PLOS ONE

RESEARCH ARTICLE

A longitudinal analysis of the public perception of the opportunities and challenges of the Internet of Things

Arkaitz Zubiaga[1]*, **Rob Procter**[1,2], **Carsten Maple**[1,2]

1 University of Warwick, Coventry, United Kingdom, **2** Alan Turing Institute, London, United Kingdom

* a.zubiaga@warwick.ac.uk

Citation: Zubiaga A, Procter R, Maple C (2018) A longitudinal analysis of the public perception of the opportunities and challenges of the Internet of Things. PLoS ONE 13(12): e0209472. https://doi.org/10.1371/journal.pone.0209472

Editor: Pablo Dorta-González, Universidad de las Palmas de Gran Canaria, SPAIN

Received: December 12, 2017

Accepted: December 6, 2018

Published: December 20, 2018

Data Availability Statement: The dataset used in this paper is available through Figshare: https://figshare.com/articles/Internet_of_Things_tweets_2009-2016_/5684548.

Funding: This work is part of the PETRAS project, which is funded by a grant from the UK Engineering and Physical Sciences Research Council (EP/N02334X/1): https://www.petrashub.org/. We wish to thank the Alan Turing Institute (EP/N510129/1), which is funded by the Engineering and Physical Sciences Research Council, for its support. The funders had no role in

Abstract

The Internet of Things (or IoT), which enables the networked interconnection of everyday objects, is becoming increasingly popular in many aspects of our lives ranging from entertainment to health care. While the IoT brings a set of invaluable advantages and opportunities with it, there is also evidence of numerous challenges that are yet to be resolved. This is certainly the case with regard to ensuring the cyber security of the IoT, and there are various examples of devices being hacked. Despite this evidence, little is known about the public perceptions of the opportunities and challenges presented by the IoT. To advance research in this direction, we mined the social media platform Twitter to learn about public opinion about the IoT. Analysing a longitudinal dataset of more than 6.7 million tweets, we reveal insights into public perceptions of the IoT, identifying big data analytics as the most positive aspect, whereas security issues are the main public concern on the negative side. Our study serves to highlight the importance of keeping IoT devices secure, and remind manufacturers that it is a concern that remains unresolved, at least insofar as the public believes.

Introduction

The Internet of Things (IoT) is a concept that refers to the networked interconnection of everyday objects, which are often equipped with ubiquitous intelligence [1]. The use of IoT devices is becoming commonplace in our daily lives given the growing presence of WiFi and 4G-LTE Internet connectivity [2, 3]. The IoT presents numerous applications in different contexts, including [2]: (1) home, e.g. entertainment, health monitoring, (2) transport, e.g. traffic control, parking management, (3) community, e.g. environment monitoring, surveillance, and (4) national, e.g. defense, remote monitoring. The utility of the IoT can range from mere personal use at home to use in the industry as well as by doctors or carers for remote assistance [4].

As well as being a valuable technology for remote and networked control of devices and data sources, the IoT also comes with the caveats of other Internet-connected devices: potential security and privacy issues linked to the use of those devices [5–8]. The fact that data associated with IoT devices is sent through the Internet and stored in the cloud can make it vulnerable

study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

[9] and can expose IoT devices to hackers [10, 11]. Security and privacy of IoT devices are especially crucial when they are being used with personal data associated with sensitive aspects such as health care [12].

There is indeed a growing concern about the security and privacy issues brought about by the IoT [13–16]. Scientists are among those calling for new regulatory approaches that will enable attacks to be intercepted, data authenticated, access controlled and the privacy of customers guaranteed [17]. Further, [18] highlight that while *"the main strength of the IoT idea is the high impact it will have on several aspects of everyday-life and behavior of potential users"*, they also warn about the numerous issues that still need to be addressed: *"many challenging issues still need to be addressed and both technological as well as social knots have to be untied before the IoT idea being widely accepted. Central issues are making a full interoperability of interconnected devices possible, providing them with an always higher degree of smartness by enabling their adaptation and autonomous behavior, while guaranteeing trust, privacy, and security."* In this work we set out to explore how these three concepts, namely trust, privacy and security, are perceived and potentially questioned by the general public.

Attitudes and perceptions of IoT security were analysed by [19] through surveying a group of actors in the energy, water and health care sectors. 18 representatives from 11 different organisations were interviewed as part of the study. Respondents in this study reported concerns about the lack of IT maturity and potential drawbacks of the IoT technology. The researchers, however, noticed a certain degree of lack of awareness among respondents, highlighting that two of the respondents had not even heard the term IoT before.

A number of studies and surveys have gathered the knowledge and opinions of experts, as well as the perceptions of actors of relevant sectors. However, the opinion of the general public has been studied to a much lesser extent. To the best of our knowledge, two studies have used social media to assess the public perceptions of the IoT. One of these two works is by [20], who analysed a small dataset of 40K tweets covering two months in 2016, examining among others the most frequent users and hashtags in the dataset. Their analysis did not identify content associated with privacy and/or security, potentially owing to the limited size of the dataset. The other work is by [21], who used larger datasets for their analysis, however, with gaps in the collection. They used a lexicon-based sentiment classification system and an LDA topic modelling approach to analyse different topics and sentiments associated with IoT. They did find that one of the topics identified in their LDA was indeed security, showing that the public has some concern about it; their study did not however further dig into the sentiments associated with these particular topics.

In this work, we perform a longitudinal analysis of public perceptions of the opportunities and challenges presented by the IoT; we use the social media platform Twitter as a data source to analyse posts from 2009 to 2016. We use a state-of-the-art sentiment analysis classifier to analyse the polarity of posts over time, and use a range of algorithms to enhance the dataset with additional dimensions to explore, including the country of origin of tweets and the gender of users, among others. Moreover, we also use an LDA topic modelling algorithm to identify the topics discussed in tweets associated with the IoT, which enables us to perform a fine-grained analysis of polarity by topic. By further examining these topics, we find that "Big data & tech" is the largest emerging topic associated with the opportunities afforded by the IoT, whereas the key negative aspect and the second by number of tweets is "Security", which we observe has been a constant concern since 2009. Our study reveals insight into public perception of the IoT, highlighting the key opportunities and challenges posed by this technology. The use of unsupervised approaches to conduct our study also enables the application of our analysis to analyse the public perception associated with other major topics discussed on Twitter.

Materials and methods

The present study was approved by the Warwick University Humanities & Social Sciences Research Ethics Committee (HSSREC), which received full ethics approval with a duration of 48 months (ref 69/13-14, date: 30.05.2014), including approval to store, analyse and publish extracts from social media datasets.

In this study, we use the microblogging platform Twitter as the data source to mine public comments associated with the IoT. Other platforms were also considered at the beginning of this study, including, for example, Reddit. These other potential sources were deemed unsuitable as the vast majority of posts were from *tech savvy* people who were dealing with technical challenges of setting up and using IoT devices, rather than more general comments from users who would be more representative of the 'average' consumer. Rather, Twitter was found to be much more suitable, as there is a variety of users of different levels of technical sophistication, discussing a wide range of issues.

Using a combination of Twitter's search interface and its REST API, we harvested a collection of tweets spanning eight years between 1st January, 2009 and 31st December, 2016. Given that Twitter's API does not give access to old tweets, we scraped the tweets from the site's search engine, which allows going back to the oldest tweets. This scraping process enabled us to collect all tweet IDs that matched our search query; these tweet IDs were then used to retrieve all tweets and metadata from the REST API. Tweets that included one of the following keywords were collected: '#iot', 'internet of things', '#internetofthings'. We filtered out non-English tweets, as well as retweets, leading to a dataset with 6,705,948 tweets. This large-scale dataset enabled longitudinal analysis of public comments associated with the IoT on Twitter.

To enable detailed analysis of this dataset by looking at additional factors that Twitter does not directly provide through its API, we pre-processed the dataset. This pre-processing included the following text mining, classification and further data collection steps that we carried out to complete the dataset:

1. **Unpacking of URLs**: Many of the links in tweets tend to be shortened by using URL shortening services such as bit.ly or tinyurl.com, which is a useful service for users, owing to the limited number of characters allowed in each tweet. Since we are interested in analysing the links that people direct to when discussing IoT, we unpacked the shortened URLs by using the cURL library [22]. In this process of unpacking URLs, we stored the ultimate URL to which a shortened URL directs to. While we also retrieved the content of the URLs, this is not used for our quantitative analysis, which focuses on the analysis of final URLs.

2. **Classification of tweets by country of origin**: We used a state-of-the-art tweet geo-location tool [23] to classify tweets by country; it is a multinomial logistic regression that combines a range of tweet content and metadata for the classification, which is publicly available [24]. We used the publicly available, geolocated dataset with over 5 million tweets [25] for training the classifier, which then applied to our dataset of IoT tweets. This enabled us to analyse what people's opinions are with respect to the IoT in different countries. Given that only a small subset of the tweets analysed were geo-tagged, we evaluated the accuracy of the classifier over those tweets, i.e. tweets that are geolocated by the user's device. The results gave an overall 81.4% accuracy in a classification task involving 217 countries.

3. **Inferring gender of users**: To infer the gender of users we used the first name of the Twitter users utilising the SexMachine Python package [26]. We extracted the first name of the author of a tweet from the "name" field of a tweet. With this first name as input, the Python package then returns one of male, female, mostly_male, mostly_female or andy (for

unknown names). We only kept the gender label for those classified as male or female (i.e. excluding those classified as mostly_male, mostly_female), labelling the rest as unknown.

4. **Sentiment analysis**: Each tweet was classified as positive, negative or neutral, using a state-of-the-art sentiment classifier [27] that determines the sentiment of a tweet with respect to IoT as a target. This is different from tweet-level sentiment classification, which classifies the overall sentiment expressed in the tweet, irrespective of the target. For instance, the tweet *"I like that there is quite a lot of research on IoT devices lately"* bears a positive overall sentiment, however, the sentiment is neutral with respect to IoT. Target-specific sentiment classification is being increasingly used where the objective is to determine sentiment towards specific topics or entities [28].

Finally, we also processed the entire dataset for topic modelling using LDA [29], for which we used the implementation provided with Gensim [30]. With LDA, we generated a set of 6 topics from our dataset, and we categorised each tweet in the dataset into one of these topics. Empirically, different numbers of topics ranging from 5 to 10 were tested to find the optimal number that would lead to as many distinct topics as possible while avoiding repeated topics. Six was found to be the optimal number of topics. Following the topic modelling based clustering approach proposed by Wang et al. [31], we summed the topic probabilities for each keyword in a tweet, choosing the topic that maximised this sum as the category for the tweet.

To validate that the assumption that the growth of Twitter from 2009 to 2016 Twitter data has not had a significant effect on the dataset, we also collected data from Google Trends. Using the search engine of Google Trends, we retrieved the trends of the "Internet of Things" as a topic between 2009 and 2016. With this, we built monthly activity counts for Google Trends, and we did the same for Twitter data. A comparison between these two lists of frequencies showed a high Pearson correlation value of $\rho = 0.8978$ ($p < 2.2e^{-16}$), suggesting the growth of mentions of IoT was commensurate with the growth on Google Trends, without a significant effect of Twitter's user base growth.

Results

Sentiment analysis

First, we examined the sentiment expressed by users from 2009 to 2016. Out of the 6,705,948 tweets in the dataset, 295,674 (4.4%) have either positive or negative sentiment, whereas the rest have a neutral sentiment. This suggests that a vast majority of the tweets are likely to be neutral stories and other comments that do not express any sentiment towards IoT. To validate the output of our sentiment classifier and to make sure that it does not have a tendency to label tweets as neutral, we also tested an alternative sentiment classifier, Vader [32], a widely used sentiment classifier for tweets and social media. A comparison of labels predicted by our classifier and by Vader led to an overlap of 95.45% of the instances, showing the prevalence of neutral tweets in the dataset. A closer look at a subset of the neutral tweets shows that in fact many of those tweets are linking to news articles, whose headlines are copied into the tweets and have neutral sentiment.

Given that our objective was to look at the benefits and concerns of the IoT that people share opinions about, we focused our sentiment analysis on the positive and negative tweets. Fig 1 shows a timeline with monthly percentages of positive and negative tweets over time. We observe that for the first years between 2009 and 2013, both positive and negative tweets went hand in hand, with very similar percentages. There are exceptions where negative tweets experienced significant spikes. This is potentially due to specific news stories that revealed negative aspects of the IoT, which we will analyse below. It is only in the last two years, 2015 and 2016,

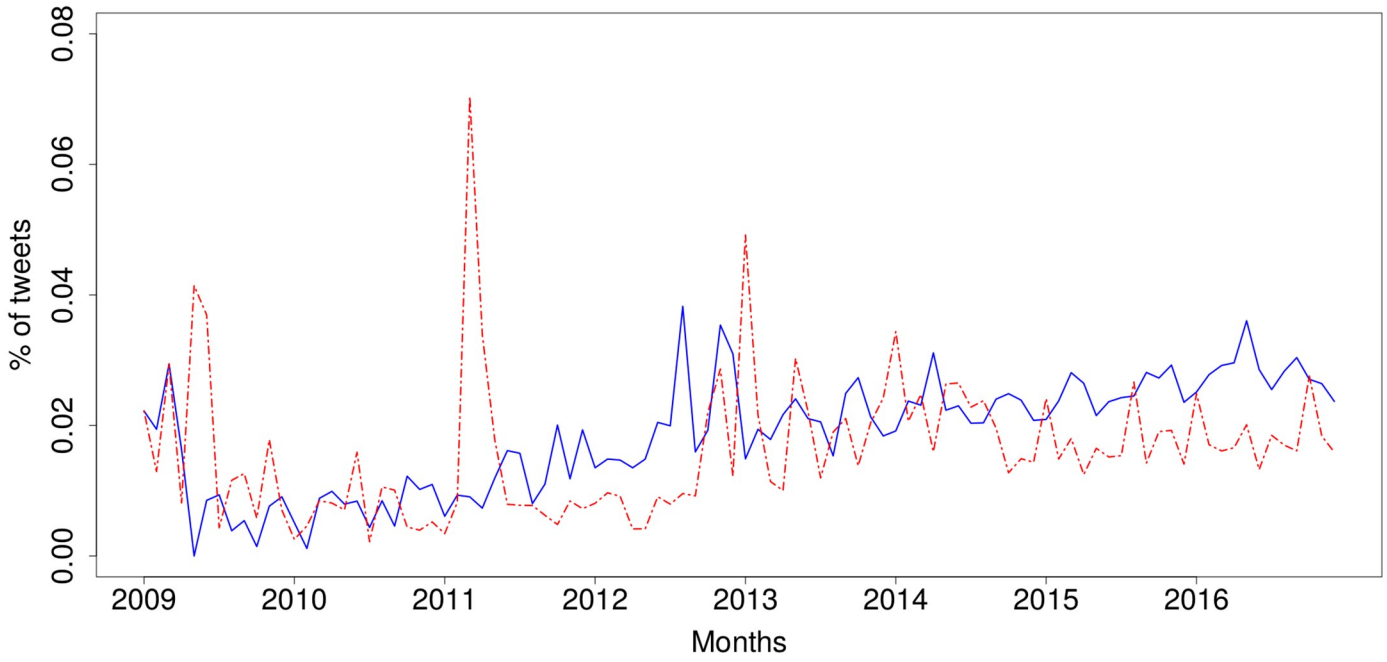

Fig 1. Sentiment of tweets discussing the Internet of Things over time (Jan 2009–Dec 2016). Blue, solid line: positive, red, dashed line: negative.

https://doi.org/10.1371/journal.pone.0209472.g001

that the positive tweets exceeded the negative tweets quite consistently, suggesting that the positivity of the IoT is improving over time.

One issue that arises from this sentiment analysis is that one may think that the results can be rigged by the presence of bots, i.e. automated accounts or 'bots' that tweet in volume positive or negative comments on the IoT [33]. This may happen, for instance, if a company wants to promote their IoT products (i.e. positive tweets) or competitors want to widely disseminate the drawbacks of IoT systems (i.e. negative tweets). To quantify and to validate our analysis, we examined the sentiment expressed by human users. To do this, we considered the subset of tweets for which the author had been identified as either male or female, i.e. having a human name. This led to a subsample with 2,602,138 tweets (38.8% of the whole) that were identified as tweets by people. Fig 2 shows the sentiment expressed over time by human users. The overall sentiment trends in Fig 1 and the human-only sentiment in Fig 2 have a Pearson correlation of 0.93 when we look at the positive tweets, and 0.95 for negative tweets, which can also be observed in the significant similarity between the charts. One striking difference that stands out is the sharper spikes seen in the negative tweets, which is more pronounced than in the positive tweets.

Topics

While the sentiment analysis above can be indicative of the overall perceptions of the IoT, we were interested in examining in detail the different issues and topics that people bring up around the IoT, and the sentiment expressed towards those topics. For this, we identified 6 different topics using a Latent Dirichlet Allocation algorithm (LDA), as described above. Fig 3 shows the keywords representing the six topics, where the importance of each keyword is represented by its font size, with larger font sizes representing more important keywords. We observe that five of these six topics are associated with either positive aspects or business opportunities of the IoT, including analytics, machine learning, big data and tech, machine to

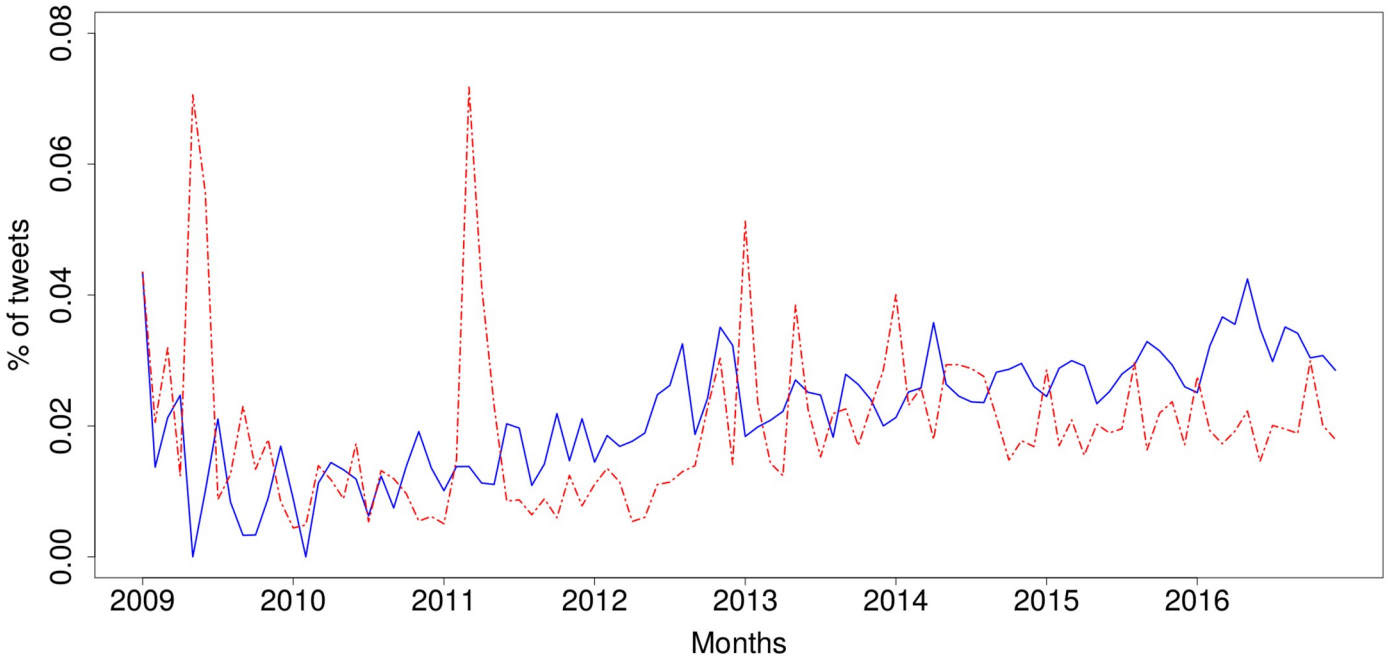

Fig 2. Sentiment of tweets posted by humans discussing the Internet of Things over time (Jan 2009–Dec 2016). Blue, solid line: positive, red, dashed line: negative.

https://doi.org/10.1371/journal.pone.0209472.g002

machine communication, and devices. There is, however, a sixth topic, where one of the possible issues of the IoT is discussed, security. The largest topics in terms of the number of tweets associated with them are first,"Big data & Tech" with nearly 30% of the tweets, and second,"Security" with over 20% of the tweets. This demonstrates that security is indeed a key concern around the IoT.

To further drill down into the analysis and quantification of topics, a temporal trend of topic sizes is shown in Fig 4. The trends show that while "Big data & Tech" is increasing in popularity. The opportunities that the IoT is bringing to the big data and technology businesses is indeed becoming increasingly popular, as is reflected in this analysis. "Security" is stable over the last few years, with a slight drop since the 2009 and remains a big issue, consistently ranking second from 2013 to 2016.

Next, we look at the positive and negative sentiment associated with these six topics. Fig 5 shows the monthly percentage of positive and negative tweets for each of the topics. In two cases we can observe that positive tweets have increased particularly in recent years, which is the case of "Analytics" and "Machine learning." In these two cases, positive tweets clearly exceed negative tweets over the last few years. Two more topics, "Big data & Tech" and "M2M communication," are showing a similar tendency, however, the difference is not as remarkable. 'Devices' is a topic that has people divided with similar amounts of positive and negative tweets, and the most negative topic is "Security," which shows an increasing tendency towards more negative tweets in recent years.

It may be surprising, at first thought, to observe that *Analytics*, *Machine learning* and *Big data & Tech* are viewed as generally positive over the period considered. However, a number of recent surveys have indicated more positive opinions about artificial intelligence and machine learning. The Royal Society report, Public views of Machine Learning [34], considered existing work on public attitudes about emerging technologies and found that "broadly speaking, the public are supportive of science and scientific developments and want to know

Fig 3. Word clouds for topics extracted using LDA, as well as percentage of the whole represented by each topic. Left to right and top down: (1) Devices (15.27%), (2) Machine to machine communication (16.14%), (3) Security (20.49%), (4) Big data & Tech (29.92%), (5) Analytics (12.09%), (6) Machine learning (6.09%).

https://doi.org/10.1371/journal.pone.0209472.g003

more about them." The report itself considered the current and near-term (5-10 years) applications of machine learning. The research involved 978 face-to-face interviews weighted to ensure the individuals selected for interview were representative of the UK population, and was supplemented with qualitative research. Though the report failed to identify that positive sentiments to machine learning significantly outweighed negative feelings, headlines in the mainstream media included "Artificial intelligence survey finds UK public broadly optimistic"

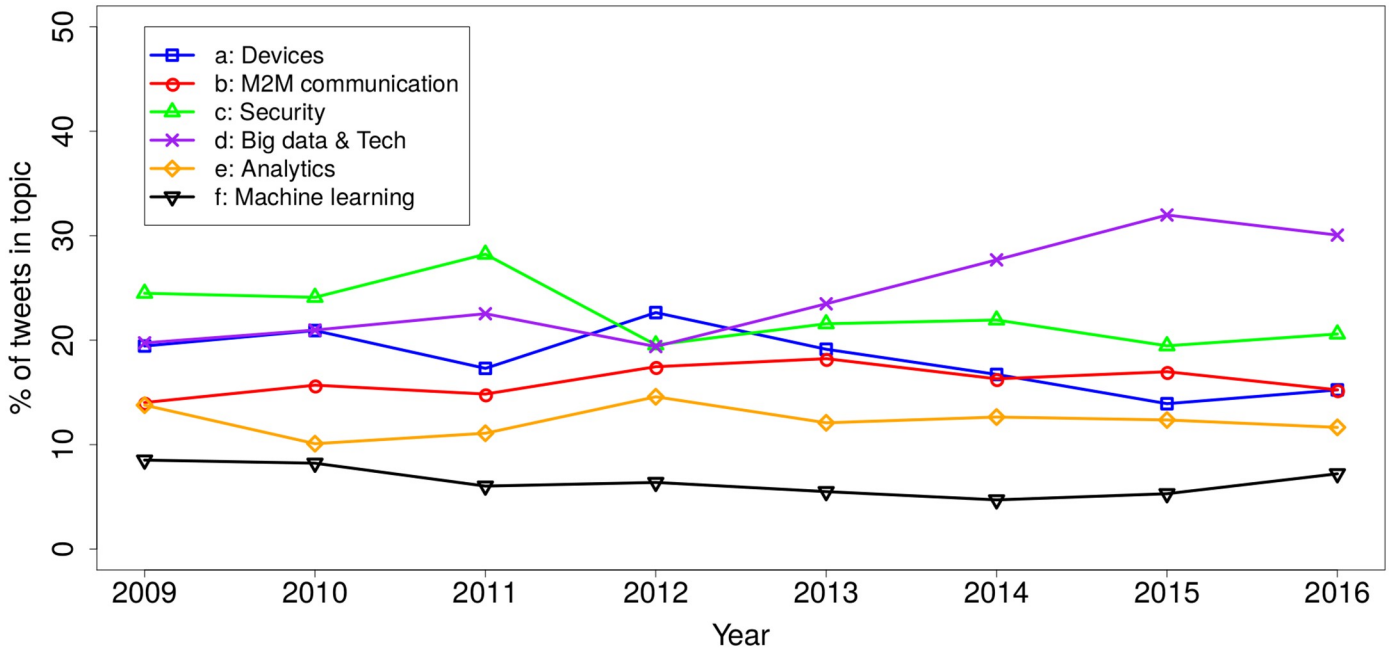

Fig 4. Popularity of each of the six topics over time.

https://doi.org/10.1371/journal.pone.0209472.g004

[35]. Academic work has found stronger support for AI and Machine learning, including work by Fast and Horvitz of Stanford University and Microsoft, respectively. Their work found that "discussion of AI has increased sharply since 2009, and that these discussions have been consistently more optimistic than pessimistic" [36]. However, the positivity seen in these algorithmic, software-based aspects of the Internet of Things does not extend to *Devices* or *Security*. The relatively less positive opinion of Devices might be explained by the public being concerned about hardware, and in particular its ability to act, rather than the more "hidden" software layer. It is not surprising that security is an area that is viewed, in aggregate, as negative. Security is one of the key concerns in the Internet of Things, and it is usually only discussed in the media, both mainstream and social, when an incident happens due to a security vulnerability.

It may also be surprising to note that there were generally more tweets that were positive about M2M than those that were negative. This might be explained by considering previous research that explored public opinions of robotics, a significant use case of M2M technology. A recent survey [37] found that 81% of respondents favoured the use of robotics in manufacturing. As such, it is less surprising that there is a slight overall positive attitude to M2M in our results.

If we look at the overall positivity of topics (Table 1), we see that the opportunities that the IoT is bringing to business are most positively perceived by people, including *Analytics*, *Machine learning*, *Big data & Tech* and *M2M communication*. The topics where the negativity exceeds the positivity include especially *Security*, which proves to be the most concerning issue linked to IoT, followed by *Devices* with a lower degree of negativity.

Countries

By having tweets geo-located by country, we were able to analyse sentiment by country towards the IoT. Fig 6 shows a heat map highlighting the extent to which tweets posted (in

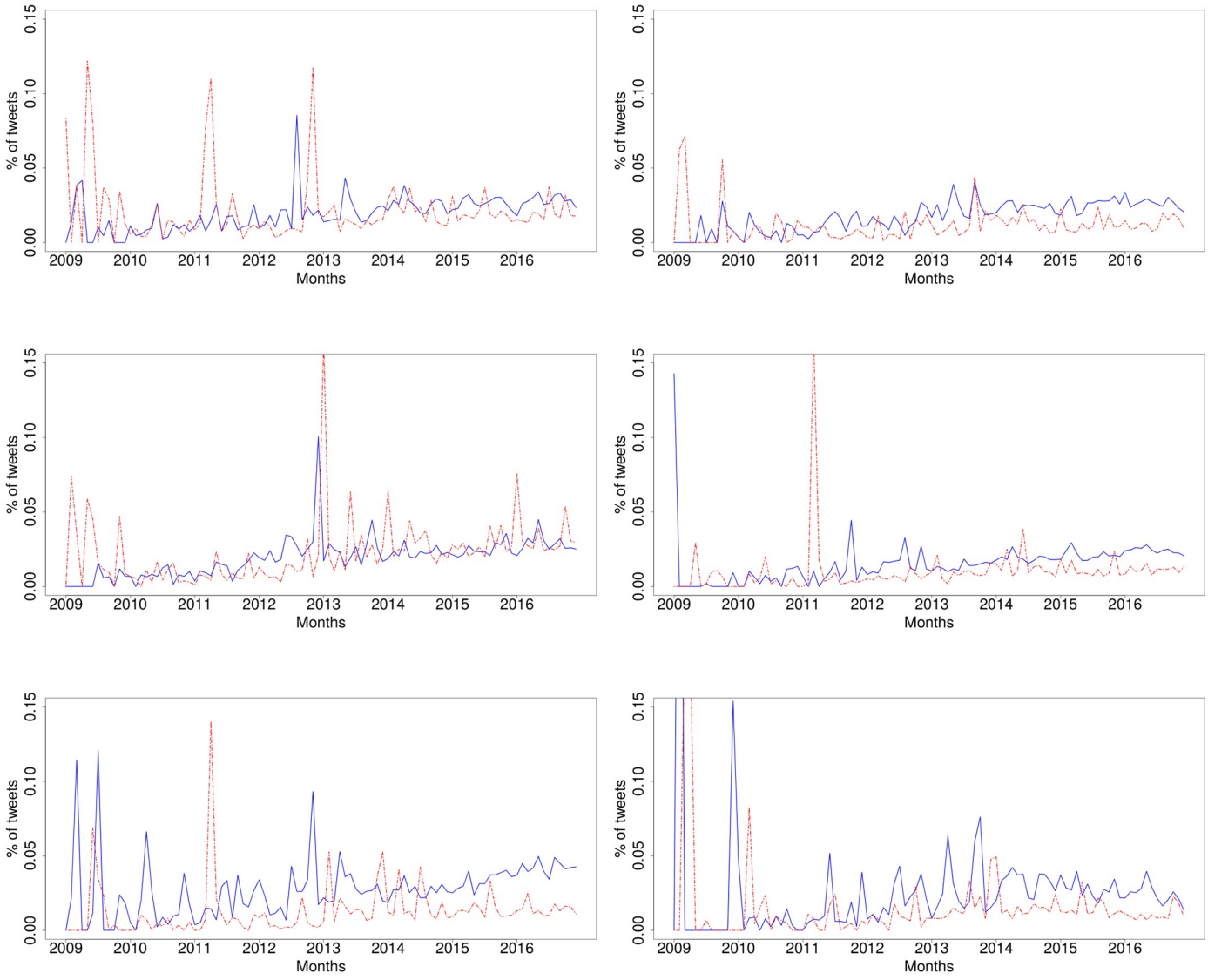

Fig 5. Timeline of sentiments for the 6 topics. Blue, solid line: positive, red, dashed line: negative. Left to right and top down: (1) Devices, (2) M2M communication, (3) Security, (4) Big data & Tech, (5) Analytics, (6) Machine learning.

https://doi.org/10.1371/journal.pone.0209472.g005

Table 1. Positivity by topic, computed as the percentage of positive tweets that exceed negative tweets. Topics are sorted by positivity, most positive first.

Topic	Positivity
(e) Analytics	+113.02%
(f) Machine learning	+75.11%
(d) Big data & Tech	+53.10%
(b) M2M communication	+53.07%
(a) Devices	-5.14%
(c) Security	-15.09%

https://doi.org/10.1371/journal.pone.0209472.t001

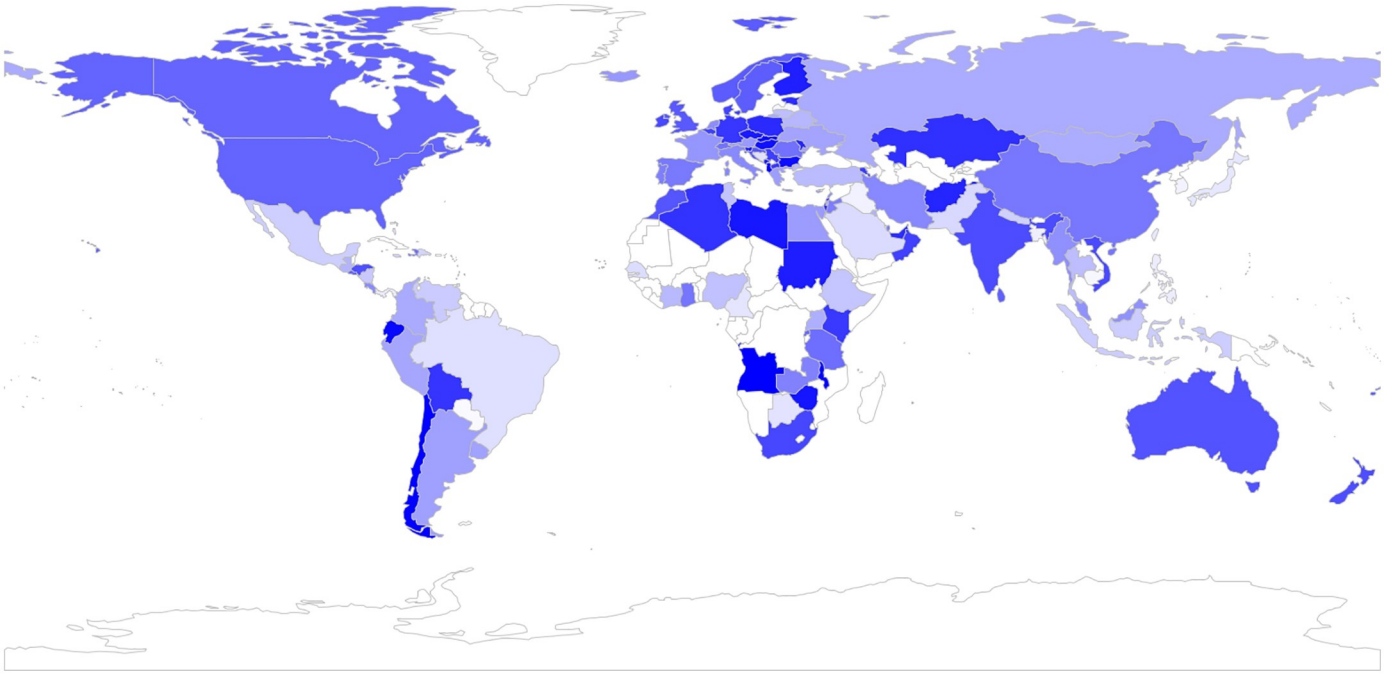

Fig 6. Tweets bearing positive sentiment, by country. Darker blue refers to more positive tweets.

https://doi.org/10.1371/journal.pone.0209472.g006

English) from a specific country are positive, where the countries with the highest percentage of positive tweets are coloured in darker blue. Likewise, Fig 7 shows the extent to which tweets posted from a country are positive, with countries with more negative tweets in darker red.

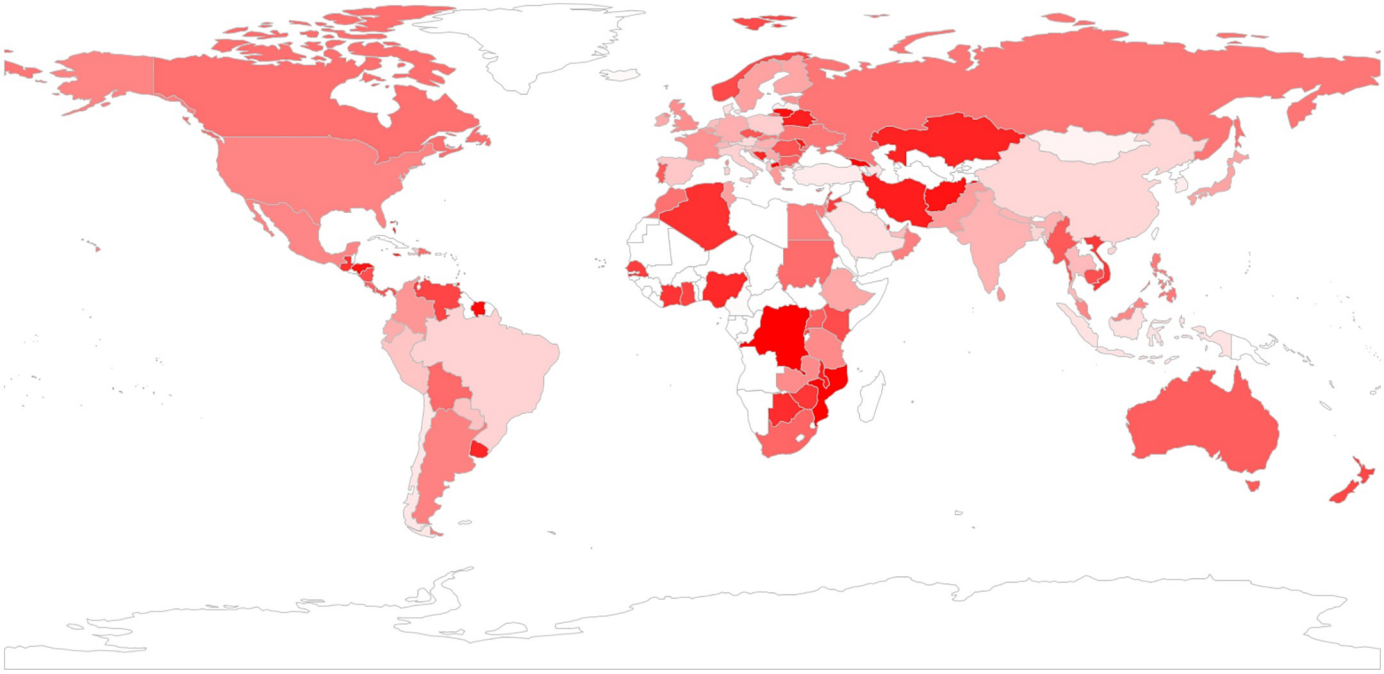

Fig 7. Tweets bearing negative sentiment, by country. Darker red refers to more negative tweets.

https://doi.org/10.1371/journal.pone.0209472.g007

Table 2. Top 10 most active countries, ranked by overall positivity towards the IoT.

	Overall	Big data & Tech	Security
Germany	+92.84%	+162.28%	+16.21%
Indonesia	+72.32%	+130.85%	+15.29%
Spain	+64.01%	+55.46%	+30.01%
Italy	+62.17%	+77.36%	+11.29%
United Kingdom	+46.32%	+90.14%	-15.68%
United States	+42.16%	+96.08%	-18.48%
Netherlands	+38.02%	+79.75%	-5.02%
Canada	+26.21%	+59.42%	-28.04%
Australia	+23.04%	+62.63%	-32.39%
France	+16.52%	+44.26%	-26.75%

https://doi.org/10.1371/journal.pone.0209472.t002

We focus our analysis on the top 10 most active countries in terms of their number of tweets, and look at their positivity, computed as the percentage of positive tweets that exceed negative tweets. The top 10 most active countries are listed in Table 2, ranked by overall positivity. The table also shows the positivity expressed by these countries towards the two most popular topics, i.e. "Big data & Tech" and "Security". All of the top 10 countries show, overall, more positivity than negativity, as shown by the positive values in the 'overall' column. This positivity varies substantially, however, from over 92% for the most positive country, Germany, to a 16.5% for the least positive country, France. The positivity also varies by topic, with an overall tendency to be more positive towards "Big data & Tech" across different countries, and an overall tendency to be more negative towards "Security." Germany stands out as one of the most positive countries, especially for "Big data & Tech." The most positive country towards "Security" is, however, Spain. While all of the countries are clearly positive towards "Big data & Tech," six of the top 10 countries are predominantly negative towards "Security," especially Australia, Canada and France, followed by the United Kingdom and the United States, and to a lesser extent by the Netherlands.

Understanding why a country such as Indonesia has a similar net positive view of IoT like Germany, whereas its neighbour, France, has a much different view may seem challenging. However, one explanation that has emerged is that positivity to IoT might reasonably be thought to correlate with a country's readiness for Industry 4.0. To examine this we have looked at the contribution to a country's GDP provided by manufacturing. The reason for considering this is that is recognised that some countries "have embraced the fourth industrial revolution more quickly and fully than many of their counterparts" because "manufacturing has consistently been seen as a critical part of those nations' economies" [38]. We collected statistics of the measurement from the World Bank [39], and the details of the statistics for the Top 10 most active countries is illustrated in Fig 8. We used the latest yearly data available for each of the Top 10 most active countries, which is 2014 for Canada, 2016 for United States and 2017 for all other eight countries. We tested also the results based on latest year which all countries data are available (the year 2014), which provides a similar result. Accepting that the contribution to a country's GDP provided by manufacturing is representative of its Industry 4.0 readiness, we can infer from Fig 8, that there is a positive correlation between people's overall positivity towards the IoT and the country's readiness to adopting Industry 4.0. This conjecture is confirmed by an additional statistical test, showing the correlation coefficient between the two factors is +0.889. Furthermore, if we assume that the readiness of a country is the causal factor for the overall positivity towards the IoT, a one percent increase in how ready a

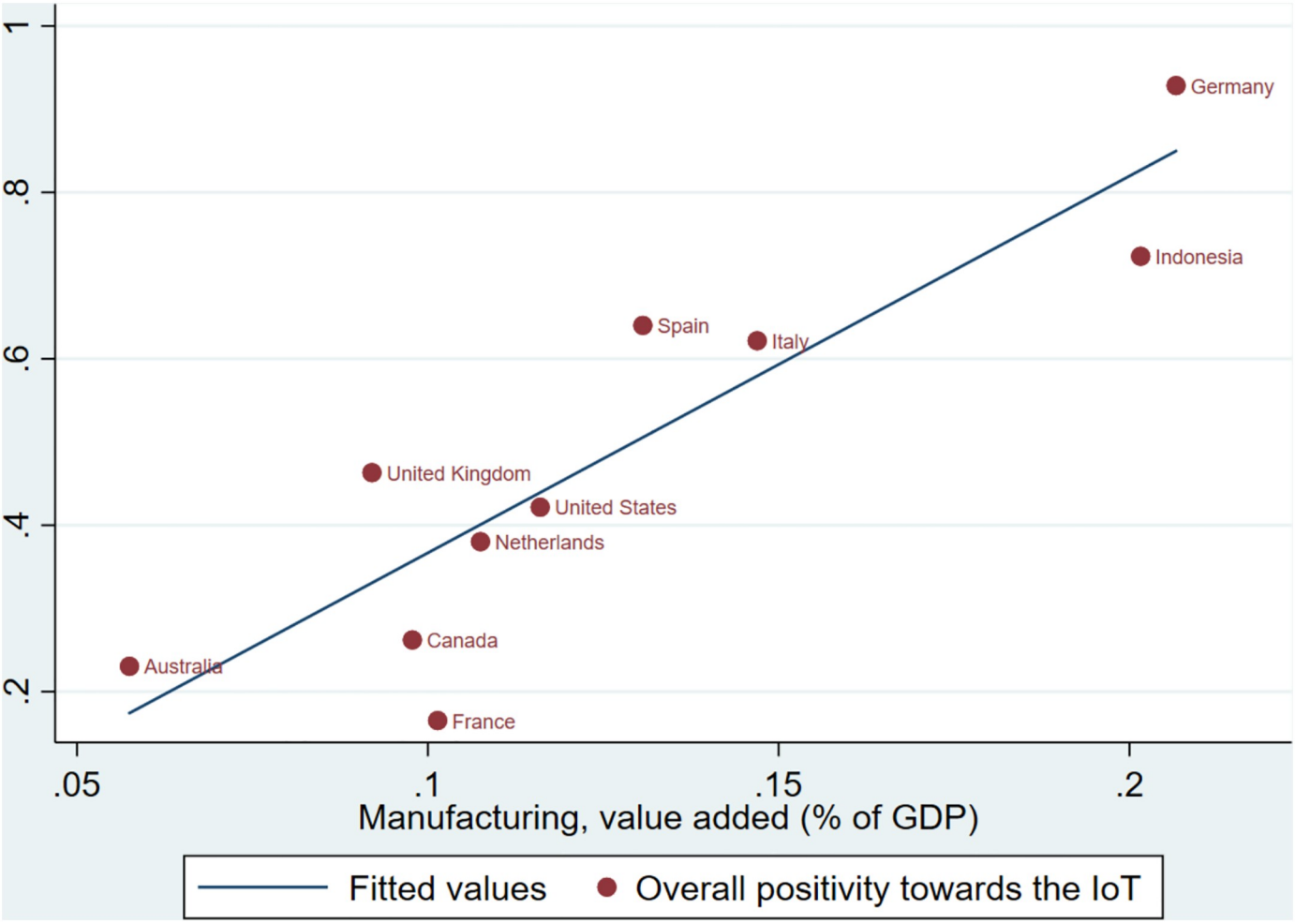

Fig 8. Overall positivity towards the IoT, by value added from manufacturing.

https://doi.org/10.1371/journal.pone.0209472.g008

country to adopting IoT changes is estimated to lead a 5.52 percent significant increase to the overall positivity (p = 0.001).

Gender

Processing of the tweets to identify those posted by accounts that could be identified as being male or female based on their first name led to a subset of 2,602,138 tweets. Given the imbalance in the number of tweets by gender (male: 1,999,934, female: 602,204), we normalised the figures for positive and negative tweets by the number of tweets posted by gender. Figs 9 and 10 show the trends in positivity and negativity for both males and females. Here we show the ratio of positive, neutral and negative tweets posted by males and females. While the charts show an overall similar tendency, an overall comparison of sentiment ratios reveals that, on average, females 9% more likely to share positive comments about IoT than males, whereas males are 10% more likely to express negativity, where percentages represent the differences between ratios for each gender. If we analyse the ups and downs in the trends, however, there is some correlation ($\rho = 0.48$ in the positive tweets and $\rho = 0.35$ in the negative tweets). Despite the similarity in the trends for both genders, the overall tendency is for females to be more

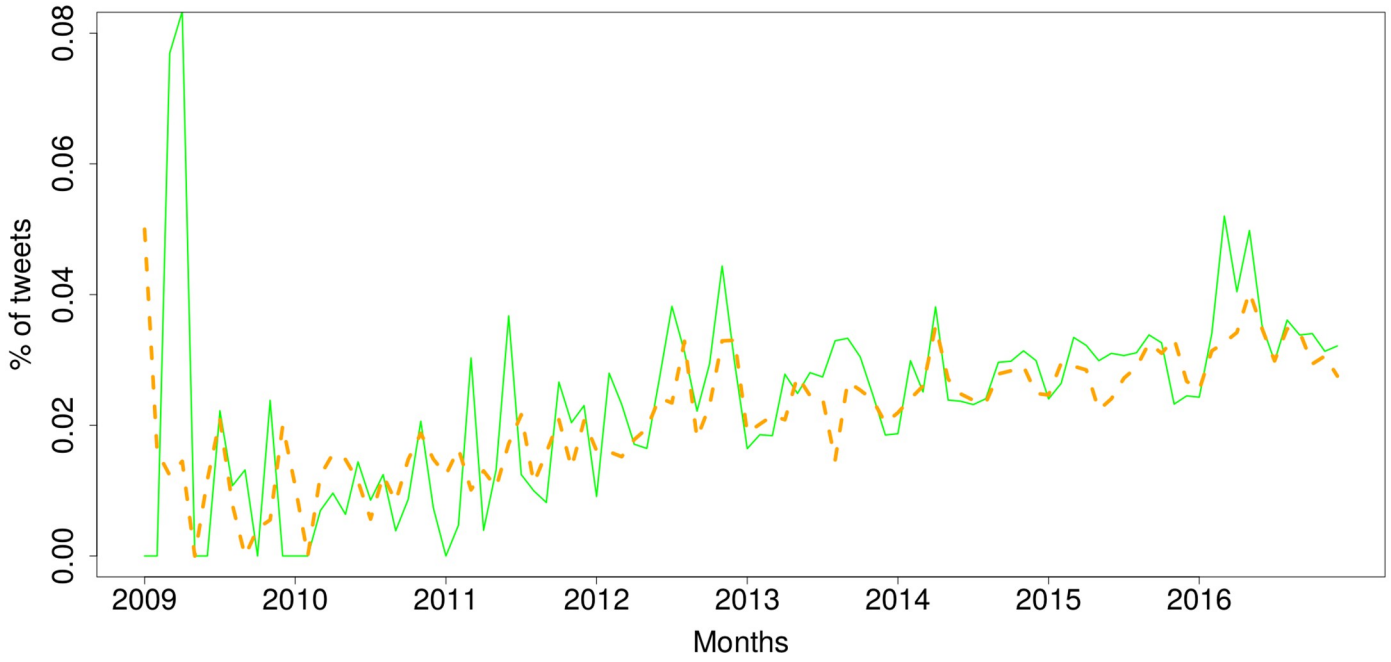

Fig 9. Tweets bearing positive sentiment, broken down by gender (green, solid line: Female, orange, dashed line: Male).

https://doi.org/10.1371/journal.pone.0209472.g009

positive about the IoT. While some studies have shown that males generally are more positive about technology than females, there is previous research on gender attitudes to technology may explain our findings. Almost 20 years ago a study found that, contrary to earlier studies, females "held more positive attitudes than males regarding the value of computers to make

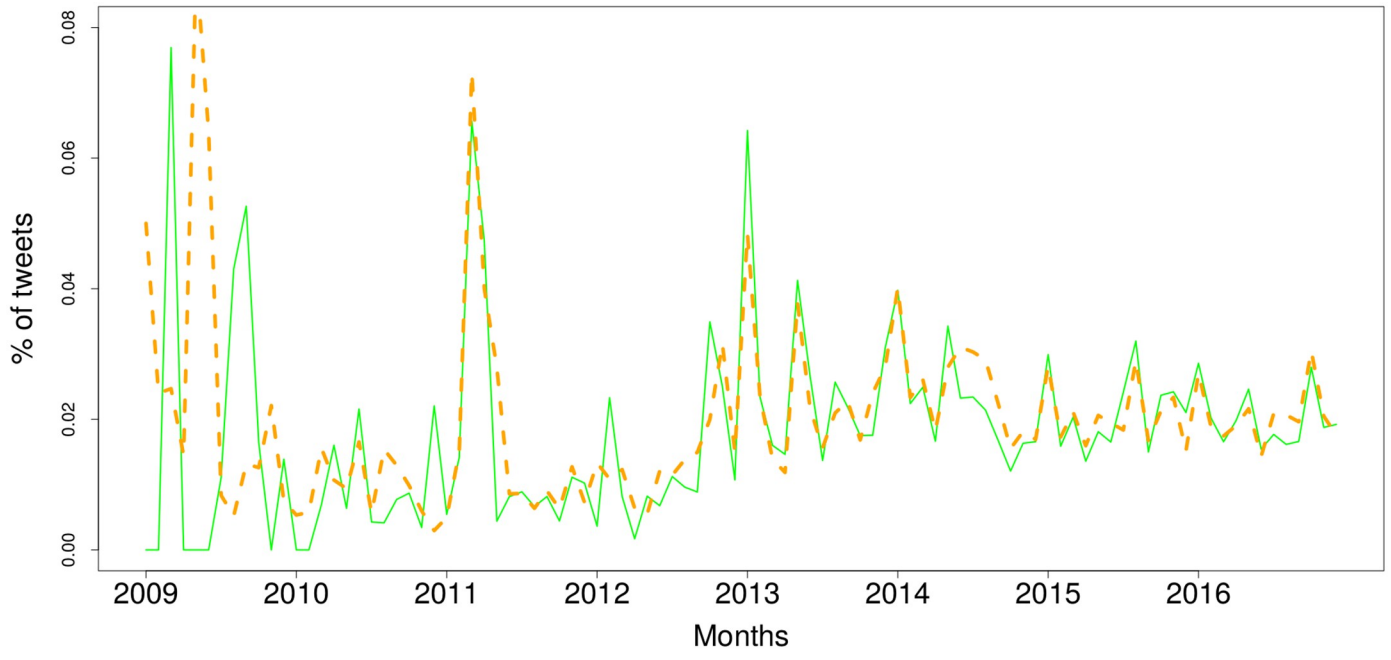

Fig 10. Tweets bearing negative sentiment, broken down by gender (green, solid line: Female, orange, dashed line: Male).

https://doi.org/10.1371/journal.pone.0209472.g010

users more productive" [40]. Obviously technology has significantly changed in 20 years, and while there has been no significant research into the impact of gender on opinion of the Internet of Things, a recent study that investigated a new model for self-management of psychological stress, based upon the use of an app found that "females rated the application more positive overall than [. . .] male participants" [41].

News stories

News stories can play an important role in shaping public opinion. Our analysis of news stories associated with IoT tweets was based on the external links that users point to in their tweets. As described above, we collect the final URL for all the links in tweets, as many of the links are shortened. In this analysis, we examined the sentiment associated with different links in tweets. First, we looked at the top news stories with the highest number of positive or negative tweets by year, which we show in Table 3. Again, a look at the top news stories suggested that the positive news stories were largely associated with the business opportunities that the IoT provides, whereas negative news were predominantly about security issues, devices being hacked, discussion on the public being scared by the IoT and the risks of the inevitable use of the cloud for the correct functioning of IoT devices.

The data shown in Fig 5 show that there are a number of spikes, showing a higher percentage of positive or negative tweets. An example is the publication of the Wired article "The Internet Of Things Has Arrived—And So Have Massive Security Issues" that was released 11 January 2013, during the Consumer Electronics Show, CES 2013 [42]. Looking closely at the tweets for that month, a lot of them related to this article highlighting concerns. This was a contributing factor to the spike in negativity observed in January 2013.

Discussion and conclusions

We have performed a large-scale, longitudinal analysis of the public perception of the different opportunities and challenges presented by the Internet of Things (IoT), which we have achieved by mining data from the social media platform Twitter. We have used the topic modelling algorithm LDA to identify the six main topics discussed by the public around the IoT. We analysed the polarity of tweets by using a state-of-the-art target-specific sentiment

Table 3. Top news headlines bearing positive or negative sentiment by year.

Year	Positive	Negative
2009	Internet of Things (IoT)—When do you think it will become a business reality?	Are you scared of the Internet of Things? Do RFID chips keep you awake at night, in unholy fear?
2010	That's it, somebody has gone and coined "Web 3.0—The Internet of Things"—. . .let the insanity begin.:/	This internet of things can be dangerous—hacker disables 100 cars remotely
2011	Inspiring The Internet Of Things: A Comic Book: The Internet of Things is one of our favorite trends at RWW.	How the Internet of Things is Changing the Way We Work
2012	Enjoying the internet of things? Thank your smartphone.	A french startup is disrupting the biz—Does the internet of things need its own internet?
2013	WiFi Bunnies and Why I Love the Internet of Things	Why the internet of things has to be not too smart and not too dumb, but just right
2014	The internet of things is great for chipmakers and a challenge for Intel	Why the Internet of Things narrative has to change. #Internet_of_things.
2015	IoT is on the cusp of something great, do you have a strategy in place?	#InternetofThings to cause major #security headaches
2016	#IoT-enabled devices with connectivity to #SAP is a great example of innovation in #IT	Why a Marriage Between the Cloud and Internet of Things Is Inevitable

https://doi.org/10.1371/journal.pone.0209472.t003

analysis algorithm, and by further exploring other dimensions such as country of origin and gender.

Among the six topics we identified in the dataset, the two most popular topics include "Big data & Tech" and "Security." This reveals that despite the business interest that the IoT presents for big data analytics, the challenges posed by the limited security of today's IoT devices are a major concern for the general public. We further confirm this with the predominantly negative sentiment that is associated with posts discussing security issues. This is again further confirmed when we look at the top news stories shared each year, with negative news being predominantly about security issues associated with IoT devices. Our study raises awareness on the importance of keeping IoT devices secure, reminding manufacturers that it is a concern that is being continually discussed.

A finer-grained analysis shows, however, variations across countries. While some countries like Germany, Indonesia and Spain tend to be generally positive about the IoT, others such as the United Kingdom, United States, Australia and France are not as positive, with a significant negative tendency towards security issues. We also observed some differences across gender groups, women being 9% more likely positive about IoT and men being 10% more likely negative.

The main finding that may impact on the large scale adoption of IoT is the public concern associated with security and IoT, suggesting that further effort is needed on the part of IoT device and service providers to convince the public that the IoT is—or will be—secure. [43, 44].

Future work

As is the case with Twitter users in general, our dataset is not a representative sample of the general population, hence the results cannot be treated as being a true reflection of public opinion. To address this issue, we are planning to use additional information extraction techniques to help determine key socio-demographic variables, such as age and education [45].

We also intend to use the findings of this work to inform further qualitative and quantitative research in a mixed-paradigm research study [46] on perceptions of the IoT, following established principles [47]. Clearly since Twitter posts are restricted in length, it is difficult to ascertain in depth information on the thoughts being presented. Using this work we intend to explore further, through focus groups, why there is a difference in perception between Analytics and Big Data and Devices and Security. Analysis of these workshops will allow us to formulate a nationally representative survey to determine national perceptions on IoT. This work would build upon our analysis and seek to find answers to open questions identified in the literature, such as [19].

Furthermore, the tools we have used are independent of the topic and thus applicable to other datasets of English tweets. In the future, we aim to collect other datasets to analyse public opinion on a range of issues of public interest, including political issues such as the United Kingdom's decision to leave the European Union (Brexit) and major societal issues such as abortion.

Acknowledgments

This work is part of the PETRAS project, which is funded by a grant from the UK Engineering and Physical Sciences Research Council (EP/N02334X/1). We wish to thank the Alan Turing Institute for its support (EPSRC grant EP/N510129/1).

Author Contributions

Conceptualization: Arkaitz Zubiaga, Rob Procter.

Data curation: Arkaitz Zubiaga.

Formal analysis: Arkaitz Zubiaga.

Funding acquisition: Rob Procter, Carsten Maple.

Investigation: Arkaitz Zubiaga.

Methodology: Arkaitz Zubiaga, Rob Procter.

Project administration: Rob Procter, Carsten Maple.

Writing – original draft: Arkaitz Zubiaga.

Writing – review & editing: Rob Procter, Carsten Maple.

References

1. Xia F, Yang LT, Wang L, Vinel A. Internet of things. International Journal of Communication Systems. 2012; 25(9):1101. https://doi.org/10.1002/dac.2417
2. Gubbi J, Buyya R, Marusic S, Palaniswami M. Internet of Things (IoT): A vision, architectural elements, and future directions. Future generation computer systems. 2013; 29(7):1645–1660. https://doi.org/10.1016/j.future.2013.01.010
3. Reiter G. Wireless connectivity for the Internet of Things. Europe. 2014; 433:868MHz.
4. Osseiran A, Elloumi O, Song J, Monserrat JF. Internet of Things. IEEE Communications Standards Magazine. 2017; 1(2):84–84. https://doi.org/10.1109/MCOMSTD.2017.7992936
5. Xu T, Wendt JB, Potkonjak M. Security of IoT systems: Design challenges and opportunities. In: Proceedings of the 2014 IEEE/ACM International Conference on Computer-Aided Design. IEEE Press; 2014. p. 417–423.
6. Farooq MU, Waseem M, Khairi A, Mazhar S. A critical analysis on the security concerns of internet of things (IoT). International Journal of Computer Applications. 2015; 111(7). https://doi.org/10.5120/19547-1280
7. Hwang YH. Iot security & privacy: threats and challenges. In: Proceedings of the 1st ACM Workshop on IoT Privacy, Trust, and Security. ACM; 2015. p. 1–1.
8. Zhao K, Ge L. A survey on the internet of things security. In: Computational Intelligence and Security (CIS), 2013 9th International Conference on. IEEE; 2013. p. 663–667.
9. Tellez M, El-Tawab S, Heydari HM. Improving the security of wireless sensor networks in an IoT environmental monitoring system. In: Systems and Information Engineering Design Symposium (SIEDS), 2016 IEEE. IEEE; 2016. p. 72–77.
10. Stanislav M, Beardsley T. Hacking iot: A case study on baby monitor exposures and vulnerabilities. Rapid 7. 2015.
11. Miessler D. IoT Attack Surface Mapping. DEFCON; 2015.
12. Islam SR, Kwak D, Kabir MH, Hossain M, Kwak KS. The internet of things for health care: a comprehensive survey. IEEE Access. 2015; 3:678–708. https://doi.org/10.1109/ACCESS.2015.2437951
13. Jing Q, Vasilakos AV, Wan J, Lu J, Qiu D. Security of the internet of things: Perspectives and challenges. Wireless Networks. 2014; 20(8):2481–2501. https://doi.org/10.1007/s11276-014-0761-7
14. Granjal J, Monteiro E, Silva JS. Security for the internet of things: a survey of existing protocols and open research issues. IEEE Communications Surveys & Tutorials. 2015; 17(3):1294–1312. https://doi.org/10.1109/COMST.2015.2388550
15. Sadeghi AR, Wachsmann C, Waidner M. Security and privacy challenges in industrial internet of things. In: Design Automation Conference (DAC), 2015 52nd ACM/EDAC/IEEE. IEEE; 2015. p. 1–6.
16. Maple C. Security and privacy in the internet of things. Journal of Cyber Policy. 2017; 2(2):155–184. https://doi.org/10.1080/23738871.2017.1366536
17. Weber RH. Internet of Things–New security and privacy challenges. Computer law & security review. 2010; 26(1):23–30. https://doi.org/10.1016/j.clsr.2009.11.008

18. Atzori L, Iera A, Morabito G. The internet of things: A survey. Computer networks. 2010; 54(15):2787–2805. https://doi.org/10.1016/j.comnet.2010.05.010
19. Asplund M, Nadjm-Tehrani S. Attitudes and Perceptions of IoT Security in Critical Societal Services. IEEE Access. 2016; 4:2130–2138. https://doi.org/10.1109/ACCESS.2016.2560919
20. Joseph N, Kar AK, Ilavarasan PV, Ganesh S. Review of Discussions on Internet of Things (IoT): Insights from Twitter Analytics. Journal of Global Information Management (JGIM). 2017; 25(2):38–51. https://doi.org/10.4018/JGIM.2017040103
21. Bian J, Yoshigoe K, Hicks A, Yuan J, He Z, Xie M, et al. Mining Twitter to Assess the Public Perception of the "Internet of Things". PloS one. 2016; 11(7):e0158450. https://doi.org/10.1371/journal.pone.0158450 PMID: 27391760
22. curl: command line tool and library for transferring data with URLs;. Accessed: 2018-10-10. https://curl.haxx.se/.
23. Zubiaga A, Voss A, Procter R, Liakata M, Wang B, Tsakalidis A. Towards real-time, country-level location classification of worldwide tweets. IEEE Transactions on Knowledge and Data Engineering. 2017;. https://doi.org/10.1109/TKDE.2017.2698463
24. GitHub: python code for tweet country classification;. Accessed: 2018-10-10. https://github.com/azubiaga/tweet-country-classification.
25. Figshare: Tweet geolocation 5m dataset;. Accessed: 2018-10-10. https://figshare.com/articles/Tweet_geolocation_5m/3168529.
26. Python package SexMachine;. Accessed: 2018-10-10. https://pypi.org/project/SexMachine/.
27. Wang B, Liakata M, Zubiaga A, Procter R. TDParse-multi-target-specific sentiment recognition on Twitter. In: Proceedings of the European Chapter of the Association for Computational Linguistics; 2017. p. 483–493.
28. Wang B, Liakata M, Tsakalidis A, Georgakopoulos Kolaitis S, Papadopoulos S, Apostolidis L, et al. TOTEMSS: Topic-based, Temporal Sentiment Summarisation for Twitter. In: Proceedings of the 8th International Joint Conference on Natural Language Processing, IJCNLP; 2017.
29. Blei DM, Ng AY, Jordan MI. Latent dirichlet allocation. Journal of machine Learning research. 2003; 3(Jan):993–1022.
30. gensim: Topic modelling for humans;. Accessed: 2018-10-10. https://radimrehurek.com/gensim/.
31. Wang B, Liakata M, Zubiaga A, Procter R. A Hierarchical Topic Modelling Approach for Tweet Clustering. In: International Conference on Social Informatics. Springer; 2017. p. 378–390.
32. Hutto C, Gilbert E. VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text. In: Eighth International AAAI Conference on Weblogs and Social Media; 2014.
33. Ferrara E, Varol O, Davis C, Menczer F, Flammini A. The rise of social bots. Communications of the ACM. 2016; 59(7):96–104. https://doi.org/10.1145/2818717
34. Castell S, Cameron D, Ginnis S, Gottfried G, Maguire K. Public Views of Machine Learning. Ipsos MORI; 2017.
35. Schreiber R, Lawton T. Comparing Global Industry 4.0 Readiness;. https://industrytoday.com/article/comparing-global-industry-4-0-readiness/.
36. Fast E, Horvitz E. Long-Term Trends in the Public Perception of Artificial Intelligence. In: AAAI; 2017. p. 963–969.
37. Castell S, Charlton A, Clemence M, Pettigrew N, Pope S, Quigley A, et al. Public attitudes to science 2014. London, Ipsos MORI Social Research Institute. 2014; 194:28.
38. Sample I. Artificial intelligence survey finds UK public broadly optimistic;. https://www.theguardian.com/technology/2017/apr/25/artificial-intelligence-survey-finds-uk-public-broadly-optimistic-mass-unemployment.
39. World Bank Open Data;. Accessed: 2018-10-10. https://data.worldbank.org/.
40. Ray CM, Sormunen C, Harris TM. Men's and women's attitudes toward computer technology: A comparison. Office Systems Research Journal. 1999; 17:1–8.
41. Wiederhold BK, Boyd C, Sulea C, Gaggioli A, Riva G. Marketing analysis of a positive technology app for the self-management of psychological stress. Stud Health Technol Inform. 2014; 199:83–87. PMID: 24875696
42. Rose A. The Internet of Things Has Arrived—and so Have Massive Security Issues;. https://www.wired.com/2013/01/securing-the-internet-of-things/.
43. Li S, Da Xu L. Securing the Internet of Things. Syngress; 2017.
44. Yager RR, Espada JP. New Advances in the Internet of Things. Springer; 2017.

45. Rao D, Yarowsky D, Shreevats A, Gupta M. Classifying latent user attributes in twitter. In: Proceedings of the workshop on Search and mining user-generated contents; 2010. p. 37–44.

46. Sale JE, Lohfeld LH, Brazil K. Revisiting the quantitative-qualitative debate: Implications for mixed-methods research. Quality and quantity. 2002; 36(1):43–53. https://doi.org/10.1023/A:1014301607592 PMID: 26523073

47. Venkatesh V, Brown SA, Bala H. Bridging the qualitative-quantitative divide: Guidelines for conducting mixed methods research in information systems. MIS quarterly. 2013; 37(1). https://doi.org/10.25300/MISQ/2013/37.1.02