calls this “moral ordering” as it involves decision-making between different sets of concerns, prioritising some whereas ignoring others. For example, even when privacy might emerge as a matter of concern, it is in relation to other concerns — viability of the product, sustainability of the business, security of the device or the storage system and so on, that it may become a priority for developers or not.

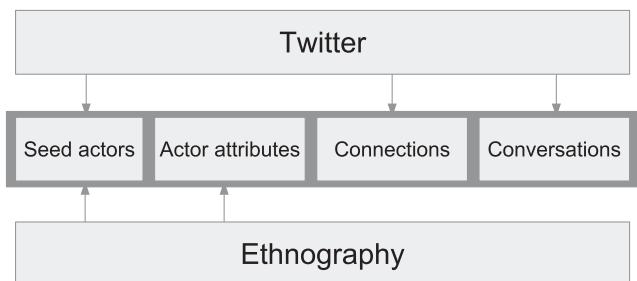
Against this background, in this paper, we pay attention to not only individual concerns [and values] expressed by actors in the IoT ecosystem, but also to the relations between them and their ordering. This implies that concerns previously not identified as “ethical concerns” might emerge as such, and other concerns might emerge to be less exemplary of the issues that are discussed and negotiated in the ecosystem.

### 3 Data Collection

Our framework and approach for data analysis were based on an iterative process. Our ethnographic data indicated

that some values, concerns and application areas of IoT were shared by the IoT actors, regardless of their location; whereas some values, concerns and technological approaches were location-specific. For instance, we found that while security and privacy were shared as concerns (and values) by a majority of IoT actors in Europe, how sustainability was understood changed from location to location. While in London, sustainability was used to refer to business sustainability of startups in addition to environmental sustainability of IoT technologies, in Amsterdam for instance, this concern for business sustainability was not raised often. Similarly, we observed that whereas in western Europe (e.g. London, Amsterdam, Copenhagen), IoT technologies included both hardware and software, in eastern Europe (e.g. Belgrade, Bled), technologies were more based on software. Based on this insight, we started the data collection process for the online social network analysis of IoT actors. In this round of data collection, we sought to both evaluate our qualitative findings, understand how widely they were applicable beyond the contexts we studied in Europe, and map out how discussions about IoT — along with particular concerns — evolved over time and space in Europe.

Figure 2 gives an overview of the data used in this work, which was collected using a methodology mixing qualitative and quantitative methods and modelled using different methods from network science, including a new approach to generate topical conversational networks from sparse Twitter data. The complexity of the questions to be answered required more nuanced and complete data than what was publicly available on Twitter or could be extracted using the Twitter API (Application Programming Interface). At the same time, field work performed at a limited number of sites carried the risk of overlooking important individuals in the IoT social network. Our mixed method approach offers a way to redress both of these limitations. Although Fig. 2 shows Twitter and Ethnography as the two main ways to gather data, Twitter is in fact a combination of multiple data sources as it includes a) Tweets written using specific hashtags, b) Tweets written by a selected number of qualitatively selected users, regardless of the hashtag, c)



**Fig. 2** Top-level data sources (Twitter, Ethnography) and main data entities

**Table 1** Specific data sources, size (order of magnitude), and their role

Source	Size	Role
Twitter IoT events (23 hashtags)	$10 - 10^4$	Selection of seed actors
Offline observations	$10^2$	Selection of seed actors, attributes
Interviews	$10^2$	Assign actor attributes
Following/follower network	$10^2 - 10^5$	Assortativity analysis
Tweets by seed actors	$10^4$	Validation, Topical analysis

following/follower networks of the same group of users. Similarly, ethnography includes several data collection activities (interviews, observations, co-creation workshops) all of which generate different kinds of qualitative data. A summary of the data collection methods used in the paper is indicated in Table 1.

A common issue in social network analysis is the so-called boundary selection problem: the choice of which actors to include in the study can have a strong influence on the results. In the type of study presented in this article this problem is particularly complex because of using Twitter as a source of data, because of the vague definition of the domain (IoT) and because of the focus on the European IoT. For these reasons it has not been possible to extract particularly significant knowledge from the large datasets. In particular, we collected tweets containing the hashtag #IoT for over one year, but the data contained a large number of unrelated tweets automatically produced by automated accounts, did not contain data not explicitly annotated with IoT-relevant hashtags despite being relevant, and contained a lot of information whose geographical location was not available. For this reason, this large dataset was not ultimately used for the fine-grained analysis and we decided to rely on other sources.

### 3.1 Seed Actors

The first part of the data described in the following was a set of European actors who were considered of particular relevance in the IoT field. The rationale behind this choice was that these actors can be used as seeds to capture the discourse about the IoT in Europe. First, we expected these actors to contribute to the conversation on IoT topics, which in the context of this study is observed through the tweets posted by them on this subject. These tweets could then be used to extract the main topics under discussion, that we use as a proxy to identify matters of concern. Second, we expected other actors interested in the IoT to mention, retweet and follow our seed users, allowing us to identify a larger portion of the IoT Twitter space. Both these hypotheses are tested below (Section 3.5).

In order to identify these seed actors, we used both qualitative and quantitative methods. We first attended

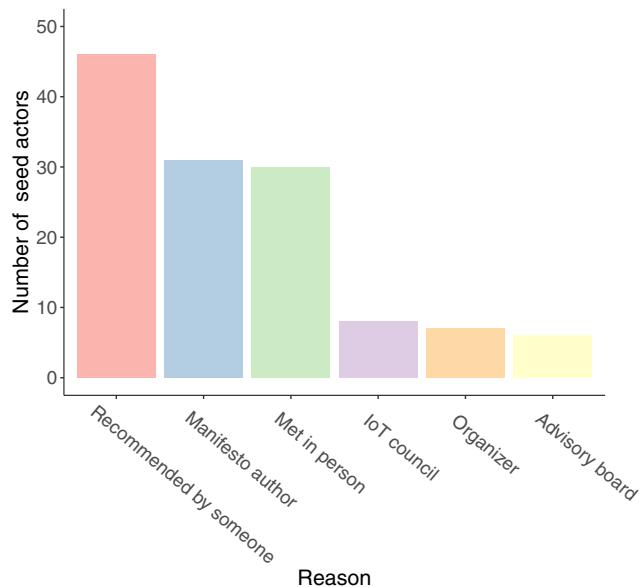
IoT-related events across eleven European countries<sup>2</sup>, and also conducted an analysis of responsible technology and IoT manifestos produced by designers and developers in Europe (Fritsch et al. 2018). Thanks to these qualitative field explorations, we identified a set of key European actors in the IoT space. This was complemented by a quantitative analysis of selected IoT events on Twitter to identify the accounts that were mentioned and retweeted more often. These accounts were manually inspected afterwards to identify additional relevant actors, through following the trail of online conversations, following hashtags, company profiles and live feeds of IoT industry conferences. As a result of these efforts, we compiled the names of 108 actors, of which 103 of them had valid Twitter accounts, when we conducted our study in mid-late 2018. We should, however, add that, the levels of online engagement among the key IoT actors differed vastly; whereas some actors were active both online and offline, some were not very active online (See Section 3.4).

Figure 3 summarizes the six main criteria used to include a seed actor in our study. These reasons are not mutually exclusive, as many actors have been selected because they fit multiple criteria. For instance, most of the actors we met during this study were interviewed because they were recommended by some other relevant actor we had previously met, were frequently mentioned in discussions of IoT-related issues or were members of key groups and organizations in this field.

### 3.2 Actor Attributes

The data about the seed actors was subsequently enriched using multiple sources. The first type of information added to the data consisted in a set of attribute values. Differently from the social network data described in the next two sections, these attributes stored data about each individual actor independently of the others. The data was collected manually, based both on information obtained from ethnographic fieldwork and interviews, but also desk research, including searches on the Internet, scanning names

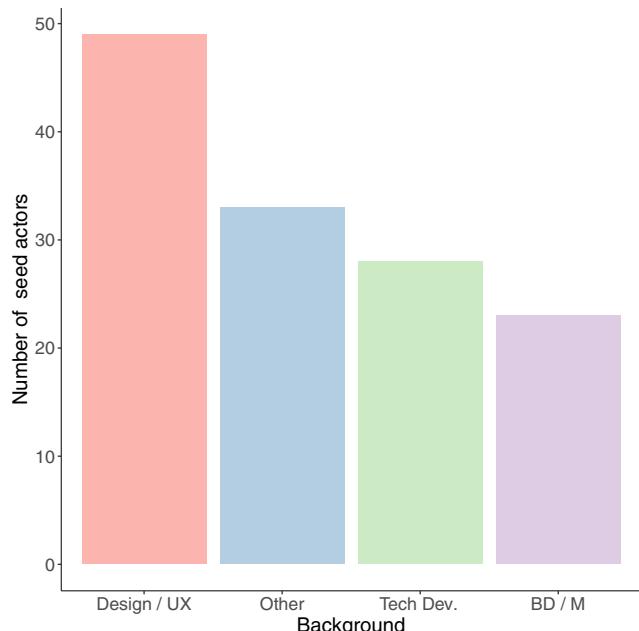
<sup>2</sup>London, Geneva, Lyon, Torino, Copenhagen, Belgrade, Bled, Malmö, Berlin, Barcelona and Amsterdam.



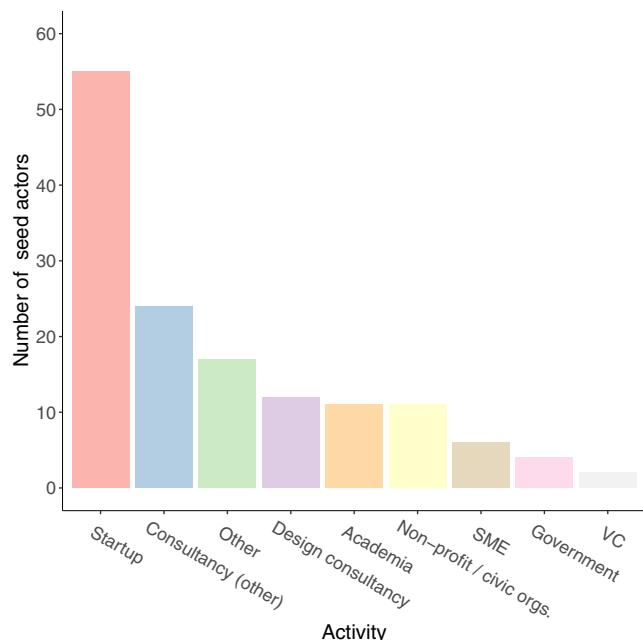
**Fig. 3** Main criteria for the selection of the seed actors

for keynotes on major IoT events in Europe and referral and mentions of some of the key actors in our list. We then collected the social media account names of these actors.

This information includes social media handles, geographical location, personal background (See Fig. 4), type of activity within the IoT space (See Fig. 5), and if the actors had shown interest in discussions about ethics in IoT. From now on we will refer to this enriched set of actors as the



**Fig. 4** Background distribution of the seed actors. BD/M: business development and/or marketing, D/UX: Design and/or UX, TD: Tech Developer (mainly engineers or computer scientists)



**Fig. 5** Activity distribution of the seed actors. SME: bigger than startups / smaller than corporations

*consolidated dataset.* Please note that due to the VIRT-EU project orientation towards IoT innovation by start-ups and small-medium enterprise actors, the consolidated dataset is skewed in that direction and away from those involved with large corporate actors in the field of IoT.

When collecting data on the seed actors, in addition to the research strategies described above, we also created a list of questions and attributes for the seed actors that would help us identify, categorise and organise their concerns and approaches in IoT. For instance, we have created a variable in the consolidated dataset for measuring “vocalness on issues of ethics” to describe how active the actors were in discussions on ethics. Some of these actors we have found the opportunity to interview, and some we have observed in industry meetings, meetups and other IoT events. Some of them, however, were outside of the geographical scope of our ethnographic fieldwork, so we categorised them only based on their level of engagement on issues of ethics online. Here, we did not consider “ethics” as a keyword only and our analysis was not a mere collection of how many times an actor used the word in their communication with others. Instead, we analysed qualitatively the types of topics, issues and events the actors engaged with and if they talked about the values which were identified as key in our ethnographic fieldwork, such as responsibility, security, privacy, trust, autonomy, accountability and non-discrimination. We classified as “high” actors who were actively discussing these topics, sharing content on Twitter expressing their concern on these issues and retweeted messages from actors which expressed ethical concerns. We

gave the “middle” score to actors who at least occasionally discussed or shared content that expressed ethical concerns. The low category was assigned to individuals who rarely, if at all, engage with issues on ethics. We need to add, however, that we did not have the same level of detail on every individual as we were unable to interview and/or observe in person all the actors. In addition, as mentioned above, not all of the key actors we identified through our fieldwork and through Twitter were actively discussing topics related to ethics.

### 3.3 Following/Follower Connections

Using the Twitter screen names of the seed actors, we connected to the public Twitter API to retrieve their full list of followers and friends. This information was then used to build a directed network of the Twitter space *surrounding* our initial set of qualitatively selected seeds. We defined three derived networks:

1. The *full network* containing all the followers and all the followees of the 103 initial Twitter users from the consolidated dataset.
2. The *reduced network* containing all the initial users from the consolidated dataset and the followers or followees connected with at least two of the initial users.
3. The *consolidated network* containing only the users from the consolidated dataset and the connections among them.

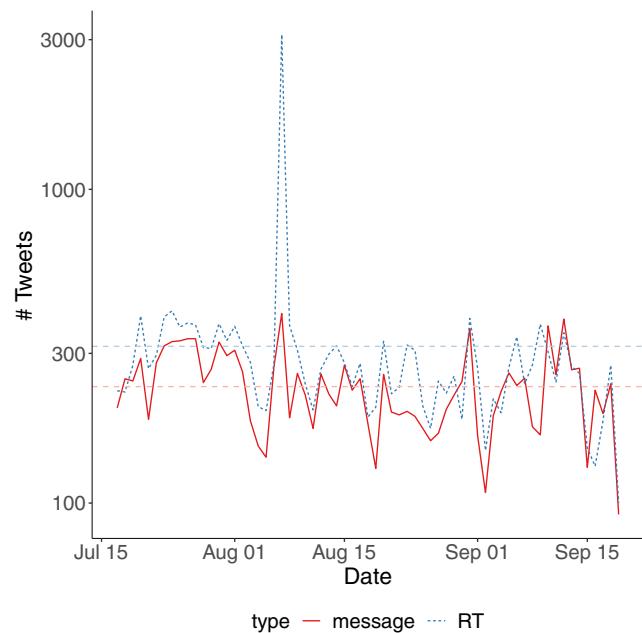
Table 2 shows how the networks are remarkably different both in terms of size as well as in terms of structure and basic topological characteristics — as expected given our sampling method expanding a core of seed users. A first consideration that can be done is that each member of the consolidated dataset has a Twitter network that only partially overlaps with the ones of the other members. The *reduced network*, that includes only the users who are connected with at least two members of the consolidated dataset is one sixth of the full network. This suggests that each member of the consolidated dataset has her own Twitter audience.

### 3.4 Topical Conversations

Finally, we used the public Twitter API to retrieve interaction information. As interactions happen by tweeting,

**Table 2** Basic statistics, three Twitter networks generated from the consolidated dataset

Network	Nodes	Edges	Density	Avg. Degree	Clust.coeff
Full network	319445	424781	0.000004	2.66	0.0001
Reduced network	51580	156916	0.00006	6.08	0.0014
Consolidated network	103	1360	0.130	26.4	0.385



**Fig. 6** Number of interactions retrieved per day. The continuous red line indicates the number of tweets/day, the dashed blue line indicates the number of retweets/day. Horizontal lines mark the average number of tweets and retweets per day

we can exploit the text of the tweets to observe whether the interactions concern specific topics. We collected the latest tweets and retweets produced by our seed actors during a period of 2 months (from 2018-07-17 to 2018-09-19). Figure 6 summarizes the number of interactions per day within the consolidated set, which contains an average of 235 messages/day and 315.9 retweets/day. These tweets correspond to the entry *Tweets by seed actors* in Table 1.

While many of the interactions we captured were relevant for our study, we also captured noisy conversations about unrelated topics. One clear example occurred on August, 7th when one of the seed actors posted a tweet referring to general interest results of a research study in climate warming that was retweeted and mentioned by more than 2,000 users. This section describes how we used the hashtags of the messages to remove such irrelevant (for this project) information.

Some of these tweets contained hashtags, some of which we used as an indication of topics of interest based on manual inspection. More than 400 hashtags were

selected and classified, in order of frequency. These selected hashtags were also grouped into a hierarchy of larger topics.

Table 3 summarizes the number of interactions (tweets and retweets) within the consolidated set that are posted by the seed actors vs. the number of interactions posted by the extended network. As we can observe, the seed actors are significantly more active than the average actors over all the consolidated set and they make a significantly higher frequency use of hashtags.

An important category for this study was ethics, conceptualized broadly and containing a wide range of topics/hashtags that are important to the development of responsible IoT technologies. As a category, ethics is split into sub-categories, including security, privacy, womenintech, sustainability, and openness. The following are selected examples of hashtags in different sub-categories of Ethics:

Ethics; security; #cybersecurity  
 Ethics; security; #iotsecurity  
 Ethics; privacy; #gdpr  
 Ethics; privacy; #dataprotection  
 Ethics; womenintech; #womenintech  
 Ethics; womenintech; #femtech  
 Ethics; womenintech; #girlsinstem  
 Ethics; trust; #trust  
 Ethics; ethics; #ethicaltechnology  
 Ethics; ethics; #responsibletech  
 Ethics; sustainability; #sustainability  
 Ethics; sustainability; #zerophone  
 Ethics; open; #opensource  
 Ethics; open; #opendata

As explained in our theoretical framework, ethical concerns cannot be studied in isolation. Therefore, we identified a number of additional categories, including business (#investment, #crowdfunding, ...), regulation (#iotmark, #gdpr, ...), concerns (#plasticgarbage, #digitaltransition, #cyberthreats, ...), application areas (#smartwatches, #smarthomes, ...), hardware (#raspberrypi, #microbit, ...), etc.

We build our taxonomy using an iterative process. First, we organised the hashtags based on the number of times they were used (including retweets). Second, we tagged

these hashtags along their socio-technical aspects, such as whether they referred to software or hardware, described an application area of the technology or signalled a particular ethical value. Third, we organized and grouped all of the tags into higher level categories so as to be able to minimise the number of categories used without sacrificing on their meaning or context. Lastly, we conducted a peer review process among the researchers to finalise the consolidation of the categories.

At the individual actor level, the identified topics allowed us to annotate the authors of the tweets with the sets of matters of concern they decided to discuss online. In addition, the hashtags were also used to produce a topical social network. A topical network is a multilayer social network (Magnani and Rossi 2011; Kivelä et al. 2014; Dickison et al. 2016) where each layer contains ties between actors corresponding to a specific topic (Vega and Magnani 2018).

To the best of our knowledge, previous studies of interaction networks on Twitter have built networks based on observations such as the one in Fig. 7a. When user1 posts a tweet mentioning user2, this is modeled as an edge from user1 to user2. However, mentions in themselves have no associated context and may lead to very sparse networks.

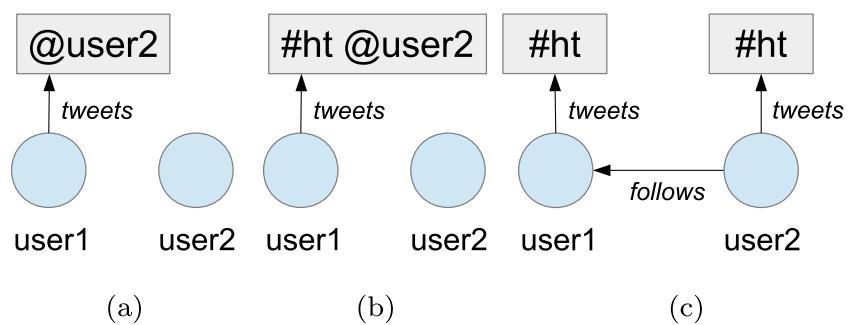
Here we extend this basic way of building networks from tweets in two ways. First, as shown in Fig. 7b and suggested by Vega and Magnani (2018), we build a separate graph (also called a layer) for different topics. If we assume that the hashtag #ht indicates a specific topic, then user1 will be connected to user2 in the layer corresponding to that topic. Then, as suggested by Hanteer et al. (2018), the usage of a topical hashtag positions the tweet inside a specific conversation, and implicitly targets other users also interested in the same topic, as shown in Fig. 7c. If user1 tweets about topic #ht and user2, who is following the updates from user1, is also interested in topic #ht (in our case, she has tweeted about it), then we also consider a connection between user1 and user2 in the corresponding topical layer.

The identification of connections not explicitly provided by the Twitter API but implicitly present in the data is a fundamental step to build networks that are not too sparse to be studied. Using only interactions of the type described

**Table 3** Summary of interactions within the consolidated dataset

Measure	Seed			Other		
	avg.	max.	min.	avg.	max.	min.
interactions / person	Tweet	126	1,288	1	1.92	50
	RT	87.3	1,377	11.64	1.64	103
hashtags / interaction	Tweet	0.325	7	0	0.051	5
	RT	0.704	12	0	0.447	12

**Fig. 7** **a** A traditional way of observing interactions on Twitter, and **b, c** two tweeting patterns resulting in an edge about topic #ht between user1 and user2 in our study



in Fig. 7a we would register 216 interactions about top-level matters of concern, some of which between the same pairs of actors. The patterns in Fig. 7b allow us to identify 852 additional interactions.

The final topical network is summarized in Table 4 and in Fig. 8. Notice that some of the interactions happen between the same pairs of actors, which explains why there are less edges than the number of interactions mentioned above.

### 3.5 Validation

The seed actors were validated to check respectively if the individuals included in the collection are indeed actively engaged in IoT discussions on Twitter and if key Twitter users were missing.

To check whether the identified actors address IoT-related topics as part of their online presence we have inspected the most frequent hashtags in their tweets, indicated in Table 5. From this list we can count the

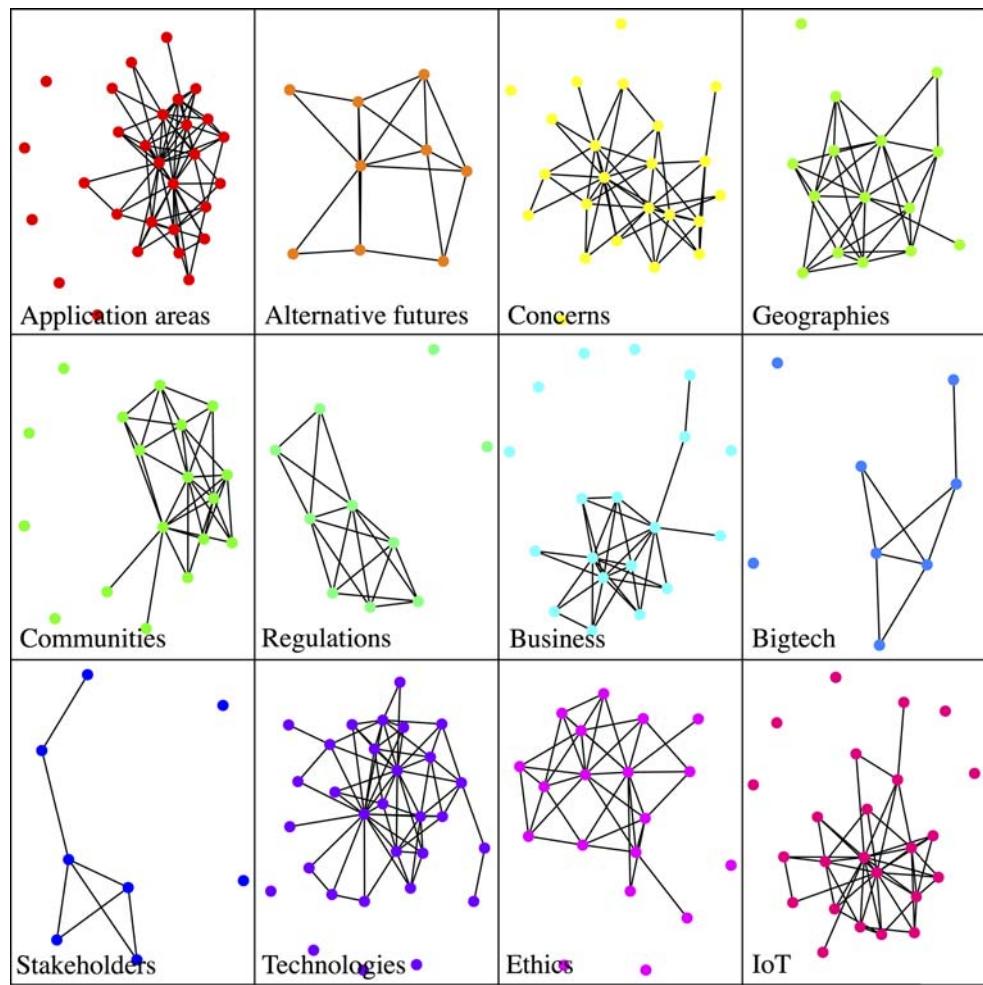
percentage of hashtags closely related to IoT after  $n$  items, also known as precision@ $n$ . Only five of the top forty hashtags are not directly related to the IoT, corresponding to a precision of .875 and supporting the hypothesis that the selected actors are highly engaged in the IoT discussions.

The conceptual dependency between the three networks can also be used to test the completeness of the consolidated dataset as a selection of relevant experts in the IoT space. Knowing, from the insights obtained from the qualitative fieldwork, that Twitter is a popular tool used to update other users about ongoing events and conversation within the professional space of IoT, it is reasonable to assume that any relevant name within that domain would be followed by a large number of the members of the consolidated dataset. Checking the data we can see that the top 100 most followed users of both the *full network* and *reduced network* belong to the initial 103 members of the consolidated dataset, showing that the qualitative selection is actually relevant for the large IoT community of Twitter.

**Table 4** Topical network with top-level categories: layer-by-layer statistics. n: number of actors, m: number of edges, nc: number of components, slc: size largest component, dens: density, cc: clustering coefficient, apl: average path length, dia: diameter. *\_flat\_* refers to a flattened network where two actors are adjacent if they are adjacent in at least one of the topical layers

	n	m	nc	slc	dens	cc	apl	dia
<i>_flat_</i>	48.00	438.00	11.00	38.00	0.39	0.35	2.17	5.00
Alternative futures	9.00	18.00	1.00	9.00	0.50	0.52	1.53	3.00
Application areas	24.00	76.00	1.00	24.00	0.28	0.43	1.80	3.00
Bigtech	6.00	8.00	1.00	6.00	0.53	0.53	1.60	3.00
Business	14.00	32.00	1.00	14.00	0.35	0.54	1.87	4.00
Communities	14.00	36.00	1.00	14.00	0.40	0.54	1.67	3.00
Concerns	20.00	50.00	1.00	20.00	0.26	0.40	1.95	4.00
Ethics	16.00	39.00	1.00	16.00	0.33	0.47	1.83	3.00
Geographies	14.00	38.00	1.00	14.00	0.42	0.52	1.64	3.00
IoT	19.00	48.00	1.00	19.00	0.28	0.42	1.84	3.00
Stakeholders	6.00	7.00	1.00	6.00	0.47	0.50	1.73	3.00
Technologies	26.00	68.00	1.00	26.00	0.21	0.42	2.14	5.00
Regulations	8.00	18.00	1.00	8.00	0.64	0.69	1.36	2.00

**Fig. 8** A multilayer topical network. The topics are, in clockwise order: Application areas, Alternative futures, Concerns, Geographies, Communities, Regulations, Business, Bigtech, Stakeholders, Technologies, Ethics, IoT. We have also included actors tweeting about an hashtag but with no followers/followees also tweeting about it, represented as disconnected nodes



## 4 Analysis

In this section we analyse both the follower/following network and the topical network extracted also using the content of the tweets.

### 4.1 Cohesion

Once we have tested the completeness and the relevance of our selection of expert users in the consolidated dataset, we can start describing how this selected subset of the IoT community on Twitter is structured. First we analyze the following/follower relations between the members of the consolidated dataset to find out the social dynamics behind the observed (online) network structure.

As expected from a network of experts in a relatively small field of tech development, our IoT network shows a high level of reciprocity (0.98) and a density of 0.13. Most of the nodes are deeply embedded within large cliques ( $k = 22$ ) in the network. One of the objectives of building a topical network is to untangle these dense relationships, organizing them according to the matters of concern under discussion.

Given the enriched information available for the original users forming the consolidated dataset, it has been possible

**Table 5** Top-40 hashtags in the tweets produced by our seed actors, ordered by number of occurrences

587	iot	40	infographic
238	blockchain	39	internetofthings
194	ai	38	fashiontech
119	cybersecurity	37	healthtech
95	leadership	35	machinelearning
80	management	34	robotics
76	bigdata	33	marketing
70	supplychain	33	5g
67	womenintech	32	artificialintelligence
63	fintech	32	ngiforum18
60	security	31	ngi4eu
60	innovation	30	iiot
59	ruiotready	28	startups
50	startup	26	healthcare
49	wearables	26	ar
49	digitaltransformation	26	vr
46	wearabletech	25	smarthomes
43	tech	25	crypto
42	technology	25	smartcity
41	lorawan	23	privacy

to explore the social dynamics behind their network structure. This has been done by studying the nominal assortativity (Noldus and Van Mieghem 2015) of the network for three specific attributes: geographic location, background, and involvement in the online conversation about IoT. All these attributes were manually specified and verified by the research team. Here by geographic location we mean the location of the actor. Figures 9 and 10 show the existing following/followers relations between users of the consolidated network and their location in Europe.

Nominal assortativity is a well known measure that quantifies the trend of a specific node to connect with other nodes with the same categorical attribute. In this case we tested nominal directed assortativity for geographic location (at a country level), users' professional background and involvement in the online conversation about ethics and IoT. The three hypotheses underlying these three tests were:

- Geographical proximity is an indicator of the presence of mutual interests between online IoT experts: users from the same geographical context will be more likely to be connected online than users from different geographical contexts. This hypothesis would result in a significant positive value of nominal assortativity if confirmed.
- Complementary background is an indicator of the presence of mutual interests between online IoT experts: users with complementary background (e.g. one in Design and one in Software development) will be more likely to be connected due to the added value of their complementarity for perspective business opportunities. This hypothesis would result in a negative value of nominal assortativity if confirmed.
- Ethical interest – the participation in the online discussion about ethics and IoT – is an indicator of the presence of mutual interests between online IoT experts: users who participate in the ongoing online discussion about IoT and ethics will be more likely to follow other users equally vocal on the issue. This hypothesis would result in a positive value of nominal assortativity if confirmed.

Figure 11 shows the level of assortativity in the network and reveals some interesting dynamics. While the degree shows a dissorative behaviour, suggesting a network organized around few hubs, nominal assortativity is always positive suggesting a small homophily effect. Among the nominal attributes, Geographical assortativity is the strongest one with a value of 0.16, while both background complementarity and participation in ethical discussion show lower values (0.05 and 0.08). These data suggest that between the three hypotheses of possible social drivers behind online connectivity only geographical proximity is

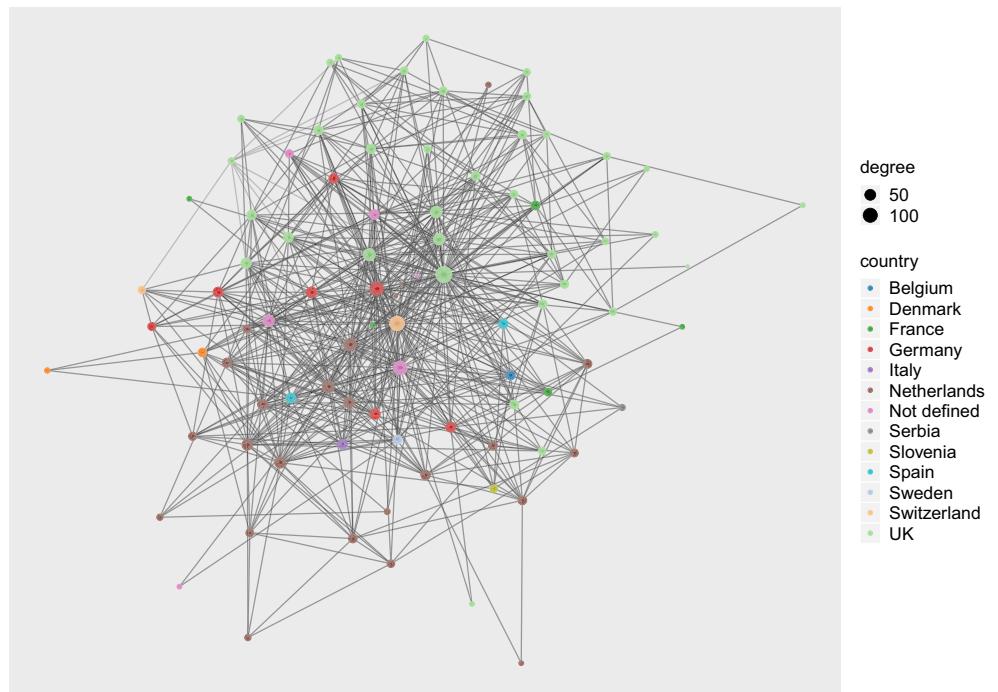
weakly supported. Professional background does not show an assortative (or dissorative) behaviour, suggesting that the reason to be connected on Twitter lies beyond the complementarity or similarity of the professional profiles. Similarly, the level of activity in online ethical discussions about IoT does not play any role in the connection process, suggesting that an interest in ethics, even if valued on an individual level, does not act as a discriminant for online connections. The only attribute that is, weakly, positive is the geographical proximity suggesting that even if there is a European IoT scene, geography still matters with the local context acting as a driving force for online connectivity.

## 4.2 Matters of Concern

In addition to the topological structure of the following/follower relations discussed in the previous paragraph we also studied the actual interactions between the members of the consolidated dataset and their online audiences. The main goal of this part of research was to investigate the presence of topical communities within the European IoT space. Since, as we showed in the previous analysis, geographical proximity still plays a significant role in existing Twitter relations, we based this analysis on the actual interactions rather than on the more stable following dynamic. Moreover we adopted the approach based on multilayer networks combined with a qualitative analysis of the most frequently used hashtags aimed at identifying how subgroups of the consolidated graph are actively engaged in thematic online conversations.

In Table 4 we already provided an overview of the topical multilayer network where we can see how topical layers are different in size and topological characteristics. All layers have a single component containing edges, in addition to a few disconnected nodes in some cases. This suggests a common conversation involving multiple users rather than isolated discussions. This is tested in Fig. 12, showing the coverage (Bródka et al. 2018) for the actors on the various layers, that is, what percentage of actors in a layer is also present in another, for each ordered pair of layers. As it can be seen by inspecting the rows *Application areas* and *Technologies and Techniques*, containing high values of coverage across the whole row, many of the users participating in discussions on other topics are subsets of the users participating in these two topics. It is worth noticing that interactions in the categories *Ethics* and *Concerns* are also frequent with respect to the size of the data, and, not surprisingly, well associated to each other as we can see in the corresponding cells of Fig. 12. Other layer correlation matrices using different measures suggested by Bródka et al. (2018) and not adding additional information are not shown in the paper.

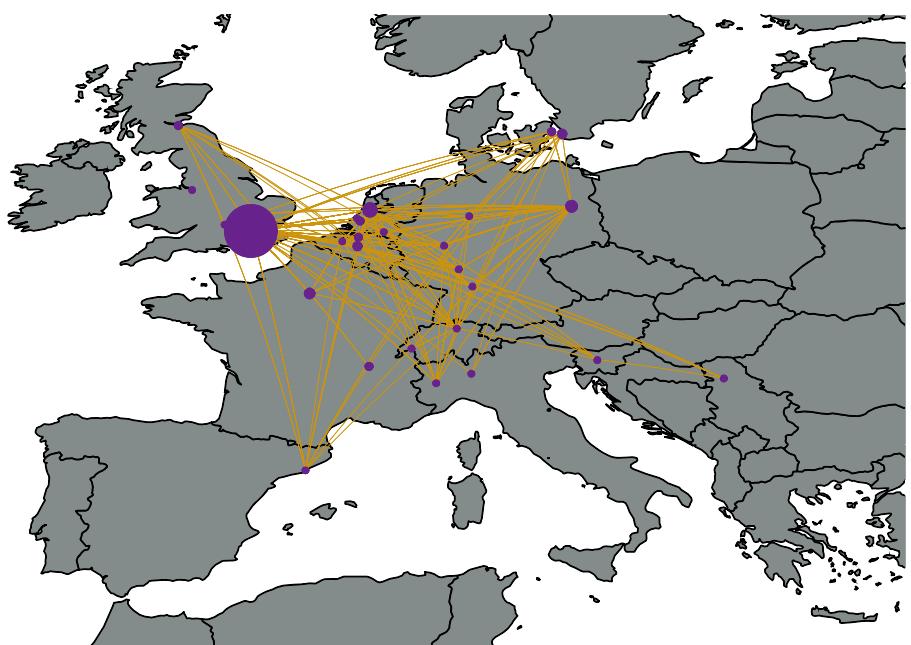
**Fig. 9** Consolidated Network.  
Colors represent users' country



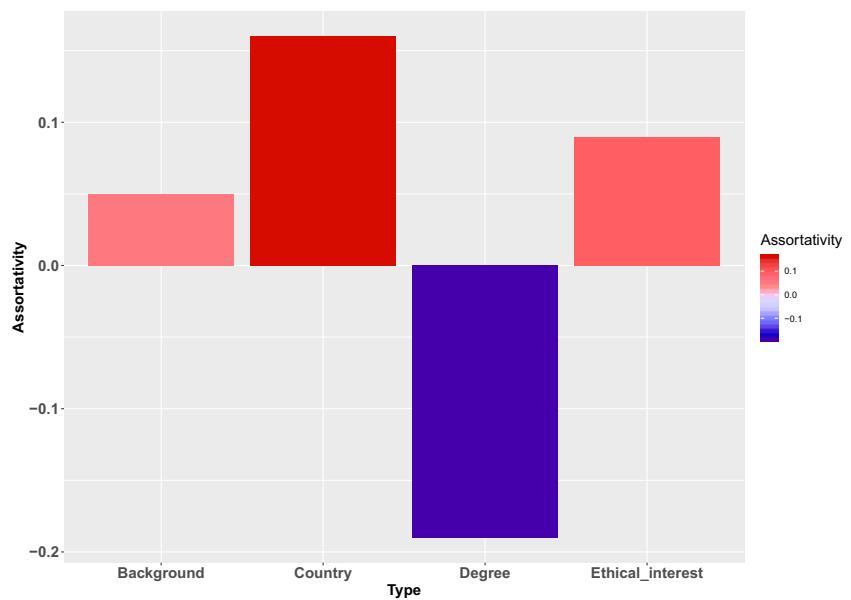
When we read this figure together with the findings from our qualitative research, we notice the central importance of application areas and technologies across all categories. One way of explaining this central “concern” is the prominent role application areas and technologies are given in business discourse. Indeed, there is a tendency to discuss technologies and their implications through their business viability and growth opportunities. This is why we see co-hashtagging as a common pattern, in the form of for

example, #healthcare, #hardware, #IoT and #startups. When we specifically look at the *Concerns* category, we notice that business emerges as a lesser concern than ethics, big tech, stakeholders and alternative futures. While this might be partly due to business and concern categories having collinearity, it might also be partly explained by the fact that when concerns are discussed, it is more common to substantiate them through their disadvantages rather than [business] opportunities. More specifically, when we

**Fig. 10** Consolidated Network.  
Connections between cities



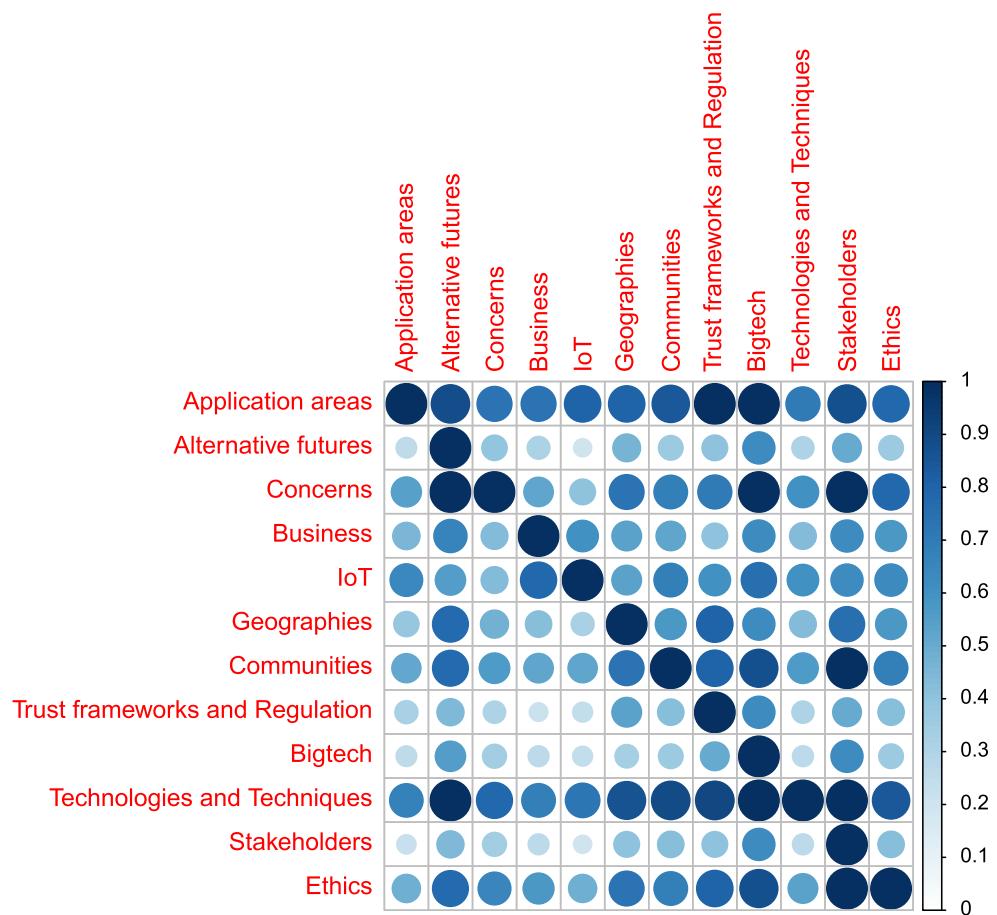
**Fig. 11** Nominal and Degree assortativity observed in the network



look at the discussions where ethics and other hashtags appear at the same time, the framing is usually done through mentioning the stakeholders (e.g. regulators, big tech, regulatory platforms) whereas with business, the focus is more on alternative futures, stakeholders and IoT.

The existence of multiple thematic subgroups within the IoT Twitter space is confirmed by the application of community detection methods to the data. Using the multiplex clique percolation method (Tehrani et al. 2018) we can observe how in the data the number of communities

**Fig. 12** Interlayer actor coverage: cell [X,Y] indicates the portion of the actors participating in the discussion on topic Y who have also participated in X



detected when we require at least two common layers in a clique ( $m = 2$ ) and cliques of at least size 5 is relatively small counting for only 3 communities with 5 to 6 members. However, two of these communities span 6 and 7 of the topical layers, indicating a very small but active core of actors covering a combination of various types of matters of concern. Communities that are present only on one of the layers (that is, dense discussion groups only interacting about one topic) are larger: up to 9 actors if we consider the same density constraints as above (5-cliques), and larger if we further relax these constraints.

## 5 Discussion and Conclusion

In this article we have presented a methodology to study the European IoT using Twitter, based on a combination of quantitative and qualitative methods, the construction of a small but rich dataset, and different types of networks generated from it, including a new type of multilayer topical network coding relationships between users not directly available from the Twitter API.

The overall picture that emerges from the study of the Twitter IoT consolidated network is complex. On the one side there is a topological common space properly defined and identified by our manual selection of the key actors. Nevertheless, beside this space of following/followers relations there is the actual space of interaction that, in itself, is subdivided into the many domains that characterize the IoT space.

Building and using a small, high-quality dataset is a crucial aspect of this work. Ideally, we would have liked to use Twitter data to perform various types of analysis: what is discussed in IoT-related tweets, where are the different topics appearing, when are they discussed (for example, when they emerge for the first time and when they reach a peak of popularity), who is leading the discussion on specific topics, and finally a joint analysis putting all these aspects together to map online conversations.

However, a large-scale automated analysis is complicated by several issues. The first problems we had to face concerned the spatial analysis of the tweets. First, we have no guarantee that the geographical distribution of the collected tweets is a good indicator of the number of tweets with the #IoT hashtag produced in different locations. This is due to uncertainty about the sampling algorithm used by Twitter and about the adoption of geo-localized tweets in different regions. A second type of problems regards the noise in the data, that is, the presence of irrelevant tweets. An example is the inclusion of tweets automatically produced by traffic cameras in Brazil in our monitoring of the #IoT hashtag. While these are clearly IoT devices, justifying their usage of the #IoT hashtag, on

top of not being in Europe these tweets are not relevant in the identification of online discussions between important actors and inside IoT communities of practice. The previous issue also highlights an additional question: which users should be considered in the analysis? If the answer is clear about traffic cameras, it is less clear whether other types of bots<sup>3</sup> should be included or not. At this time the application of bot detection methods has not produced results that are accurate enough to be included in this article.

Another challenge to be able to perform the aforementioned analyses is the identification of topics. Our experiments with automated topic detection methods (Blei et al. 2003) confirmed that the special language and the short texts in Twitter do not allow an accurate topic extraction. Grouping tweets into larger texts did not help either. This, in addition to other challenges in the state of the art on automated topic detection (such as the selection of the number of topics to be retrieved and the presence of non-deterministic results) forced us to rely once more on manual pre-processing, where the qualitative team checked the list of most common hashtags and defined a set of rules to automate their normalization.

**Funding Information** Open access funding provided by Uppsala University.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Bijker, W.E., Hughes, T.P., Pinch, T. (1987). *The social construction of technological systems*. Cambridge: MIT Press.
- Blei, D.M., Ng, A.Y., Jordan, M.I. (2003). Latent dirichlet allocation. *The Journal of Machine Learning Research*, 3, 993–1022. <http://dl.acm.org/citation.cfm?id=944919.944937>.
- Bródka, P., Chmiel, A., Magnani, M., Ragozini, G. (2018). Quantifying layer similarity in multiplex networks: a systematic study. *Royal Society open science*, 5(8), 171747.
- Dickison, M.E., Magnani, M., Rossi, L. (2016). *Multilayer social networks*. Cambridge: Cambridge University Press.
- Fritsch, E., Shklovski, I., Douglas-Jones, R. (2018). Calling for a revolution: An analysis of IoT manifestos. In *Proceedings of the 2018 CHI conference on human factors in computing systems* (pp. 1–13). <https://doi.org/10.1145/3173574.3173876>.

<sup>3</sup>AI-based computer programs simulating users

- Hanteer, O., Rossi, L., D'Aurelio, D.V., Magnani, M. (2018). From interaction to participation: the role of the imagined audience in social media community detection and an application to political communication on twitter. In *ASONAM* (pp. 531–534): IEEE Computer Society.
- Kitchin, R. (2017). Thinking critically about and researching algorithms. *Information, Communication & Society*, 20(1), 14–29.
- Kivelä, M., Arenas, A., Barthelemy, M., Gleeson, J.P., Moreno, Y., Porter, M.A. (2014). Multilayer networks. *Journal of Complex Networks*, 2(3), 203–271. <https://doi.org/10.1093/comnet/cnu016>.
- de La Bellacasa, M.P. (2011). Matters of care in technoscience: Assembling neglected things. *Social Studies of Science*, 41(1), 85–106.
- Latour, B. (2004). Why has critique run out of steam? from matters of fact to matters of concern. *Critical inquiry*, 30(2), 225–248.
- Latour, B. (2005). *From realpolitik to dingpolitik. Making things public: Atmospheres of democracy*. Cambridge: MIT Press.
- Magnani, M., & Rossi, L. (2011). The ML-model for multi layer network analysis. In *IEEE international conference on advances in social network analysis and mining*. Los Alamitos: IEEE Computer Society.
- Noldus, R., & Van Mieghem, P. (2015). Assortativity in complex networks. *Journal of Complex Networks*, 3(4), 507–542.
- Powell, A.B. (2018). Moral orders in contribution cultures. *Communication Culture & Critique*, 11(4), 513–529.
- Tehrani, N.A., Magnani, M., Afsarmanesh, N., Magnani, M. (2018). Partial and overlapping community detection in multiplex social networks. In *Social informatics*, (Vol. 11186 pp. 15–28): Springer. Lecture notes in computer science.
- Ustek-Spilda, F., Powell, A., Shklovski, I., Lehude, S. (2019). Peril v. promise: Io-t and the ethical imaginaries. In *CHI'19 Workshop Proceedings*. arXiv:1906.10378.
- Vega, D., & Magnani, M. (2018). Foundations of temporal text networks. *Applied Network Science*, 3(1), 25:1–25:26.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Funda Ustek Spilda** is a post-doctoral researcher at the Oxford Internet Institute, University of Oxford. She works for the Fairwork Project, analysing working conditions in the emerging gig economy. She previously held post-doc positions at VIRT-EU and ARITHMUS: How data make a people. She studies ethics, fairness and justice in data and technology, with a focus on labour, gender and migration.

**Davide Vega** is a lecturer at the Department of Information Technology at Uppsala University, working in the intersection of computational social science and network science. He is a member of the Uppsala Information Laboratory, where he develops analytical models and methods to study online communication and its influence on our collective opinions and ideas.

**Matteo Magnani** is an associate professor and distinguished university teacher at Uppsala University. He leads the Uppsala Information Laboratory, doing research on the analysis of online information networks. He is one of the authors of the book *Multilayer Social Networks* and of the multinet library for multilayer network analysis.

**Luca Rossi** is Associate Professor of Digital Media and Networks at the Department of Digital Design of IT, University of Copenhagen. He is member of the Networks Data and Society (NERDS) research group and of the Digital Platforms and Data research group. He also coordinates the Data Science & Society research lab. His interdisciplinary research tries to connect traditional sociological theory with computational approaches. Within this line of research he has developed research in the context of online participation, online activism, political campaign and election studies.

**Irina Shklovski** is Professor of Communication and Computing at the University of Copenhagen working across the disciplines of computer science and communication. Her projects address responsible technology design, data governance, data leakage on mobile devices and the sense of powerlessness people experience in the face of massive personal data collection. Most recently she coordinated project VIRT-EU -<http://www.vrtieuproject.eu/servicepackage>

**Sebastian Lehude** is a PhD researcher at the Department of Media and Communications at the London School of Economics. His doctoral thesis interrogates the geopolitics of international data collaborations from a decolonial perspective. Sebastian worked as research assistant for Virt-EU, conducting participant observation of Internet of Things meetups and hackathons. He also taught on digital communication at the Faculty of Journalism of the Pontifical Catholic University of Chile.

**Alison Powell** is an Associate Professor in the Department of Media and Communication at the London School of Economics and Political Science. She is also the Director of the JUST AI Network, a humanities-led project investigating data and AI ethics in partnership with the Ada Lovelace Institute. Her research investigates ethics in practice in technology development, with a focus on concerns of care and justice.

## Affiliations

Funda Ustek-Spilda<sup>1</sup> · Davide Vega<sup>2</sup> · Matteo Magnani<sup>2</sup> · Luca Rossi<sup>3</sup> · Irina Shklovski<sup>4</sup> ·  
Sebastian Lehuede<sup>5</sup> · Alison Powell<sup>5</sup>

Funda Ustek-Spilda  
funda.ustek@oii.ox.ac.uk

Davide Vega  
davide.vega@it.uu.se

Luca Rossi  
lucr@itu.dk

Irina Shklovski  
ias@di.ku.dk

Sebastian Lehuede  
s.a.lehueude-bravo@lse.ac.uk

Alison Powell  
a.powell@lse.ac.uk

<sup>1</sup> University of Oxford, Oxford Internet Institute, Oxford, UK

<sup>2</sup> InfoLab, Department of Information Technology,  
Uppsala University, Uppsala, Sweden

<sup>3</sup> IT University Copenhagen, Copenhagen, Denmark

<sup>4</sup> University of Copenhagen, Copenhagen, Denmark

<sup>5</sup> The London School of Economics and Political Science,  
London, UK