Original Paper

Using Twitter to Understand the Human Bowel Disease Community: Exploratory Analysis of Key Topics

Martín Pérez-Pérez 1,2,3, PhD; Gael Pérez-Rodríguez 1,2,3, PhD; Florentino Fdez-Riverola 1,2,3, PhD; Anália Lourenço 1,2,3,4, PhD

1 Department of Computer Science, University of Vigo, Escuela Superior de Ingeniería Informática, Ourense, Spain
2 Biomedical Research Centre, Campus Universitario Lagoas-Marcosende, Vigo, Spain
3 Next Generation Computer Systems Group, School of Computer Engineering, Galicia Sur Health Research Institute, Galician Health Service - University of Vigo, Vigo, Spain
4 Centre of Biological Engineering, Campus de Gualtar, University of Minho, Braga, Portugal

Corresponding Author:
Anália Lourenço, PhD
Department of Computer Science
University of Vigo
Escuela Superior de Ingeniería Informática
Edificio Politécnico
Campus Universitario As Lagoas s/n
Ourense, 32004
Spain
Phone: 34 988 387 013
Fax: 34 988 387 001
Email: analia@uvigo.es

Abstract

Background: Nowadays, the use of social media is part of daily life, with more and more people, including governments and health organizations, using at least one platform regularly. Social media enables users to interact among large groups of people that share the same interests and suffer the same afflictions. Notably, these channels promote the ability to find and share information about health and medical conditions.

Objective: This study aimed to characterize the bowel disease (BD) community on Twitter, in particular how patients understand, discuss, feel, and react to the condition. The main questions were as follows: Which are the main communities and most influential users?; Where are the main content providers from?; What are the key biomedical and scientific topics under discussion? How are topics interrelated in patient communications?; How do external events influence user activity?; What kind of external sources of information are being promoted?

Methods: To answer these questions, a dataset of tweets containing terms related to BD conditions was collected from February to August 2018, accounting for a total of 24,634 tweets from 13,295 different users. Tweet preprocessing entailed the extraction of textual contents, hyperlinks, hashtags, time, location, and user information. Missing and incomplete information about the user profiles was completed using different analysis techniques. Semantic tweet topic analysis was supported by a lexicon-based entity recognizer. Furthermore, sentiment analysis enabled a closer look into the opinions expressed in the tweets, namely, gaining a deeper understanding of patients’ feelings and experiences.

Results: Health organizations received most of the communication, whereas BD patients and experts in bowel conditions and nutrition were among those tweeting the most. In general, the BD community was mainly discussing symptoms, BD-related diseases, and diet-based treatments. Diarrhea and constipation were the most commonly mentioned symptoms, and cancer, anxiety disorder, depression, and chronic inflammations were frequently part of BD-related tweets. Most patient tweets discussed the bad side of BD conditions and other related conditions, namely, depression, diarrhea, and fibromyalgia. In turn, gluten-free diets and probiotic supplements were often mentioned in patient tweets expressing positive emotions. However, for the most part, tweets containing mentions to foods and diets showed a similar distribution of negative and positive sentiments because the effects of certain food components (e.g., fiber, iron, and magnesium) were perceived differently, depending on the state of the disease and other personal conditions of the patients. The benefits of medical cannabis for the treatment of different chronic diseases were also highlighted.
Conclusions: This study evidences that Twitter is becoming an influential space for conversation about bowel conditions, namely, patient opinions about associated symptoms and treatments. So, further qualitative and quantitative content analyses hold the potential to support decision making among health-related stakeholders, including the planning of awareness campaigns.

(J Med Internet Res 2019;21(8):e12610) doi:10.2196/12610

KEYWORDS
inflammatory bowel diseases; irritable bowel syndrome; social media; communication; data mining; natural language processing

Introduction
Mass access and diffusion of health-related news and information have dramatically changed in recent years. In addition to traditional media, the internet has become a pivotal instrument for sharing knowledge [1]. In this context, the use of social media has increased exponentially over the last years. More and more people and institutions, including governments and health organizations, use social networks, blogs, content-sharing sites, and wikis on a regular basis [2-4]. Therefore, social media have become a major source of information [5]. Social media creates the opportunity for users to interact among large groups of people that share the same interests and suffer the same afflictions. In particular, these channels promote the ability to find and share information about health and medical conditions [6,7]. Notably, many people who suffer from chronic diseases resort to groups or communities in social networks to share experiences and stand by for news about their afflictions [8-10]. One recurrent example is inflammatory bowel disease (IBD), a chronic, relapsing, and remitting autoimmune disorder with 2 main conditions, Crohn disease and ulcerative colitis [11]. The worldwide incidence of these conditions has been increasing over the last few decades [12]. IBD is frequently diagnosed in the second to fourth decades of life, with a high incidence during the peak female reproductive years [13]. Communities all around the world elected May 19 as the World IBD Day, a reference date to raise awareness of this chronic disease and its associated symptoms [14]. Nowadays, there is an increasing interest in analyzing and studying several factors related to health topics in social media, especially on Twitter [15-17]. Namely, text mining (TM) and natural language processing methods and techniques are being applied to systematically identify the topics under discussion as well as study the relationships among individuals and topics. Indeed, over recent years, an increasing number of publications reported health-related studies in social media. For example, how promotional health information about Lynch syndrome impacts laypeople’s discussions [18]; diabetes-related participation on Twitter by describing the frequency and timing of diabetes-related tweets, the geography of tweets, and the types of participants [19]; understanding the use of social media by patients with different types of cancer [17]; and the emergence of health online communities of practice (ie, group of people who share experiences) [16]. Regarding IBD, only a few studies exist, and these are focused on the analysis of particular user communities or particular user details. For example, the study by Guo et al studied the use and quality of social media in patients with IBD [20], and the study by Keller et al discussed how individuals taking IBD medication during reproductive periods made decisions about their medication use [21]. So, the aim of this study was to gain a better understanding of the communities involved in the dissemination of information about bowel disease (BD). A collection of over 24,000 tweets related to BD enabled the identification and geolocation of active users and the analysis of external events that determine tweet content and discussion, as well as the relations between the biomedical and scientific topics being mentioned. The obtained results are useful to understand the most relevant topics and communities and, specifically, how patients discuss, feel, and react to symptoms, changes in habits, and medication. That is, to learn how to enhance the dissemination of information and raise awareness among patients, which are the 2 main objectives of health-related stakeholders.

Methods
Twitter Communication Environment
Currently, the architecture of Twitter supports different user actions in response to a tweet (ie, replies, retweets, and favorites), each of them holding a specific meaning in terms of communication capabilities. Notably, replies represent the specific response to a sent tweet, retweets stand for the reposting of tweets (which is useful to quickly share and promote content), and favorites indicate that the content of a tweet is highly appreciated by the community and can be seen as a user tweet bookmark. Figure 1 exemplifies the aforementioned communication modes. On the one hand, replies and retweets (ie, user relations) allow identifying users who support or discuss any sent message. For example, users B and C are interested in a tweet published by user A. This allows to identify how the information is spread and which are the most influential users. On the other hand, the number of retweets and favorites (ie, tweet interactions) helps to measure the relevance of the new content to the community. For example, user D likes the original tweet sent by user A and saves it for future reference.

http://www.jmir.org/2019/8/e12610/
**General Workflow**

Figure 2 depicts the workflow implemented in this study to retrieve, process, and analyze BD-related tweets, which consisted of 2 fundamental phases: (1) data collection and filtering and (2) corpus processing and analysis.

From February 1, 2018, to August 31, 2018, tweet data were retrieved via the Twitter application programming interface (API). Tweet contents and associated information were processed: whenever possible, users were geolocated, their gender was determined, and they were identified as organization or patient; tweet contents were cleaned (hashtags and mentions to user accounts were removed) and prepared for further text processing; and, entity recognition and sentiment analysis were applied.
Data Collection and Filtering

Data collection accounted for tweets containing terms as Inflammatory bowel disease, Irritable bowel disease, Irritable colon, Ulcerative colitis, Ileocolitis, Ileitis, Crohn, Granulomatous, and Jejunoileitis. The Java library Twitter4J [22] was used to perform such collection. From all the retrieved tweets, only those written in English were considered to ensure the consistency of further examination. A total of 4.10% (1055/25,689) tweets written in other languages, such as French, Spanish, or Italian, were eliminated. The final dataset comprised 24,634 unique tweets written in English by 13,295 different users.

Corpus Processing and Analysis: User Characterization

All the tweets were automatically labeled by tweet creator. Data present in user profiles were collected and further validated (see details in the next subsections). Specifically, user characterization involved gender determination, differentiation of organizations and patients, and geolocation.

Unfortunately, it was not possible to determine the age of the users because of the low precision of current prediction models and because several studies point that most Twitter users fall into a small range of years [23].

Face and Gender Recognition

Gender identification entailed a 2-step strategy based on a gender-name dictionary [24] and a convolutional neural network model (this model was trained over more than 500,000 public face images extracted from IMDb and Wikipedia and was reported to have nearly a 90% of accuracy) [25]. First, the user name was checked against the dictionary. If there was a perfect match, that is, a unique gender associated with the name, the gender was resolved. Otherwise, and if there was a user profile picture, the deep learning model was applied. If there was no user profile picture, or the model could not recognize a single face in the image, the gender was set as unknown.

User Identification

Whenever possible, user accounts were categorized as belonging to organization, individual (ie, patient and medical expert), and unknown. The strategy to identify users was focused on the analysis of the user account (ie, name and description).

First, to identify users as organizations, several cascade steps were followed, namely, (1) the account had a country, country code, or a continent in the user name; (2) the account had a URL domain in the name (eg, .org); and (3) usage of regular expressions to check for nonpersonal keywords in the description (eg, official, news, info, or pharma).

If the user was not identified as an organization, the following steps were applied to check if the user was an individual: (1) the account had a recognized gender; (2) the account was recognized by Twitter as a contributor or a translator; (3) the description was written using first-person pronouns and their variant forms (ie, possessive and reflexive); (4) the description had emojis or emoticons; and (5) usage of regular expression to check for person abbreviations (eg, Ms or Mr).

Finally, to differentiate BD experts (eg, doctors, medical staff, or researchers) from patients, all accounts identified as individuals were processed with an additional recognition step to check for expert-related keywords (eg, Dr, Prof, MD, or PhD). Whenever the strategy could not help determine the type, the user was labeled as an unknown type.

Geolocation

Twitter does not require users to specify the location. When users introduce such data, it is in the form of free text, which often raises consistency issues in further analysis (eg, a user can enter NYC and others may identify the same location as New York City). Another issue to take into account is the existence of cities in different countries that share the same name (eg, Guadalajara is a city both in Spain and Mexico).

Thus, the applied location identification method took into consideration information about the time zone, the Coordinated Universal Time (UTC) offset, and the location text. These data were used in combination with the GeoNames database [26], which contains over 10 million geographical names and is accessible through a free Web service. The data extracted from Twitter were searched against the GeoNames database. If the data were not accurate enough, that is, matching multiple database entries, the time zone and the UTC offset were used to help resolve the location. In those cases where different cities shared the same name and time zone but were located in different countries, the location was set as unknown.

Corpus Processing and Analysis: Tweet Characterization

To analyze the content of the generated corpus, it was essential to be able to properly recognize the relevant (topic related) terms mentioned in the tweets. For this purpose, several text preprocessing techniques were applied and then an in-house–developed named entity recognizer supported the annotation of terms pertaining to the selected BD-related semantic categories.

Data Cleaning

As a first step, the following preprocessing tasks were applied to the content of the tweets in the dataset:

- Removal of special characters that did not provide useful information (eg, &, (”, “), *, +, <, or >)
- Identification of replies and mentions to other users (represented with @) and extraction of URLs
- Removal of the symbol # in hashtags, and split of hashtags in multiple words (if possible) with the goal of revealing relevant terms to the analysis (eg, InflammatoryBowelDisease to Inflammatory Bowel Disease). All these operations were carried out using the Twitter text library [27]
- Deletion of repeated letters if there are 3, or more, consecutive and identical characters (eg, haaaappppy to haaaaappy)
- Correction of spelling errors using the Hunspell dictionary [28], a collection of specific medical terms [29] obtained from the OpenMedSpel [30], and the MTH-Med-Spel-Check tool [31]. The correction was done...
automatically by selecting the suggested word with the highest similarity with the original (incorrect) term. The similarity was calculated using the normalized Levenshtein algorithm [32].

- Expansion of abbreviations and shorthand terms, which were not included in the Hunspell dictionary (eg, SBBOS to small bowel bacterial overgrowth syndrome). Although Twitter has increased the maximum length of tweets from 140 to 280 characters, the use of abbreviations is still very common. Therefore, a custom dictionary of abbreviations was constructed in house, comprising terms extracted from multiple locations [33-35].

Then, a new round of text preprocessing tasks prepared the tweets for named entity recognition, namely, tokenization (ie, breaking a stream of text up into words, phrases, or other meaningful elements), stop word removal (ie, removal of too frequent, not content-bearing tokens), part of speech tagging (ie, to assign a lexical category to each token), and lemmatization (ie, to obtain the lexicem form of the token). Beside single word tokens (unigrams), bigrams and trigrams, that is, contiguous of 2 or 3 sequences of tokens, were also considered in entity recognition. All the aforementioned tasks were implemented using the Stanford CoreNLP pipeline [36].

Named Entity Recognition

The semantic lexicon applied in named entity recognition was mostly retrieved from the repository of biomedical ontologies BioPortal [37] as follows: the Human Disease Ontology (DOID) [38], which provides descriptions of human disease terms, phenotype characteristics, and related medical vocabulary; the Ontology For Nutritional Studies (FoodOn) [39], which covers human food raw ingredients, food products, and product types and develops semantics to food production, culinary, nutritional, and chemical ingredients and processes; the Symptom Ontology (SYMP) [40], which covers disease symptoms, with symptoms encompassing perceived changes in function, sensations, or appearance reported by a patient indicative of a disease; and the branch Intervention or Procedure of the National Cancer Institute Thesaurus (NCIT) [41], which describes treatments or actions taken to prevent or treat disease or improve health in other ways. The DrugBank ontology supported the recognition of chemical, pharmacological, and pharmaceutical terminology, that is, approved small molecule drugs, approved biotech (protein and peptide) drugs, nutraceuticals, and experimental drugs [42].

As a whole, the lexicon supporting the entity recognition encompassed a total of 217,468 term entries. For the sake of simplicity, the results of the semantic annotations are presented and discussed in terms of the meta categories, that is, Drag encloses all the classes encompassed by DrugBank, Food and Diet refers to the food ingredients and food products classified by FoodOn, Symptom relates to the symptoms as presented by SYMP, Treatment refers to the treatments classified by NCIT, and Disease groups together disease terms, phenotype characteristics, and related medical vocabulary, as described by DOID.

The named entity recognition pipeline was implemented in house and entailed dictionary lookup, as well as pattern- and rule-based recognition. To be able to match the lexicon with tweet contents, the lexicon required some processing, namely, convert all terms to lowercase, remove extra whitespaces, remove small and long terms (ie, less than 2 characters and terms longer than the maximum tweet length), replace special characters by a whitespace, and remove terms associated with more than one category.

An inverted recognition technique was used in actual entity recognition [43]. This technique uses the words in the text as patterns to be matched against the lexicon. This was a valid approximation for this study because the number of words in the tweets were much smaller than the number of terms in the lexicon, that is, a fewer number of patterns to match. Recognition preference was given to the longest possible n-grams. In addition, the recognizer accepted perfect matches as well as lexical variations of the terms (ie, lemmatized entries and abbreviations).

Sentiment Analysis

The sentiment of the tweets was analyzed using the Valence Aware Dictionary and sEntiment Reasoner (VADER) API for Python [44]. VADER is a lexicon- and rule-based sentiment analysis tool that is specifically attuned to sentiments expressed in social media. The predicted sentiment (ie, compound score) is computed by summing the valence scores of each word in the lexicon, adjusted according to emotion-related rules, and then normalized to have values between −1 (most extreme negative emotion) and +1 (most extreme positive emotion).

Network Representation and Analysis

Graph analysis was performed to measure the relevance of individual terms as well as term-term pairs. Specifically, this analysis was applied to user interactions, via mentions and retweets (ie, directed mentions and retweets), and co-occurrence of semantically meaningful terms (ie, whenever 2 terms were found in the same tweet, these 2 terms were considered to share a link).

Networks were generally described in terms of the number of nodes and edges, as well as metrics of degree, characteristic path length, clustering coefficient, and the average number of neighbors [45]. In more detail, graph connectedness was measured by degree centrality, betweenness centrality, and closeness centrality [46-48]. Briefly, degree centrality measured the total amount of direct links with the other nodes (ie, higher degree implies the node is more central), betweenness centrality measured the mediation role of the nodes (ie, if other nodes have to go through the node to ensure communication, then the node is likely important and has a high betweenness centrality), and closeness centrality measured the convenience and ease of connections between each node and the rest of the nodes (ie, if the average shortest path of the node is small, then the node has a high closeness centrality) [49,50].

The clustering coefficient was used to measure the degree to which nodes tend to cluster together. Evidence suggests that social network nodes tend to create tight groups, which are characterized by a relatively high density of links; this likelihood tends to be greater than the average probability of a link randomly established between 2 nodes [51,52].
Results

Overview
The Results section was structured following the logical order of the proposed questions. Figure 3 illustrates how particular research questions were answered by a certain analysis and identifies the main insights provided by each of these analyses. Most notably, results were structured such that the basic characterization of the BD-related tweets in terms of user activity and topics of interest were presented first and, then, the interrelation of topics in the conversations was detailed along with the visibility of external sources of information.

The ability to identify the most active users (ie, users that post more tweets) and topic-specific communities (eg, gluten-centered community) is pivotal to gain a better understanding about how to adapt or finetune the communication (ie, reach a broader audience or focus on a specific community) as well as discover influencers (ie, vessels of information distribution, within or across the communities). The analysis of user activity (different time windows may apply to different communities) is interesting as a means to plan the best time to post a new tweet (ie, when to expect major user attention).

The semantic annotation of tweet contents goes a step forward, providing knowledge about the topics being discussed (ie, individually and in combination). Likewise, the identification of tweets linking to external information sources is relevant to bring forward the visibility these sources are receiving through Twitter.

A health-related stakeholder is a typical example of someone that can benefit from the insights provided by the overall study. Say that the aim is to plan an awareness campaign about BD and food habits. Likely, this stakeholder wants to study the target audience in terms of topics that are attracting more attention and tweeting habits. The campaign may be planned to receive short-term attention (eg, announcing the launch of a novel drug or a new food supplement) or promote awareness throughout a longer period of time (eg, promote healthier food habits).

Figure 3. A roadmap showing the usefulness of the proposed analyses for the study of bowel diseases communities on Twitter. Q: question; R: result.

Bowel Disease Communities
User relations, that is, mentions and retweets, were represented in a network to study communication interplay, namely, to identify target communities and influential users (ie, individuals and/or institutions with high audiences). Figure 4 depicts this network such that the nodes denote the users (ie, account name), the node size is based on the node in-degree (ie, the users that received most communication are represented by bigger nodes), and the edges account for the number of mentions and retweets. The 4 different background areas in the figure denote the communities, mostly characterized by account descriptions: gluten-, nutritional-, BD-, and food-related relations. In this study, only communities with more than 5 connected users were considered.
Figure 4. User relation network representation. Two users are related if they share at least one mention or retweet. Node size is based on the node degree. Background color identifies the community.

In this network, the out-degree of the nodes (i.e., users initiating communication with other users) is much lower than the in-degree of the nodes (i.e., users receiving significant tweets). The user accounts @CrohnsColitisUK, @CrohnsColitisFn, and @HealioGastro, which belong to disease-specific organizations and have primarily informative/educational goals, were among the accounts receiving more communication (i.e., higher in-degree). This conveys the rationale that individuals typically prefer to ask for health information to trusted organizations [6,53,54]. Conversely, accounts of BD patients (e.g., @colitisandme) and experts in bowel conditions or nutrition (e.g., @IBDMD and @charlie_lees) were among those showing highest out-degree.

Public lists of BD-related influencers [55] supported the identification of organization accounts, such as @CrohnsColitisUK, @HealioGastro, and @ACCUcatalunya, and personal accounts, such as @colitisandme, @IBDMD, and @EdwardLoftus2, that are typically reached for medical advice. Figure 5 shows an example of a tweet exchange: the user @CrohnoId, a Crohn disease patient, asks the user @ibddoctor, an expert in BD, about 2 possible diagnostic techniques, that is, magnetic resonance imaging and computed tomography; @ibddoctor explained the advantages and disadvantages of these techniques, and another user @SandraZelinsky also intervened, pointing out limited access.
Figure 5. Example of a tweet exchange among an irritable bowel disease expert and 2 laymen.

Demographic Distribution and Hot Zones
Knowledge about how the community is demographically distributed is important to carry out public information campaigns and study the impact of government and institutional actions in different demographic areas. In this analysis, 59.98% of the users (7975/13,295) were geographically distributed all around the world.

Figure 6 shows the geographical distribution of the BD communities. The size of the nodes represents the number of users located in each country (ie, the bigger the circle, the higher is the number of users) and colors identify the continents (ie, blue for America, purple for Europe, brown for Africa, green for Asia, and red for Australia).

In general, most of the users were located in the United States, the United Kingdom, Canada, and Australia, which is consistent with current knowledge about the prevalence of these conditions [56]. The highest reported prevalence is in Europe (with the highest prevalence of ulcerative colitis in Norway and of Crohn disease in Germany) and North America (with the highest prevalence of ulcerative colitis in the United States and Crohn disease in Canada). The prevalence of IBD exceeded 0.3% in North America, Oceania, and many countries in Europe.

The number of users from India, Bangladesh, and the Philippines was also noticeable and may be explained by the rising incidence of BD conditions in newly industrialized countries in Africa, Asia, and South America.

Figure 6. Geographical distribution of the bowel disease-related users. The bigger the node, the higher the number of users in the country. Node color denotes the continent.
Biomedical and Scientific Topics

Understanding what the user community is talking about was important to identify the topics that received more attention and to be able to better align new information/communication strategies with the interests of the community. Table 1 presents the top 25 most mentioned terms along with their corresponding semantic category (ie, Food and Diet, Disease, Treatment, Symptom, and Drug) and sorted by the number of including tweets. Explicit mentions to IBS and IBD (eg, Inflammatory Bowel Disease or Ulcerative Colitis) and noncontent–bearing, generalist terms (eg, Disease or Food), were at the top of term mentions but were not listed in the table.

The #Tweets column indicates the number of tweets in which a term is mentioned, the #Favorites column indicates the number of times that a tweet containing a term was selected as a favorite, the #Retweets column indicates the number of times that a tweet containing a term was retweeted, and the #Hashtags column indicates the number of times that a term was used in hashtags, including term variations (eg, anxiety disorder like #anxiety).

The volume of retweets is useful to understand how the information flows among users, whereas the number of favorites can be understood as a metric to measure the usefulness of the tweets containing the term [57,58]. In addition, the number of times a term appears in hashtags is interesting to understand how the user wishes the tweet to be indexed (to be easily found by others with similar interests as well as to bring attention to certain topics). This can be understood as a metric to measure the impact of the term in the community [59].

As illustrated in Table 1 (and in the topological analysis of the co-occurrence network of terms in Multimedia Appendix 1), the most mentioned terms related to Disease (35.64%, 11,688/32,794) and Food and diet (25.43%, 8342/32,794) categories, followed by terms from Symptom (17.71%, 5811/32,794), Treatment (14.83%, 4864/32,794), and Drug (6.37%, 2089/32,794) categories, respectively.

Diarrhea (10.33%, 1208/11,688 of the mentions to disease terms) and constipation (8.94%, 1046/11,688 of the mentions to disease terms) were the most mentioned disease-related terms. This was somewhat expected considering that many IBD and IBS patients have diarrhea as a side effect of the disease and many others claim to have problems of constipation [60]. Interestingly, constipation is one of the terms most often included in hashtags, which indicates that the topic is meaningful and timely to the community. Other diseases typically associated with BD conditions were also discussed, notably cancer (3.37%, 395/11,688 of the mentions to disease terms), anxiety disorder (3.04%, 356/11,688 of the mentions to disease terms), depression (2.49%, 292/11,688 of the mentions to disease terms), arthritis (1.71%, 200/11,688 of the mentions to disease terms), and asthma (1.46%, 171/11,688 of the mentions to disease terms). Chronic inflammation is known to be a major risk factor for the development of gastrointestinal malignancies and is often associated with inflammatory arthritis. Notably, as the population of patients with IBD grows older, with longer periods of chronic inflammation and longer exposure to immunosuppression, there is an increased risk of developing cancer [61] and arthritis [62]. Moreover, studies show that the rates of anxiety and depression tend to be higher among patients with Crohn disease or ulcerative colitis compared with those with other diseases as well as the general population [63,64]. Finally, interest in discussing asthma is justified by the association between this disease and early- and late-onset ulcerative colitis, particularly because of shared environmental risk factors [65]. Posts discussing BD and other diseases often show high retweet rates, and it is noticeable that depression is currently at the center of attention (ie, those tweets have the greatest number of favorites).

Symptoms typically associated with the different stages of the disease such as bloating (4.40%, 256/5811 of the mentions to symptom terms), flatulence (3.20%, 186/5811 of the mentions to symptom terms), and abdominal pain (2.97%, 173/5811 of the mentions to symptom terms) were also discussed with considerable frequency and the containing tweets were among the most retweeted, that is, people within this community find it relevant to broadcast information related with the symptomatology of BD conditions (eg, symptom-disease evidence and symptom relief therapies) [60].

The BD community is also sharing information about diet-based treatments and shows particular interest in gluten-free dietary interventions [66] and probiotics [67] (both with a 7.32%, 611/8342 and 544/8342 of the mentions to food and diet terms). Although the characteristics of gluten sensitivity in IBD remain unclear, gluten is known to generate peptides that can alter intestinal permeability and affect the immune system [68]. Recent studies investigated the effects of a wheat gluten–containing diet on the evolution of sodium dextran sulfate–induced colitis [69] and evaluated the usefulness of a low fermentable oligosaccharides, disaccharides, monosaccharides, and polyols diet on patients with IBS, nonactive IBD, and celiac disease compared with a gluten-free diet [70]. Likewise, some prebiotics, such as germinated barley foodstuff, psyllium, or oligofructose-enriched inulin, might provide some benefit in patients with active ulcerative colitis or ulcerative colitis in remission [71]. Other studies suggest that the VSL#3 probiotic may be effective in inducing remission in active ulcerative colitis [72]. The high number of favorites, retweets, and hashtags reflect the interest of the BD community in knowing more about these dietary interventions and discuss the pros and cons of different diets and foods.
In general, drugs and nondiet-based treatments are less mentioned than the other categories. However, medical cannabis (36.38%, 760/2089 of the mentions to drug terms) and hypnotherapy (3.57%, 174/4864 of the mentions to treatment terms) raise some attention as an alternative, nonconventional treatments. Although these results are still inconclusive, recent studies show improvements in some BD-related symptoms. Medical cannabis is being tested for the treatment of gastrointestinal disorders such as abdominal pain and diarrhea. Experimental tests show that single ingredients from cannabis, such as tetrahydrocannabinol and cannabidiol, are responsible for these effects [73,74]. Conversely, the use of hypnotherapy is being validated in the treatment of gastrointestinal symptoms such as reducing fasting distal colonic motility or reducing systemic and rectal mucosal inflammatory responses [75]. Although, the results of using these alternative therapies remain unclear. As it stands, users show greater interest for cannabis-related posts, as denoted by the high number of tweet favorites and retweets as well as the inclusion of hashtags. In addition, users find these tweets interesting as they choose to disseminate them among their followers and make the topic indexation easier.

### Topics of Patient Communications

A closer look into the tweets posted by BD patients was relevant to better understand how patients feel about and deal with BD symptoms, associated diseases, changes in habits, and medication. This analysis considered only the tweets posted by user accounts identified as patients (58.63%, 7795/13,295 of the total of user accounts). Moreover, it was possible to identify 2800 females (35.92%, 2800/7795 of the user accounts identified as patients) and 3860 males (49.51%, 3860/7795 of the user accounts identified as patients). Typically, the 11,098 tweets (45.05%, 11,098/24,634 of the total of tweets) posted by patients expressed a negative sentiment (51.99%, 5770/11,098), but there were also positive tweets (31.99%, 3551/11,098) and some tweets with a neutral sentiment (13.99%, 1553/11,098). The high amount of negative opinions is explained by the fact that patients are known to use social platforms to vent out their emotions, namely, their frustrations when it comes to diseases that do not have a cure, such as in the case of IBD [76,77]. Table 2 describes the tweets of patients in terms of the recognized semantics terms and the tweet sentiment.

### Table 1. Top 25 mentioned terms sorted by the number of tweets.

<table>
<thead>
<tr>
<th>Term</th>
<th>Category</th>
<th>Number of tweets</th>
<th>Number of favorites</th>
<th>Number of retweets</th>
<th>Number of hashtags</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diarrhea</td>
<td>Disease</td>
<td>1009</td>
<td>487</td>
<td>271</td>
<td>45</td>
</tr>
<tr>
<td>Constipation</td>
<td>Disease</td>
<td>971</td>
<td>359</td>
<td>213</td>
<td>152</td>
</tr>
<tr>
<td>Pain</td>
<td>Symptom</td>
<td>701</td>
<td>682</td>
<td>267</td>
<td>112</td>
</tr>
<tr>
<td>Medical cannabis</td>
<td>Drug</td>
<td>630</td>
<td>1134</td>
<td>614</td>
<td>100</td>
</tr>
<tr>
<td>Gluten</td>
<td>Food and Diet</td>
<td>519</td>
<td>1952</td>
<td>191</td>
<td>62</td>
</tr>
<tr>
<td>Probiotic</td>
<td>Food and Diet</td>
<td>498</td>
<td>936</td>
<td>344</td>
<td>142</td>
</tr>
<tr>
<td>Celiac disease</td>
<td>Disease</td>
<td>376</td>
<td>542</td>
<td>221</td>
<td>16</td>
</tr>
<tr>
<td>Cancer</td>
<td>Disease</td>
<td>367</td>
<td>691</td>
<td>424</td>
<td>50</td>
</tr>
<tr>
<td>Anxiety disorder</td>
<td>Disease</td>
<td>338</td>
<td>724</td>
<td>269</td>
<td>52</td>
</tr>
<tr>
<td>Depression</td>
<td>Disease</td>
<td>283</td>
<td>1107</td>
<td>501</td>
<td>51</td>
</tr>
<tr>
<td>Bloating</td>
<td>Symptom</td>
<td>264</td>
<td>158</td>
<td>80</td>
<td>52</td>
</tr>
<tr>
<td>Fibromyalgia</td>
<td>Disease</td>
<td>213</td>
<td>362</td>
<td>161</td>
<td>74</td>
</tr>
<tr>
<td>Dietary supplement</td>
<td>Food and Diet</td>
<td>196</td>
<td>602</td>
<td>146</td>
<td>9</td>
</tr>
<tr>
<td>Arthritis</td>
<td>Disease</td>
<td>190</td>
<td>872</td>
<td>375</td>
<td>32</td>
</tr>
<tr>
<td>Myocardial infarction</td>
<td>Disease</td>
<td>188</td>
<td>122</td>
<td>77</td>
<td>17</td>
</tr>
<tr>
<td>Allergic hypersensitivity disease</td>
<td>Disease</td>
<td>179</td>
<td>471</td>
<td>241</td>
<td>22</td>
</tr>
<tr>
<td>Flatulence</td>
<td>Symptom</td>
<td>175</td>
<td>482</td>
<td>35</td>
<td>6</td>
</tr>
<tr>
<td>Abdominal pain</td>
<td>Symptom</td>
<td>172</td>
<td>801</td>
<td>129</td>
<td>6</td>
</tr>
<tr>
<td>Multiple sclerosis</td>
<td>Disease</td>
<td>172</td>
<td>480</td>
<td>243</td>
<td>15</td>
</tr>
<tr>
<td>Asthma</td>
<td>Disease</td>
<td>170</td>
<td>486</td>
<td>259</td>
<td>12</td>
</tr>
<tr>
<td>Obesity</td>
<td>Disease</td>
<td>152</td>
<td>370</td>
<td>186</td>
<td>24</td>
</tr>
<tr>
<td>Vitamin D</td>
<td>Food and Diet</td>
<td>148</td>
<td>96</td>
<td>49</td>
<td>26</td>
</tr>
<tr>
<td>Heart disease</td>
<td>Disease</td>
<td>148</td>
<td>289</td>
<td>142</td>
<td>33</td>
</tr>
<tr>
<td>Autistic disorder</td>
<td>Disease</td>
<td>144</td>
<td>435</td>
<td>210</td>
<td>29</td>
</tr>
<tr>
<td>Hypnotherapy</td>
<td>Treatment</td>
<td>131</td>
<td>210</td>
<td>131</td>
<td>29</td>
</tr>
</tbody>
</table>
Table 2. Distribution of the sentiment of tweets by semantic category.

<table>
<thead>
<tr>
<th>Category</th>
<th>Negative tweets, n (%)</th>
<th>Neutral tweets, n (%)</th>
<th>Positive tweets, n (%)</th>
<th>Total number of tweets, N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disease</td>
<td>5691 (54.00)</td>
<td>1686 (16.00)</td>
<td>3161 (30.00)</td>
<td>10,538</td>
</tr>
<tr>
<td>Symptom</td>
<td>1152 (67.02)</td>
<td>103 (5.99)</td>
<td>464 (26.99)</td>
<td>1719</td>
</tr>
<tr>
<td>Food and Diet</td>
<td>1305 (47.01)</td>
<td>222 (8.01)</td>
<td>1247 (44.98)</td>
<td>2772</td>
</tr>
<tr>
<td>Drug</td>
<td>237 (36.02)</td>
<td>118 (17.93)</td>
<td>303 (46.05)</td>
<td>658</td>
</tr>
<tr>
<td>Treatment</td>
<td>710 (43.99)</td>
<td>339 (21.00)</td>
<td>565 (35.01)</td>
<td>1614</td>
</tr>
</tbody>
</table>

An important part of the tweets expressing positive emotions was related to gluten-free diets and probiotic supplements. In this line, there were more positive tweets posted by females than males (57.98%, 2274/3922 of the female tweets against 51.99%, 2390/4597 of the male tweets). However, in general, the tweets containing mentions to foods and diets (Figure 7) showed a similar distribution in terms of negative and positive sentiments. The main reason was that the effects of certain food components (eg, fiber, iron, and magnesium) were perceived differently, depending on the state of the disease and other personal conditions of the patient.

Disease and symptom (Figure 7) were the semantic categories with the highest number of mentions in negative tweets. These tweets discussed the bad side of BD conditions, that is, symptoms such as pain, fatigue, and migraines along with the co-occurrence of other conditions, namely, depression, diarrhea, and fibromyalgia [64,78]. In contrast, several tweets containing mentions of drugs (Figure 7) showed positive emotions, namely, the tweets highlighting the benefits of medical cannabis for the treatment of different chronic diseases.

As a means to look into these tweets from another perspective, Figure 8 depicts a subgraph of the semantic co-occurrence network reconstructed from the patient tweets (see details on the topological analysis of the complete network in Multimedia Appendix 2). This subgraph shows the relations between diseases and drugs. The size of the nodes is based on the node degree (ie, bigger nodes represent the terms mentioned in more tweets), whereas the edge size expresses the strength of co-occurrence (ie, thicker edges represent a higher number of term-term occurrences). Red nodes represent drugs and yellow nodes represent diseases, whereas the edge color stands for the tweet sentiment (ie, red indicates a majority of negative tweets, green indicates a majority of negative tweets, and black indicates neutral sentiment). Although medical cannabis (and its components such as cannabidiol) was the most discussed drug, it was also possible to track down patient discussions about commercial drugs such as l-glutamine, lactulose, loperamide, and Plantago seed. For example, patients reported the positive effect of l-glutamine on diarrhea and muscular atrophy, but they also expressed their concern with glutamate leading to the occurrence of anxiety disorders. Glutamine is used to protect the mucous membrane of the esophagus and intestines and can boost immune cell activity in the gut [79,80]. Conversely, glutamine helps to create gamma-aminobutyric acid, a neurotransmitter that can stabilize the mind but can also produce glutamate, an excitatory neurotransmitter that can overstimulate the brain [81].

Figure 7. Examples of positive and negative patient tweets. (Left panel) Tweets related to foods and diets; (middle panel) tweets describing diseases and symptoms; (right panel) tweets mentioning drugs.

J Med Internet Res 2019 | vol. 21 | iss. 8 | e12610 | p.11
http://www.jmir.org/2019/8/e12610/

Pérez-Pérez et al

JOURNAL OF MEDICAL INTERNET RESEARCH

XSL-FO

RenderX
Figure 8. Subnetwork representing the co-mention of diseases and drugs in patient tweets. The size of the nodes is based on the number of tweets mentioning the term, whereas node color represents the corresponding semantic category. The edge color indicates the sentiment of the majority of the associated tweets.

Patients discussed the use of laxatives in treating constipation, but not all of the mentioned laxatives were recommended. Notably, lactulose is a sugar that cannot be digested in the gut and thus, tends to cause or aggravate IBS symptoms, such as gas, bloating, discomfort, and cramping [82]. In turn, loperamide was suggested as an effective astringent to treat diarrhea and constipation. Indeed, a previous study reported a significant improvement in stool frequency and consistency [83].

Finally, patients were interested in the beneficial healing properties of medicinal plants such as Plantago (in various forms, such as roasted seeds, decoction, or syrup), namely, anti-inflammatory, laxative, and astringent properties [84].

Temporal Analysis of User Activity

Table 3 describes the number of tweets and the volume of tweet interactions (ie, retweets and favorites) from February 1, 2018, to August 31, 2018. It was interesting to identify the periods of time when users are more likely to tweet and, in particular, how specific events, such as the IBD day, a scientific conference, or an informational campaign, could affect such activity.

In particular, the celebration of the World IBD Day (on May 19, 2018), which is a worldwide event to raise awareness about BD conditions and to urge governments and health care professionals to take action and show support to the sufferers, motivated an increase in tweet interactions, that is, retweets and favorites (ie, 15,820 and 18,158 tweet interactions in May and June 2018, respectively). The average number of retweets during these months increased by 174% (3314/1902) compared with the average number of retweets during the rest of the year. Regarding tweet favorites, the increment was still noticeable (329%, 13,585/4130) compared with the average number of favorites during the rest of the year.
Table 3. Monthly activity of posting and interaction during the analyzed period.

<table>
<thead>
<tr>
<th>Month</th>
<th>Number of tweets, n</th>
<th>Number of tweet interactions, n</th>
</tr>
</thead>
<tbody>
<tr>
<td>February</td>
<td>2930</td>
<td>4285</td>
</tr>
<tr>
<td>March</td>
<td>4239</td>
<td>5822</td>
</tr>
<tr>
<td>April</td>
<td>3992</td>
<td>6693</td>
</tr>
<tr>
<td>May</td>
<td>4202</td>
<td>16,506</td>
</tr>
<tr>
<td>June</td>
<td>3422</td>
<td>18,298</td>
</tr>
<tr>
<td>July</td>
<td>2366</td>
<td>7955</td>
</tr>
<tr>
<td>August</td>
<td>3952</td>
<td>4468</td>
</tr>
</tbody>
</table>

External Sources of Information

The hyperlinks shared via tweet provided useful information about current health research and development initiatives (public and private), health promotion actions, and other events that are promoted via Twitter as means to reach out to and engage more people. Tables 4 and 5 summarize the results obtained by grouping the URLs into 4 categories: (1) informative or awareness campaigns about BD; (2) scientific sites, articles, and conferences, usually covering new treatments; (3) health Web pages related to a specific disease; and (4) commercial sites. The #Tweets column indicates the number of tweets in which the URL is mentioned, the #Retweets column indicates the number of times that a tweet containing the URL was retweeted, and the #Favorites column indicates the number of times that a tweet containing the URL was selected as a favorite.

Only a small number of external links included in tweets posted by organizations got highly retweeted and labeled as favorite. Notably, the Guts4life Web page (a portal about IBD) got the highest number of retweets and favorites. This sort of analysis can be of aid in identifying the external sources that the community finds most useful/interesting, especially considering that most of them are related to pages describing BD symptomatology and potential treatments.

Regarding the external links included in the tweets shared by patients, their information flow was at the same level compared with that of links posted by organizations (an average of 49 retweets against an average of 49 retweets, respectively). In turn, the links shared by patients had a lower average number of favorites compared with those shared by organizations (ie, an average of 44 favorites against an average of 92 favorites, respectively).

Looking into the linked contents, most of the resources were related to posted articles that belonged to highly prestigious journals, namely, *Nature* and *British Medical Journal*, and reported recent research in BD topics (with no particular focus). The presence of commercial links to pages selling stoner- and other drug-related products that do not require a medical prescription was also noteworthy.

Table 4. Top 10 external sources of information mentioned in the bowel disease tweets posted by organizations. The URLs are sorted by the corresponding sum of the number of tweets, retweets, and favorites.

<table>
<thead>
<tr>
<th>Source of information</th>
<th>Category</th>
<th>Number of tweets, n</th>
<th>Number of retweets, n</th>
<th>Number of favorites, n</th>
</tr>
</thead>
<tbody>
<tr>
<td>Guts4life [85]</td>
<td>Health Web page</td>
<td>7</td>
<td>164</td>
<td>1918</td>
</tr>
<tr>
<td>Are Your Digestion Troubles Irritable Bowel Syndrome? [86]</td>
<td>Informative or campaign</td>
<td>5</td>
<td>128</td>
<td>231</td>
</tr>
<tr>
<td>About Crohn’s Disease [87]</td>
<td>Informative or campaign</td>
<td>1</td>
<td>69</td>
<td>63</td>
</tr>
<tr>
<td>Can You Treat Irritable Bowel Syndrome with Cannabis? [88]</td>
<td>Informative or campaign</td>
<td>4</td>
<td>34</td>
<td>84</td>
</tr>
<tr>
<td>How to Manage Irritable Bowel Syndrome with Your Brain [89]</td>
<td>Informative or campaign</td>
<td>3</td>
<td>30</td>
<td>77</td>
</tr>
<tr>
<td>Inflammatory Bowel Disease (IBD) [90]</td>
<td>Health Web page</td>
<td>10</td>
<td>45</td>
<td>51</td>
</tr>
<tr>
<td>Why I Get Excited When You Say You Know Someone With IBD [91]</td>
<td>Informative or campaign</td>
<td>3</td>
<td>27</td>
<td>65</td>
</tr>
<tr>
<td>New Treatment Options for Inflammatory Bowel Diseases [92]</td>
<td>Scientific</td>
<td>1</td>
<td>30</td>
<td>57</td>
</tr>
<tr>
<td>Chronic Inflammation [94]</td>
<td>Informative or campaign</td>
<td>5</td>
<td>48</td>
<td>26</td>
</tr>
</tbody>
</table>
Regarding more commercial drugs, patients expressed positive emotions with the usage of L-glutamine on diarrhea and muscular atrophy but also reported negative sentiments because of the production of glutamate and its influence on anxiety disorders. Another notable group of drugs in the discussion were the laxatives, for example, the usage of lactulose was associated with negative emotions because it tends to cause or aggravate IBS symptoms, whereas loperamide was noted to be an effective astringent to treat diarrhea and constipation.

Users were more active during and after the World IBD Day, which shows that these types of initiatives are raising public awareness about these diseases as well as indicates that social networks are part of the routine communication of the BD community. The external resources being shared in tweets by organizations aim to draw people’s attention to awareness/informational sites, whereas those shared by patients typically point to more scientific contents (eg, scientific articles on BD) and alternative treatments (such as cannabis).

**Limitations**

The most immediate limitation arises from the fact that the capture of raw data was carried out using the free Twitter API, that is, the identification of users that select one tweet as a favorite as well as the number of available tweets is restricted, with no assurance of a random or representative sample [105]. For this reason, it was not possible to perform a more exhaustive analysis of the social BD communities. Thus, a full data retrieval through automated dashboard vendors, or using a paid service of the Twitter API, may provide further insights.

It should also be noted that this study was based on the assumption that the data entered by Twitter users are true. It is not possible to detect if certain data (eg, the user profile picture or the user location) are reliable. This limitation impacts mainly the analysis of the demographic distribution and the conclusions inferred for a particular gender. That being said, the obtained

### Discussion

**Principal Findings**

The objective of this paper was to characterize and study the BD community on Twitter. To do so, a dataset of tweets related to BD was collected, processed, and analyzed. The dataset covered a consecutive period of 8 months, from February 1, 2018, to August 31, 2018. As a whole, this analysis provided new insights into 6 main questions: Which are the main communities and most influential users?; Where are the main content providers from?; What are the key biomedical and scientific topics under discussion? How are topics interrelated in patient communications?; How do external events influence user activity?; What kind of external sources of information are being promoted?

Health organizations and BD experts (eg, @CrohnsColitisUK and @IBDMD) were the users that received more tweets, typically looking for trusted information about the conditions. Patients shared experiences among themselves or asked for medical advice. Moreover, the most active users were located in the United States and the United Kingdom, which are among the demographic regions with highest BD prevalence.

Most of the tweets talked about BD symptoms, related diseases, foods, and diets. Specifically, diarrhea, constipation, and pain were the symptoms that raised more concern (in general as well as among patients), whereas gluten and probiotics were among the most discussed dietary interventions (including a high number of favorites and retweets). In this line, females showed higher positive emotions about these dietary interventions than males. Medical cannabis was the most commented drug, notably in the tweets that raised the highest number of favorites, and patients actively discussed the beneficial effects of cannabis (and its components) in mitigating common BD symptoms. Regarding more commercial drugs, patients expressed positive emotions with the usage of L-glutamine on diarrhea and muscular atrophy but also reported negative sentiments because of the production of glutamate and its influence on anxiety disorders. Another notable group of drugs in the discussion were the laxatives, for example, the usage of lactulose was associated with negative emotions because it tends to cause or aggravate IBS symptoms, whereas loperamide was noted to be an effective astringent to treat diarrhea and constipation.

Users were more active during and after the World IBD Day, which shows that these types of initiatives are raising public awareness about these diseases as well as indicates that social networks are part of the routine communication of the BD community. The external resources being shared in tweets by organizations aim to draw people’s attention to awareness/informational sites, whereas those shared by patients typically point to more scientific contents (eg, scientific articles on BD) and alternative treatments (such as cannabis).

### Table 5.

Top 10 external sources of information mentioned in the bowel disease tweets posted by patients. The URLs are sorted by the sum of the corresponding number of tweets, retweets, and favorites.

<table>
<thead>
<tr>
<th>Source of information</th>
<th>Category</th>
<th>Number of tweets, n</th>
<th>Number of retweets, n</th>
<th>Number of favorites, n</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ginger for Nausea, Menstrual Cramps and Irritable Bowel Syndrome [95]</td>
<td>Informative or campaign</td>
<td>1</td>
<td>62</td>
<td>122</td>
</tr>
<tr>
<td>Symptoms of Ulcerative Colitis [96]</td>
<td>Commercial</td>
<td>1</td>
<td>178</td>
<td>1</td>
</tr>
<tr>
<td>Medicinal Marijuana as a Treatment for IBD Inflammatory Bowel Disease [97]</td>
<td>Commercial</td>
<td>8</td>
<td>45</td>
<td>94</td>
</tr>
<tr>
<td>A Starbucks barista called 911 [98]</td>
<td>Informative or campaign</td>
<td>1</td>
<td>24</td>
<td>99</td>
</tr>
<tr>
<td>Advances in Inflammatory Bowel Disease Pathogenesis: Linking Host Genetics and the Microbiome [99]</td>
<td>Scientific</td>
<td>1</td>
<td>47</td>
<td>70</td>
</tr>
<tr>
<td>Fungal Microbiota Dysbiosis in IBD [100]</td>
<td>Scientific</td>
<td>1</td>
<td>47</td>
<td>70</td>
</tr>
<tr>
<td>Murine Colitis Reveals A Disease-Associated Bacteriophage Community [101]</td>
<td>Scientific</td>
<td>1</td>
<td>47</td>
<td>70</td>
</tr>
<tr>
<td>I Am LOVING These Probiotics! [102]</td>
<td>Informative or campaign</td>
<td>1</td>
<td>83</td>
<td>27</td>
</tr>
<tr>
<td>Effects of Prebiotics vs a Diet Low in FODMAPs in Patients With Functional Gut Disorders [103]</td>
<td>Scientific</td>
<td>2</td>
<td>28</td>
<td>56</td>
</tr>
<tr>
<td>Acute GI Bleeding [104]</td>
<td>Health Web page</td>
<td>2</td>
<td>31</td>
<td>50</td>
</tr>
</tbody>
</table>
results were in accordance with the common knowledge of these communities. Language is another aspect of analysis to take into consideration. This study was only focused on tweets written in English. If the applied techniques are extended to support a greater variety of languages, such as Chinese or Spanish, it may provide complementary findings.

Finally, this study was focused on Twitter. However, the set of social networks might be expanded to analyze a richer dataset from a wide variety of sources. Considering other studies in the literature [21,106,107], Facebook and Instagram would be also sources of interest, although public data access is greatly limited.

Conclusions and Further Research
In this study, tweets related to BD were analyzed to characterize the user community and the exchanged contents. According to the obtained results, it was possible to detect communities and to describe the most discussed topics among these communities. The large and increasing volume of tweets demonstrates that Twitter is becoming a space for online conversation about BD, namely, associated symptoms and alternative treatments. In addition, the location of users indicates that conversations are happening at a global scale and, motivated by this, health-related stakeholders are using the platform to reach out to a larger audience on a daily basis.

In terms of future research, it would be interesting to perform user classification, that is, being able to identify experts, researchers, and companies, as well as patients. Thus, it would be possible to apply different sentiment analysis and TM approaches to the tweets to explore user-specific motivations, questions, and concerns. For example, it would be interesting to discover the opinion of patients about different treatments and specific symptoms.

Acknowledgments
SING group thanks CITI (Centro de Investigación, Transferencia e Innovación) from the University of Vigo for hosting its information and technology infrastructure. This study was partially supported by the Consellería de Educación, Universidades e Formación Profesional (Xunta de Galicia) under the scope of the strategic funding of ED431C2018/55-GRC Competitive Reference Group, the Portuguese Foundation for Science and Technology (FCT) under the scope of the strategic funding of UID/BIO/04469/2013 unit and COMPETE 2020 (POCI-01-0145-FEDER-006684). The authors also acknowledge the PhD grants of MPP and GP-R, funded by the Xunta de Galicia.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Topological analysis of the co-occurrence network of terms for all types of users.
[PDF File (Adobe PDF File), 369KB - jmir_v21i8e12610_app1.pdf ]

Multimedia Appendix 2
Topological analysis of the co-occurrence network of terms for patient users.
[PDF File (Adobe PDF File), 263KB - jmir_v21i8e12610_app2.pdf ]

References


Abbreviations

API: application programming interface
BD: bowel disease
Edited by G Eysenbach; submitted 30.10.18; peer-reviewed by J Groshek, Z He; comments to author 13.12.18; revised version received 23.01.19; accepted 26.04.19; published 15.08.19.

Please cite as:

©Martín Pérez-Pérez, Gael Pérez-Rodríguez, Florentino Fdez-Riverola, Anália Lourenço. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 15.08.2019. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.