# Twitter data analysis for studying communities of practice in the media industry☆,☆☆

Marlen Komorowski[a,*], Tien Do Huu[b], Nikos Deligiannis[b]

[a] *Vrije Universiteit Brussel Imec, Studies on Media, Information and Telecommunication (SMIT), Pleinlaan 2, B-1050 Brussels, Belgium*

[b] *Vrije Universiteit Brussel Imec, Department of Electronics and Informatics (ETRO), Pleinlaan 2, B-1050 Brussels, Belgium*

**Abstract**

Today, more and more physical communities of practice, a concept that describes a group of people that share a passion and interact regularly at events to exchange knowledge, utilize social media, such as Twitter. Brotaru, for instance, is such a physical community of practice for media professionals in Brussels. It is a monthly meet-up of videogame developers in various locations in Brussels. Furthermore, Twitter becomes widely acknowledged as important instrument for learning and community formation in the virtual world. But, do these communities of practice use Twitter only to promote their physical activities of learning? Or, are the activities of the physical communities further extended into the virtual world meaning that virtual communities of practice emerge from them? This article suggests a novel mixed-methods approach based on qualitative and quantitative data to measure the role of Twitter for physical communities of practice. The method applies different statistical measures and analysis on harvested Twitter data and additionally brings two of the most used methods in Twitter analysis together, social network analysis and text data analysis (a.k.a., content analysis). Four different communities of practice in Brussels' media industry and their activities and followers on Twitter have been analysed. The findings showed that the activities of the communities of practice extend into the Twitter sphere as the online communities are characterised by (1) a shared domain, (2) a lively community and (3) shared practices. The analysis further revealed that Twitter offers three main opportunities for the activities of communities of practice: it offers geographical extension; it gives temporal autonomy; and, it can be used to diversify the practices.

*Keywords:* Twitter, Community of practice, Media industry, Learning, Mixed-methods, Topic modelling

## Acknowledgements

The research was conducted within the Media Clusters Brussels (MCB) project (`www.mediaclusters.brussels`) and financed by Innoviris under the Anticipate programme (Prospective Research Anticipate 66 2014/2018) and supported by FWO (Grants G0A2617N and G025615).

## Suggested Citation

Suggested Citation: Komorowski, M., Huu, T.D. and Deligiannis, N., Telematics and Informatics (2017), `http://dx.doi.org/10.1016/j.tele.2017.11.001`

☆Preprint version of online publication in Telematics and Informatics.

☆☆`https://doi.org/10.1016/j.tele.2017.11.001`

Received 22 June 2017; Received in revised form 22 September 2017; Accepted 1 November 2017;

Available online 9 November 2017 

*Corresponding author

*Email addresses:* `mkomorow@vub.be` (Marlen Komorowski), `thdo@etrovub.be` (Tien Do Huu), `ndeligia@etrovub.be` (Nikos Deligiannis)

## 1. Introduction

The social learning approach describes the idea that there are other ways to learn, receive and share knowledge than in schools during classes or lectures, or within companies during trainings. Wenger (1998) based the concept of communities of practice on this idea. He defines them as "groups of people who share a concern or a passion for something they do and learn how to do it better as they interact regularly" (Wenger, 2006, p. 1). Wenger (2006) highlights communities such as a band of artists seeking new forms of expression, a group of engineers working on similar problems, a clique of pupils defining their identity in the school, a network of surgeons exploring novel techniques, a gathering of first-time managers helping each other cope and other groups that meet and interact regularly in different settings and forms.

The development of communities of practice can also be observed in the media industry. Brussels-based media professionals have several possibilities to participate in physical events of communities of practice because Brussels is a major media cluster, an agglomeration of media activities (Komorowski, 2016), of local and international media (Komorowski, 2017). Brotaru, for instance, is such a physical community of practice for media professionals in Brussels. It is a monthly meet-up of videogame developers in various locations in Brussels.

Besides communities of practice, more recently new information technologies are acknowledged as tools for transforming how people share their knowledge and interact with each other and Twitter has been recognised as tool for such transformation (Newgarden, 2009). The needs and goals of users have therefore also shaped Twitter from a "simple status updating platform into a versatile networking and learning tool" (Newgarden, 2009). While there is plenty of research on communities of practice in the media industry (e.g. Grabher, 2002; Roberts, 2006; Schmitz Weiss and Domingo, 2010) only very recently was Twitter considered in research as a tool for communities of practice (Gilbert, 2016). As argued, Twitter has become an important instrument for learning and forming communities in the virtual world. Therefore, the following questions rise: Do communities of practice use Twitter only to promote their physical activities of learning? Or, are the activities of the physical communities further extended into the virtual world meaning that virtual communities of practice emerge from them? To summarize: what role plays Twitter for physical communities of practice and how can we test this?

This article suggests that qualitative approaches combined with data-driven analysis are necessary to understand the extension of physical communities of practice into the virtual world. As such, we seek to develop new forms of data analysis and visualization that can prove and measure such extensions. The analysis is built on a mixed-methods approach combining qualitative data and data harvested from Twitter. Several techniques are applied including statistical measures, tweet text and metadata analysis, and network analysis. This article analyses four different physical communities of practice in the media industry in Brussels and their activities and followers on Twitter. Our hypothesis is that social media platforms give new opportunities to physical communities of practice through virtual communication tools. Furthermore, communities of practice that are both virtual and physical can alter the patterns of activities and prompt the organizers to make valuable changes to their activities. There is a pertinent need to be able to measure what role Twitter plays for physical communities of practice because of several reasons. First, of all, there has not been any research yet bringing Twitter analysis and the communities of practice approach together. Existing analysis methods do not fit for this purpose and little knowledge on the matter exists. Second, while Twitter is widely used by physical communities of practice, organizers of the communities have no insights on their Twitter activities at the moment (as no measurement techniques exists). Insights from such analysis can be used by organizers to modify their activities on Twitter and also by the physical communities to better serve the needs of their members and the goals of the communities of practice.

The article is structured as follows. In Section 2, we give insights into background information, which delineate what communities of practice and their characteristics are, introduce the analysed communities of practice in Brussels' media industry, and describe how Twitter can be facilitated as a research tool. In Section 3, we describe in detail the method applied for the analysis and the data retrieved. In Section 4, we present and discuss the findings of our analysis. Finally, in Section 5, we elaborate on implications and considerations of the findings and the developed analysis methods.

## 2. Background

### *2.1. Communities of practice in the media industry*

The concept of communities of practice has its origin in the work of Wenger and Lave (1991). They noticed that workers are voluntarily sharing knowledge outside of the usual framework (such as classes, lectures, trainings, workshops, etc.) and describe this process of learning as not vertical but horizontal (in contrast to a top-down teacher to student approach). Plazy and Patriarche (2015) summarize existing research on communities of practice and relate the concept to Brussels' media industry. This article extends their elaborations and employs Wenger's approach (1998), who argues that there are three crucial common characteristics for communities of practice (Wenger, 1998; Wenger, McDermott, and Snyder, 2002):
*(1) A shared domain*: The members of a community of practice have a shared domain of interest, to which they are committed. They also have shared competences that distinguish them from other people and a certain level of expertise. In the case of communities of practice in the media industry in Brussels the shared domain can be based on the media industry, in which the members have their (a) interest, the (b) geographical domain of Brussels and the shared (c) language.
*(2) A lively community*: Activity and interaction is at the core of the community. Having a profession does not make a community of practice unless members interact and learn together. In pursuing their interest in their domain, members engage in (a) frequent activity, (b) different kind of activities and (c) interactions.
*(3) A practice*: For Wenger (2006, p. 2), "members of a community of practice are practitioners. They develop a shared repertoire of resources: experiences, stories, tools, ways of addressing recurring problems - in short a shared practice". This practice can include three possible forms: (a) sharing of information, (b) planning of events and (c) asking for help or offering help to others.

Besides these common characteristics there are also many different forms of communities of practice. The literature usually identifies three main categories of communities of practice depending on their organizational form, including self-driven (informal and created by members), artificial (created by managers), and virtual communities of practice (which can be both, self-driven and artificial) (Cohendet, Roberts, and Simon, 2010; Snyder and Briggs, 2003; Roberts, 2006; Wenger, 2006). For the purpose of this research, we especially differentiate between physical communities of practice, that are formed through physical events, where members regularly meet (e.g. conferences, workshops, networking events, meetings, hackathons), and virtual communities of practice, that are formed online. To describe physical communities of practice in more detail, we can give several examples from Brussels' media industry, which are also the analysis objects of this study (see Section 3.1 for more information):

- *Brotaru*: Brotaru is a monthly meet-up of game developers in Brussels' cafés or restaurants and other various locations. The organizers are young game developers themselves and the events are open to everyone. Members have the opportunity to go to the events to present their on-going work, get feedback and network.

- *Café Numérique*: Café Numérique is a growing tech community that started in Belgium with a focus on organizing technology-related events and is now being held also in more than 20 different cities. Anybody interested in the concept can apply to host the events with focus on the entertainment sector but also other sectors, such as health. The shared domain is new technologies and experts in this domain are invited to participate and present at the events.

- *VRT Sandbox*: VRT Sandbox is a technology accelerator of the Flemish public broadcaster, VRT, in Belgium. The members of the accelerator can participate in workshops and have access to VRT's infrastructure and technological resources. The community of practice is open to the members of the invited accelerator projects and professionals using the co-working space.

- *UrLab*: UrLab is a "hackerspace" of the Université Libre Bruxelles (ULB). It is a community within the university and therefore organized by students and for students of the university with an interest in computers, electronics or technology. UrLab provides a place of facilities that are open for use and for the members to interact and collaborate. UrLab also organizes regular conferences and presentations.

As observed above, communities of practice can take various forms. With the upsurge of the new information technologies also virtual communities of practice get more and more prominence. Virtual communities of practice can be defined as self-organizing, voluntary, and open participation systems that are created and sustained through computer-mediated communications (Wasko, Teigland, and Faraj, 2009). However, it is uncertain what the role of Twitter for physical communities of practice is. Still, the influence technology has on the way people interact and share knowledge is apparent.

*2.2. Twitter as a Research Tool*

Indeed, the usage of such technologies also influences research. The rise of social media platforms, particularly Twitter, provides new opportunities to collect real-time data in large quantities directly from users (Abascal-Mena, Lema, and Sèdes, 2015). By the end of 2016, Twitter reached an average of 317 million monthly active users (Statista, 2017). On Twitter, users can write brief texts of a maximum length of 140 characters and can include links to photos, videos and websites and publish them online on their Twitter account. These messages are called tweets. Users can follow each other meaning that they receive the tweets of another user in their feed being able to respond to them, like them or re-share (re-tweet) them on their own account. Users of Twitter have usernames (using the symbol "@"), which can be used to identify users and to address them within tweets. Users can also use hashtags (using the symbol "#") before a relevant keyword or phrase in their tweet to categorize the content and make tweets more easily searchable.

Because of the high number of users, functionalities, facilitation through users and the open API, the application programming interface (which is much more restricted for other social media platforms), over the last view years, Twitter gained prominence as tool for research. Various studies, from different research fields, using different methods utilize Twitter as key element of the analysis:

- There are various research fields that consider Twitter. Bruns and Stieglitz (2013) summarize existing research on Twitter into the fields of political communication and activism (e.g. Larsson and Moe, 2011), crisis communication (e.g. Hughes and Palen, 2009), brand communication (e.g. Stieglitz and Krüger, 2011), engagement around shared experiences and audience research (e.g. Deller, 2011) and every day interpersonal exchanges (e.g. Yardi and Boyd, 2010). The fields of education and knowledge creation is not yet prominent. This article closes this gap and will combine the communities of practice approach with Twitter analysis.

- However, the formation of communities and networks on Twitter outside of the communities of practice approach has already been acknowledged widely. Honey and Herring (2009) showed that Twitter is not only used in one direction of communication but generally as a media for conversation. Huberman, Romero and Wu (2008) examined the social network of friends and follower relationships. Cunha et al. (2011) studied the propagation of hashtags within speech communities. Romero et al. (2013) studied how hashtags spread on a network defined by the interactions among users. Tiryakian et al. (2011) analysed the possibility that Twitter can form interlinked personal communities. It is surprising that while the community formation on Twitter seems widely acknowledged, the communities of practice concept is overlooked in that context. Gilbert (2016) was the first to bring the community of practice approach together with the Twitter sphere by analysing a community around a certain hashtag. The use of hashtags to conduct research on community creation on Twitter around specific issues is popular in research (Abascal-Mena et al., 2015). While data is most often collected through search terms (e.g. Hughes and Palen, 2009) or specific hashtags (e.g. Mendoza, Poblete, and Castillo, 2010), this study chooses a different approach and analyses communities on Twitter through the activities of certain community leading accounts. Studies that explore communities built on a network rather than topics are rare.

- Existing research on Twitter applies different analysis methods. Within these methods, besides presentation of statistics such as number of followers, re-tweets and more, we see two main streams, (1) content analysis and linguistics, and (2) social network analysis. On the one hand, content analysis is used to classify a given text (the tweets). This method can be based on different kind of analysis either by qualitative and manual coding (Naaman, Boase, and Lai, 2010) or through available software (Stieglitz and Dang-Xuan, 2012). Social network analysis on the other hand defines nodes representing a social actor, and the relationship between two nodes as expressed by edges. Through comprehensive mappings of links on Twitter, the aim of such studies is to describe networks through specific interconnections (Abascal-Mena et al., 2015). We suggest in this study a novel mixed-method that goes beyond existing literature and combines both qualitative data with quantitative data and not only content but also social network analysis. Our novel method of analysis has the aim to find new ways of representation of data and visualizations that help to understand if and how the practices of communities of practice are extended into the Twitter sphere.

## 3. Methodology and Data

### *3.1. Data Selection and Collection*

Such an aim - to study the extension of activities of communities of practice into Twitter - poses methodological challenges. We propose quantitative and qualitative methodologies in a mixed-methods approach. Several data sources were selected and different data collection methods were applied. These include (1) qualitative data collection methods and (2) quantitative data through Twitter crawling:

*(1) Qualitative data collection methods*: First, communities of practice in the media industry in Brussels needed to be identified. These were identified through a survey. The data was collected from July 2016 until January 2017. Media professionals in the Brussels' media industry were asked to identify communities of practice in Brussels in which they participate. In total, 19 potential communities of practice in Brussels were identified through the respondents ($n = 423$). The possible communities of practice were afterwards investigated through desk-research if they fulfil the following requirements: (a) they organize regular events, (b) these events happen in Brussels and are related to the Brussels' media industry and (c) they have a dedicated Twitter account that is active at the moment of the research (February 2017). Four communities of practice fulfilled these requirements. Observations and semi-structured interviews took place (from May 2016 to February 2017). Several events organized by the identified communities of practice were observed through active participation and semi-structured interviews were conducted with the organizers ($n = 4$). These interviews and observations had the purpose to better understand the structures of the identified communities of practice in Brussels' media industry. The results are displayed in Table 1.

*(2) Quantitative data through Twitter crawling*: We built a Twitter crawling engine to collect data from the identified Twitter accounts of the communities of practice using the Twitter REST APIs, which can return up to 3.200 of a users most recent tweets. We decided to employ the REST APIs, because we wanted to have an insightful view on the historical data of these accounts. The crawler, as described by Sechelea et al. (2016), runs a Python script that queries the Twitter REST APIs, and is able to parse the content of each response. For simplicity and for easy import to analytic tools, we stored the collected data using well-formatted text files. Each file contains plenty of data lines, and each line has multiple data fields separated by tab characters. The data retrieved integrated account information of the Twitter accounts and all its followers including the account screen name ("@"), the account description, the language and location of the account, the number of followers (the number of accounts that follow the account), following (the number of accounts the account follows) and the number of tweets and likes. This information was collected from 7.233 following accounts and from the four communities of practice accounts. Additionally, all tweets (original and re-tweets) from the four communities of practice (4.577 tweets) and all tweets of the followers ($\sim$7 million tweets) were retrieved including information on whether the tweet is original or a re-tweet, how often the tweet was re-tweeted and liked, which accounts were mentioned, which hashtags were used, the date and time tweeted and the category of the content (text, URL, photo, video, etc.). Duplicate tweets (that were re-tweeted) were removed, which resulted in more than 4 million original tweets. Finally, the "relationship" - actions between the accounts were retrieved that resulted in 410.160 connections. This includes 38.131 "like" actions, 135.292 "mentioning" actions and 63.635 "re-tweeting" actions. The information of the Twitter accounts of the communities of practice identified are summarized in Table 2.

Table 1: The identified communities of practice in Brussels' media industry and their offline facilitation (information gathered through desk research, observations and interviews).

| **Community of practice** | **Short description** | **Observation/ interview** | **Domain** | **Community/practice** |
|---|---|---|---|---|
| **Brotaru** | Networking event of game developers | Interview with founder | **Shared profession / domain:** Video games **Location:** Saint Gilles in Brussels **Language:** Not specified (official communication in English) | **Activities:** Monthly networking events in various bars / restaurants **Average number of attendants:** 25-35 |
| **Café Numérique** | Tech community in various cities | Interview with organizer | **Shared profession / domain:** Technology, entertainment, various **Location:** Various locations in Brussels **Language:** French | **Activities:** Regular events (workshops, networking, etc. in various bars / restaurants) several times per month **Average number of attendants:** 30-50 |
| **VRT Sandbox** | Technology accelerator of the public broadcaster | Interview with organizer | **Shared profession / domain:** Technology and audio-visual sector **Location:** Schaerbeek in Brussels **Language:** English (also Dutch) | **Activities:** Co-working space and regular events (workshops, getaways) about once a month **Average number of attendants:** 80-100 |
| **UrLab** | Hackerspace of a university | Interview with organizer | **Shared profession / domain:** Technology **Location:** Elsene in Brussels **Language:** French | **Activities:** Hackerspace space and regular events (presentations, conferences, workshops) several times per month **Average number of attendants:** 15-25 |

### *3.2. Proposed Mixed-Methods Approach*

The mixed-methods approach presented in this article reflects the communities of practice concept (see Section 2.1). The scraped data from Twitter needed to be cleaned, coded and parsed to establish the new theoretical framework for analysis. We apply (1) different statistical measures and analysis on the harvested Twitter data and additionally bring two of the most used methods in Twitter analysis together, the (2) social network analysis and the (3) text data analysis (a.k.a., content analysis).

*(1) Statistical measures and analysis*: Information about the followers of each communities of practice Twitter account and the tweets and information on the tweets for each communities of practice account were analysed with statistical measures. To understand the geographic distribution of Twitter accounts in the communities of practice, we looked at the location information indicated in the accounts' profiles. Technically, we have called Twitter REST APIs to obtain profile data of Twitter accounts where the data is available. The location information of the Twitter accounts needed cleaning as there is no restriction in how to make indications in Twitter. The location information given by the Twitter users had three possible forms. First, the information was clearly given and could be mapped by the used software directly. Second, the information was ambivalent and included for example spelling mistakes, only a city and not a country or several locations. In this case, the location information was manually cleaned. If there were several locations indicated, the first information given was chosen (e.g. for "Berlin & Dublin" Berlin was chosen). Third, the information was not clearly indicated (e.g. "Across Europe", "Cloud 9"). In this case, the information was not included in the analysis. The country and if possible the city was extracted from the given information on Twitter. Out of the 7.233 follower's accounts, 5.144 gave a usable indication for the location on Twitter. Statistical analysis and visualization of the data from the followers' accounts and the tweets was performed using the commercially available Tableau Software (`www.tableau.com`) that not only enables data visualization into graphs but also geographic mapping. Additionally, word frequency analysis

Table 2: The Twitter accounts of the communities of practice in Brussels' media industry (translation is provided by the authors).

| **Community of Practice and Twitter username** | **Description on Twitter** (partly translated by authors) | **No of followers** | **No of tweets** | **No of likes** | **Language** |
|---|---|---|---|---|---|
| **Brotaru: @otaru_bxl (Joined Twitter July 2014)** | BROTARU is a monthly meet up of videogame developers and passionates in Brussels! `facebook.com/bxl.otaru` | 412 | 727 | 79 | en |
| **Café Numérique: @CafeNBxl (Joined Twitter September 2009)** | Curious and passionate about new technology and innovation? `cafenumerique.org/bruxelles/` | 6.129 | 3.234 | 403 | fr |
| **VRT Sandbox: @VRTSandbox (Joined Twitter June 2015)** | VRT Sandbox is an international joint platform for collaborative innovation. Put together by VRT, EBU and iMinds. `sandbox.vrt.be` | 500 | 303 | 186 | en |
| **UrLab: @UrLabBxl (Joined Twitter March 2012)** | Laboratory for students and Hackerspace of the ULB (Université Libre Bruxelles) `urlab.be` | 292 | 313 | 7 | fr |
| **Total** | | 7.233 (100 shared) | 4577 | 675 | 675 |

was performed using Nvivo (`www.qsrinternational.com`), a software that is useful for data import, corpus handling, pre-processing, data management and for the creation of term-document matrices.
*(2) Social network analysis*: The aim of the social network analysis was to move beyond descriptive statistics and study the interaction between users, shedding light not only on the volume and forms of action, but also on how activities relate (or not) and therefore have an influence. In order to visualize the interaction between users, we compiled a list of followers and the communities' accounts that have been defined as nodes and a list of relations between the accounts and tweets defined as edges (for reasons of simplicity we only consider accounts as nodes that do belong only to one community). As discussed, interaction between accounts can happen through four different actions, namely, following, liking, re-tweeting and mentioning. In order to assess if the potential communities around certain Twitter accounts have a lively interaction, we needed first to assess if these actions (i.e., following, liking, re-tweeting and mentioning) can influence each other. This will show if any action of one account might actually lead to a re-action and therefore, interaction. To this end, for each follower account, we count the number of followers as well as the number of times the account is liked, mentioned or re-tweeted by other accounts. Suppose that the random variables $X_1$, $X_2$, $X_3$ and $X_4$ denote the number of followers, likes, re-tweets and mentions for a user. This means we have $n = 7.237$ samples for each random variable, with $n$ being the total number of Twitter accounts (4 main accounts and 7.233 followers). To measure the linear correlation between a pair of these random variables we consider the well-established Pearson correlation coefficient. For example, the linear correlation between $X_1$ and $X_2$ is given by:

$$\rho_{X_1 X_2} = \frac{\mathrm{cov}(X_1, X_2)}{\sigma_{X_1}\sigma_{X_2}} \tag{1}$$

In (1), $\mathrm{cov}(X_1, X_2) = E\{(X_1 - \mu_{X_1})(X_2 - \mu_{X_2})\}$ is the covariance between $X_1$ and $X_2$, $\mu_{X_1}$ (resp., $\mu_{X_2}$) is the mean value of $X_1$ (resp., $X_2$), and $\sigma_{X_1}$ (resp., $\sigma_{X_2}$) is the standard deviation of $X_1$ (resp., $X_2$). The value range for $\rho_{X_1 X_2}$ is between $-1$ and $+1$. The closer to $+1$, the more positive is the linear correlation. The closer to $-1$, the more negative is the linear correlation. A value close to zero means that the two random variables are not correlated.

In order to derive a unified metric for accounts' interactions, we put further emphasis on the actions of liking, mentioning and re-tweeting as they show strong interaction. Specifically, we add the frequencies of liking, mentioning and re-tweeting of an account $p$ towards a target account $q$ to express their interaction. Since the account $p$ interacts with a set of accounts, normalization is necessary, leading to the following expression:

$$\text{Interaction}(p,q) = \frac{f_{\text{rtwt}}(q)}{\sum_{i=1}^{\|\mathcal{P}_{\text{rtwt}}\|} f_{\text{rtwt}}(x_i)} + \frac{f_{\text{like}}(q)}{\sum_{i=1}^{\|\mathcal{P}_{\text{like}}\|} f_{\text{like}}(x_j)} + \frac{f_{@}(q)}{\sum_{i=1}^{\|\mathcal{P}_{@}\|} f_{@}(x_k)} \tag{2}$$

In (2), $\mathcal{P}_{\text{rtwt}}$, $\mathcal{P}_{\text{like}}$, and $\mathcal{P}_{@}$ are the sets of target accounts that are re-tweeted, liked and mentioned by the account $p$, and $x_i$, $x_j$ and $x_k$ are members of these sets. Moreover, $f_{\text{rtwt}}(q)$, $f_{\text{like}}(q)$ and $f_{@}(q)$ are re-tweet, like and mention frequencies of account $p$ towards account $q$.

To visualize the interactions inside a community and also between the four communities of practice, we can build weighted directed graphs from the Twitter accounts of these communities. A node in a graph corresponds to an account while an edge depicts the interaction (like, mention or retweet) between two accounts. The weight of an edge is equal to the value of the unified metric in (2). Furthermore, many useful properties of the graphs can be computed, revealing insightful facts about the communities. For instance, we can measure the relative importance of the accounts in the graph using the PageRank algorithm (Page, Brin, Motwani, and Winograd, 1999), which was originally used to rank web pages in the Google search engine. In PageRank, the rank of a page depends on the number of in-coming connections (other pages that direct to the page) and the rank of these referring pages. Concretely, the rank of a page is defined by:

$$Pr(u) = 1 - d + d \sum_{v \in B_u} \frac{Pr(v)}{L(v)} \tag{3}$$

where $u$ is the page of which the rank we wish to calculate and $v$ is the page having links to $u$. Furthermore, $B_u$ is the set of all pages with connections to $u$, $d$ is a damping factor, and $L(v)$ is the number of links from page $v$. Having built the graphs of Twitter accounts, we can directly apply the PageRank algorithm to identify the most important accounts in the communities. We used the weighted version of PageRank (Xing and Ghorbani, 2004) with the interaction's values forming the weights.

*(3) Text data analysis*: There has been a lot of linguistic research applied on tweets (e.g. Sumner, Byers, Boochever, and Park, 2012; Zappavigna, 2011). Many studies pre-process the language in tweets with sophisticated algorithms by applying grammatical rules and classifying tweets into categories (see Section 2.2). However, mood or sentiment analysis to detect meaning in tweets was not applied here as a lot of limitations have been found (Abascal-Mena et al., 2015). Instead, the goal of this analysis is to find the topics that are established in the communities on Twitter. By the term "topic", we mean a set of words that appear together frequently bearing a certain semantic. To this end, we employ Latent Dirichlet Allocation (LDA), a generative statistical topic model, which assumes that a document is a mixture of topics (where the mixture is drawn from a Dirichlet distribution) and each topic is drawn from a multinomial distribution across words (Blei, Ng, and Jordan, 2003).

Prior to analyzing content using LDA, we did text pre-processing, which is an important step before any downstream analysis, including text data analysis. We tokenized the text content of the collected tweets, namely we split the text into meaningful tokens. We employed the tweet tokenizer from the most common natural language processing library named NLTK (Bird, 2006; Perkins, 2010), because it is designed specifically for noisy text from Twitter. After that, stop words, which appear frequently without contributing much to the semantics of the text, were eliminated. Because the communities of practice are based on Brussels and its surroundings, we considered stop words in French, English and Dutch. For this task, we also used NLTK, which provides the lists of stop words in different languages.

The next step was to extract features from the corpus of the text. There are several methods for that purpose including: Bag-of-words (Sriram, Fuhry, Demir, Ferhatosmanoglu, and Demirbas, 2010), which constructs output binary vectors (0 or 1) or word-count vectors (frequency of words); Sequences (Salton and McGill, 1986) which include output vectors containing indexes of words in the vocabulary; and TF-IDF (Salton and McGill, 1986), that can be compared to the methods of bag-of-words, but adds weights to words. In this article, we decided to use the TF-IDF feature for topic modeling.

Technically, to find the topics, we used the LDA class provided by the scikit-learn (Pedregosa et al., 2011) Python library, a popular scientific package for data analysis and machine learning. The found topics were visualized using pyLDAvis (Karpovich, Smirnov, Teslya, and Grigorev, 2017), a package for visualizing topics on iPython notebook. The visualization is built on the classical multi-dimensional scaling

(MDS) (Cox and Cox, 2000), which assigns topic circles coordinates in the two-dimensional space such that their distances in the high-dimensional space are preserved. The distance between two topics is calculated using the Jensen-Shannon divergence (Lin, 1991).

It is known that in a corpus of documents, the most probable (frequent) words are not necessarily the most important ones. Specifically, more generic words may bring ambiguities into topics. To address this problem, Chuang Manning and Heer (2012) proposed a metric to measure the discrimination of words named Saliency. The saliency of a word is calculated based on its frequency and distinctiveness:

$$\text{distinctiveness}(w) = \sum_t p(t|w) \log \frac{p(t|w)}{p(t)} \tag{4}$$

$$\text{saliency} = p(w) \times \text{distinctiveness}(w) \tag{5}$$

In (4), $p(t|w)$ is the conditional probability of topic $t$ given word $w$, $p(t)$ is the marginal probability of topic $t$, which is the likelihood that a randomly-selected word is generated by topic $t$. In (5), $p(w)$ is the probability of word across the corpus.

The ultimate purpose of topic modelling is understanding the text using quantitative methodologies. Having found the topics containing lists of terms, one has to interpret the meaning of topics using terms from these lists. Normally, up to top thirty frequent terms are chosen for the interpretation (Sievert and Shirley, 2014). However, the common terms across the corpus dominate these lists making it hard to differentiate their meanings. To assess the relevance of terms for a given topic, we notice the fact that if a term appears very frequently across documents, that term does not contribute significantly to any specific topic. We therefore use the following formula from Sievert's work (Sievert and Shirley, 2014):

$$\text{relevance}(w|t) = \lambda p(w|t) + (1 - \lambda)p(w|t)p(w) \tag{6}$$

In (6), $\lambda$ is the weight parameter having value between 0 and 1. Setting $\lambda$ to 1 means we only consider the term's frequency for a given topic, while setting it to 0 results in penalizing a lot on very common terms. According to Sievert Sievert and Shirley (2014), the optimal value of $\lambda$ is 0.6, but fine-tuning might be needed to obtain better interpretations.

In the following chapter, we present the findings of our mixed-methods analysis. We look not only into the whole network and structure of the identified communities of practice on Twitter but also focus on the comparison between the communities. When relevant, certain communities of practice activities are visualized and discussed in detail.

## 4. Findings

### *4.1. The Domain*

In order to have a community of practice on Twitter, the members of the community need to have a shared domain of interest, to which they are committed (see Section 2.1). The analysis of the account descriptions based on word frequency analysis and statistical measures revealed that the followers share certain (1) interests, their (2) location and (3) language and therefore share a domain. Looking at the 50 most frequent words in the 7.233 account descriptions of the followers of the communites of practice's accounts, we see that in 4.614 times the interests of the account holders were expressed, in 971 instances the occupation is mentioned and 907 times a location is stated.

*(1) Interests*: The most shared interest is based on "technology" and includes terms such as "web", "digital", "innovation", "SEO", "data", "web design" and similar words. In total 1.688 times a related word to this interest occurred. Other interests that are shared include "design", "society", "creativity", "travelling" and "food". This is expressed in the account descriptions for instance, like this:

> *@JessicaVdGucht: "Communication professional with a passion for technology, travel and food" (follows @CafeNBxl).*

Besides technology as an interest, the followers also share an interest in certain domains, most significantly "media" but also "advertising" and "culture" among others. In total, "media" and similar words of media were mentioned 266 times. The followers are also interested in commercial activities including "business", "consultancy", "start-up" and more. If we look at the occupation, we find that most prominently we have followers that are "managers", "entrepreneurs" and "freelancers". Also, interesting is that in 907 instances, Brussels or Belgium is mentioned as location. For example:

*@kurt_vermeersch: "Globetrotter. Kinerd. TED-enthousiast. Concert-goer. R&D Manager @ Inteno Belgium" (follows @CafeNBxl).*

Looking at the frequent words within the descriptions of the followers' accounts, we can support that words related to technological topics are the most prominent. In total, the word "digital" has a percentage of 0,6 %, meaning that in all words used in all descriptions, 0,6% of the words is "digital" or similar, such as "digitalization". The hierarchy of the most frequently used words within the descriptions of the Twitter accounts and their average can be found in Table 3. Figure 1 visualizes the most frequently used words in the descriptions in the form of a word cloud. While these technological issues are re-occurring between all followers of each communities of practice Twitter account, we can also see that depending on which account they follow, certain topics that are highly relevant for the communities of practice are also re-presented in the interests expressed in the descriptions of the followers' Twitter accounts. For instance, the followers of the Twitter account of BROTARU, which is a monthly meet-up of videogame developers in Brussels, typically use "gaming" and related words (5,3 % of all words used). Other high ranked words and their related words include "development", "indie" and "gamedev". Followers of VRTSandbox on the other hand use words in their descriptions such as "broadcasting", "founder" and "innovative". This is not surprising as VRTSandbox is the start-up accelerator of the public broadcaster. Followers of UrLab's Twitter account use the words "ULB", "student" and "hacker" most frequently mirroring that UrLab is a community of practice, that is a Hackerspace at the university ULB.

Table 3: The most frequently used words in the descriptions of the followers of the analysed Twitter accounts and their percentage of occurrence.

| **Word** | **%** | **@otaru_bxl** | **%** | **@CafeNBxl** | **%** | **@VRTSandbox** | **%** | **@UrLabBxl** | **%** |
|---|---|---|---|---|---|---|---|---|---|
| digital | 0,6 | gaming | 5,3 | web | 0,7 | digital | 1,2 | makers | 0,6 |
| web | 0,6 | development | 1,8 | digital | 0,6 | media | 1,1 | Brussels | 0,6 |
| managing | 0,6 | designer | 1,2 | marketing | 0,6 | designer | 0,8 | developer | 0,5 |
| designs | 0,5 | artist | 1,0 | managing | 0,6 | managing | 0,7 | tweets | 0,5 |
| communicator | 0,5 | indie | 0,9 | communicator | 0,5 | technology | 0,7 | student | 0,5 |
| Brussels | 0,4 | video | 0,7 | designs | 0,5 | broadcasting | 0,6 | ulb | 0,5 |
| developing | 0,4 | dev | 0,5 | Brussels | 0,4 | creativity | 0,6 | science | 0,4 |
| media | 0,4 | music | 0,5 | tweets | 0,4 | innovative | 0,6 | web | 0,4 |
| tweet | 0,4 | digital | 0,5 | media | 0,4 | marketing | 0,6 | hacker | 0,4 |
| gaming | 0,4 | gamedev | 0,5 | sociales | 0,4 | founder | 0,5 | software | 0,4 |

*(2) Location*: Besides the word-frequency analysis, it is also possible to map the indicated location of the followers' Twitter accounts. In total 69,1% are located in Brussels. This is somehow surprising as the communities of practice only meet and are formed in Brussels. Still, three out of ten Twitter accounts are not located in the city. 82,0 % are still located within Belgium. Especially the followers of the Twitter account of VRTSandbox are not necessarily located in Belgium's capital. The reason is that Dutch is only spoken in 16 % of Brussels' households and most Flemish people working in Brussels, live actually outside of the city and commute to work (Janssens, 2008). Therefore, VRTSandbox has as many followers with accounts indicated in Ghent as in Brussels. Also, only every second follower of BROTARU's Twitter account indicate Brussels as location. Besides Belgium as location of the accounts, there are accounts in European countries such as France, UK, The Netherlands but also accounts all around the world. The distribution of accounts outside of Belgium is visualized in Figure 2.

*(3) Language*: The most indicated language of the follower's Twitter accounts is French (57,2 %) followed by English (35,4 %) and Dutch (4,7 %). BROTARU's and VRTSandbox's followers indicate however, English as their language most often. The Twitter accounts of these communities of practice are also in English and, at the gatherings, often English is the most prominent language. Therefore, this is also shared by the followers of the accounts. The distribution of accounts' location and indicated language is summarized in Table 4.


Figure 1: Word cloud of the most frequently used words in the descriptions of the followers of the analysed Twitter accounts based on their occurrence.

Table 4: The distribution of the location of the accounts in Belgium and Brussels and the indicated language of the followers of the analysed Twitter accounts.

| **Twitter account** | **% of followers in Belgium** | **% of followers in Brussels** | **% of accounts in fr** | **% of accounts in en** | **% of accounts in nl** |
|---|---|---|---|---|---|
| @otaru_bxl | 77,5 | 51,1 | 38,4 | 51,2 | 7,5 |
| @CafeNBxl | 82,9 | 72,9 | 63,4 | 31,4 | 1,6 |
| @VRTSandbox | 78,6 | 28,0 | 2,4 | 58,6 | 36,6 |
| @UrLabBxl | 79,3 | 81,1 | 44,9 | 46,6 | 3,4 |
| Total | 82,0 | 69,1 | 57,2 | 35,4 | 4,7 |

### *4.2. The Community*

A community of practice in the digital sphere, like Twitter, needs a lively community (see Section 2.1). In order for a community to be lively there needs to be (1) frequent activity, (2) different kind of activities and (3) especially interactions taking place. The analysis based on statistical measures and the social network analysis revealed that there are different and frequent activities taking place and that interaction happens on different levels. Therefore, a lively community is formed.

*(1) Frequency of activities*: In order to find out if frequent activities take place, we can first look at the tweets of the communities of practice, 4.577 tweets from all four accounts together. @CafeNBxl published the most tweets as the account already exists since 2009. The account @VRTSandbox only exists since June 2015. In order to compare the accounts' tweets, we therefore only consider tweets published during the last year, from 1 February 2016 to 1 February 2017, leading to a number of 1.076 tweets being analyzed. On average the communities of practices Twitter accounts tweet 0,8 tweets per day in this period. @CafeNBxl is the most active with on average 1,5 tweets per day and @UrLabBxl the least active with only about 3 tweets per month. During a day, the most favorite time to tweet by the accounts is 9:00 in the morning (9,9%). Most notably is that the accounts tweet at any day of the week and any time of the day. Even at


Figure 2: The distribution of the location of the accounts of the followers of the analysed Twitter accounts.

23:00 1,5 % of the tweets are tweeted.
*(2) Kind of activities*: Analyzing the tweets showed that only 52,1 % of all tweets are original tweets. The rest of tweets are re-tweeted content. @CafeNBxl posts the most original content with 73,2 % and @otaru_bxl the least with only 17,8 % of tweets being original. On average, every tweet of the accounts gets 0,8 re-tweets and likes. @VRTSandbox gets most interactions in likes with 1,1 likes per tweet. If we look at what kind of tweets are tweeted by the accounts we see that 31,4 % are pure text tweets. But the accounts also tweet photos, videos and animated gifs. Another way to classify the content of tweets is the existence of hashtags. More than half of all tweets of the accounts do not have a hashtag. When tweeting, the accounts can also interact with other Twitter accounts by mentioning other users. The accounts use this more often than hashtags as almost two out of three tweets mention one or more other users. Table 5 summarizes the findings based on the analysis of tweets for each communities of practice account.

Table 5: The analysis of the tweets of each communities of practice Twitter accounts from Feb 2016- 1 Feb 2017.

| **Twitter account** | **Tweets** | **Average tweets per day** | **% of original tweets** | **Average number of re-tweets per tweet** | **Average number of likes per tweet** | **% of tweets with only text** | **% of tweets that have no hashtag** | **% of tweets that do not mention another account** |
|---|---|---|---|---|---|---|---|---|
| @otaru_bxl | 315 | 0,9 | 17,8 | 0,8 | 0,7 | 38,7 | 41,0 | 7,9 |
| @CafeNBxl | 534 | 1,5 | 73,2 | 0,7 | 0,7 | 24,9 | 61,1 | 41,4 |
| @VRTSandbox | 200 | 0,5 | 50,0 | 0,8 | 1,1 | 38,0 | 48,5 | 22,0 |
| @UrLabBxl | 27 | 0,1 | 51,9 | 0,7 | 0,7 | 25,9 | 85,2 | 40,7 |
| Total | 1.076 | 0,8 | 52,1 | 0,8 | 0,8 | 31,4 | 53,4 | 28,0 |

*(3) Interactivity*: Even though we can show that there is activity on a Twitter account, we can even further explore the connections of the activities. The pure activity of one account is not enough to prove the existence of a community of practice in the Twitter sphere. Based on the Pearson correlation coefficient

Table 6: The Pearson correlation coefficient between interactions on Twitter of all the identified accounts.

| | **Follow** | **Like** | **Mention** | **Retweet** |
|---|---|---|---|---|
| **Follow** | 1 | 0,61 | 0,65 | 0,63 |
| **Like** | 0,61 | 1 | 0,75 | 0,73 |
| **Mention** | 0,65 | 0,75 | 1 | 0,90 |
| **Retweet** | 0,63 | 0,73 | 0,90 | 1 |

we can see that all actions on Twitter have a positive correlation with other actions, meaning that within the potential communities, interaction is taking place. Mentioning shows the strongest correlation with re-tweeting. Following only shows a moderate positive correlation with other metrics, as the Pearson correlation coefficients are always below 0.7 according to Ratner (2009). This shows that more means than simply following other accounts are necessary to promote interaction; especially, the mentioning of other accounts and the re-tweeting promote strong interactions. The results are shown in Table 6.

The next step is to go beyond the study of interaction and assess the importance of accounts within the network. Our assessment is based on a visualization of the nodes and the edges in the network of the communities of practice. Figure 3 shows this network of the followers of the four communities of practice. It is evident that @CafeNBxl has the largest community based on the number of followers. It is worth noting that within the network there are many interactions not only with the Twitter account of the communities of practice but also between the different following accounts. Interaction happens not only in one way, top-down (from the communities of practice‘s accounts to the followers) but also the other way around as well as between the communities' members. This supports the assumption that communities of practice are formed here on Twitter, because they are functioning horizontally and learning takes place between members and not only in a top-down relationship (from teacher to students) (see Section 2.1). Additionally, we can see that based on the size of some nodes, not only the communities of practice's Twitter accounts have an influential position, but also other accounts interact highly with the community, meaning that the four primary accounts are not necessarily the most important accounts in their ego network.

To exemplify this, Figure 4 shows the interactions in the community of @otaru_bxl's followers. We can see that there are some important accounts in the community. It can be seen that @otaru_bxl is not the most important account in its network. In fact, @FLEGAvzw, which is the official Flemish Games Association (FLEGA) supporting the Flemish games industry, is more important. Moreover, we can detect some other influential accounts in the network including @thehouseofindie and @andreadst.

### *4.3. The Practice*

A shared practice is the last characteristic a community on Twitter needs to have in order to be a community of practice (see Section 2.1). This practice can include three possible forms: sharing of information, planning of events and asking for help or offering help to others. Whether these practices are applied can be analysed by (1) looking at the tweets of the communities of practice's accounts or the (2) content of all tweets based on topic modelling. The analysis revealed that all three practices are applied by the communities of practice on Twitter but also by their followers.

*(1) The tweets of the communities of practices accounts*: The 4.577 tweets of the communities of practice have been analysed based on the 100 most frequent words appearing (after the text was cleaned from hashtags, users mentioned, and stop words). Within the 100 most frequent words, we then looked for words that fit into the three categories, namely, information, events, and help. We have found that many tweets can be identified based on frequent words that do share the practice of sharing of information, planning an event or asking and offering of help. A list of these words is given in Table 7. This can be exemplified on certain tweets. Within the 100 most frequent words, we have found many words that relate to time, including for instance "today", "tonight", "Wednesday", "tomorrow" and more. When such a word is used, the tweet most often aims at informing the readers about an event that is happening or will happen. Other words have been identified in the most frequent words that share information including questions such as "how" and demands such as "check out". More tweets could be identified that share information. For instance, the words "presentation", "project" and "article" are within the most frequent words. When they are used, the tweet aims at sharing information such as an article or a new project:

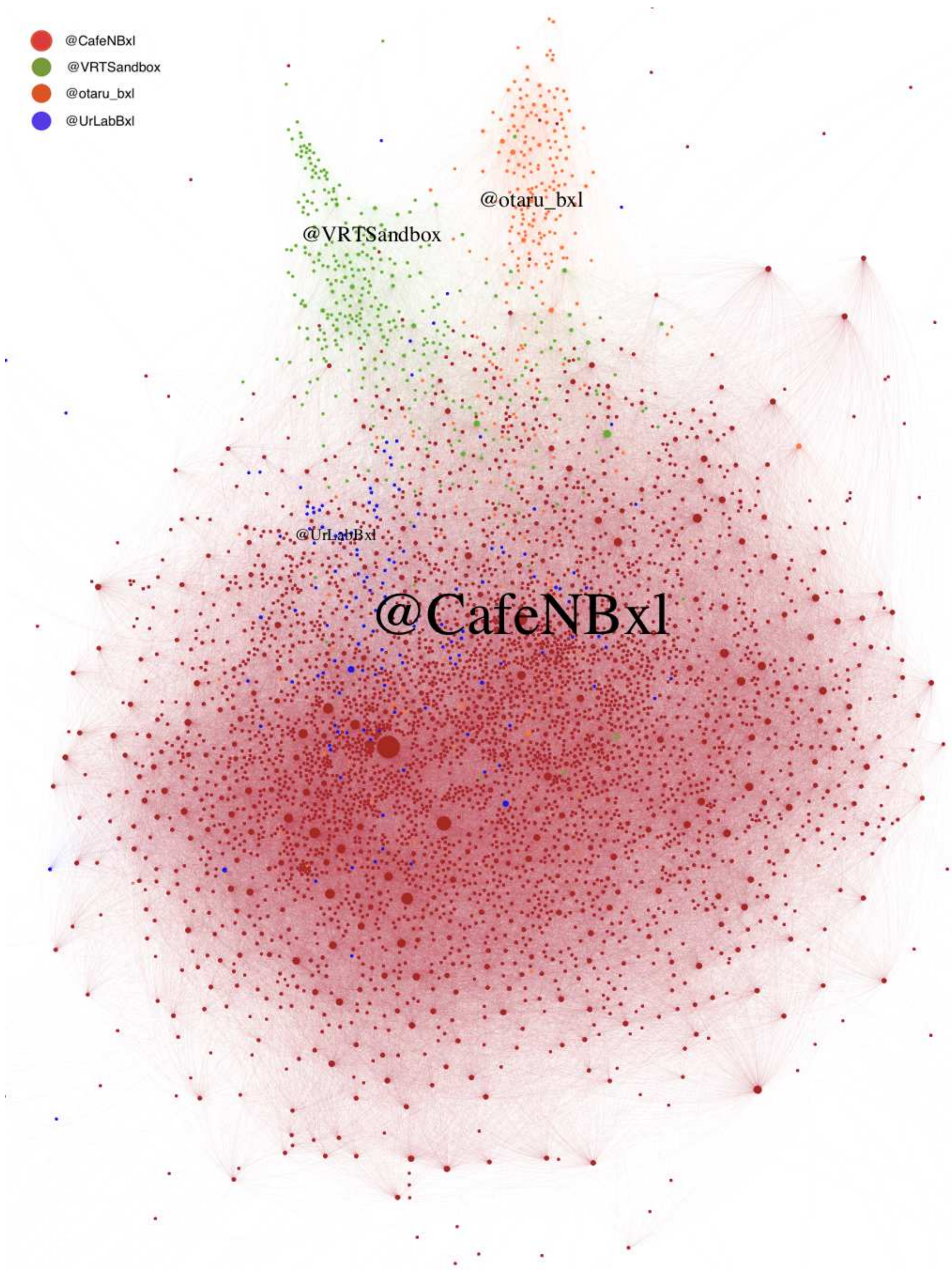


Figure 3: Visualization of the communities' networks on Twitter with a node representing a Twitter account and an edge denoting the interaction between two accounts (the size of the nodes corresponds to their importance over the network).

*@VRTSandbox: "Today at 10am CET: the multicam LiveIp debate, live, over IP. How crazy is that?..."*
*@CafeNBxl: "Missed the @AppsMarathon? Watch the videos of the presentations on our website"*
*"#OpenData: a researcher of the UCL published an article on geological data"*

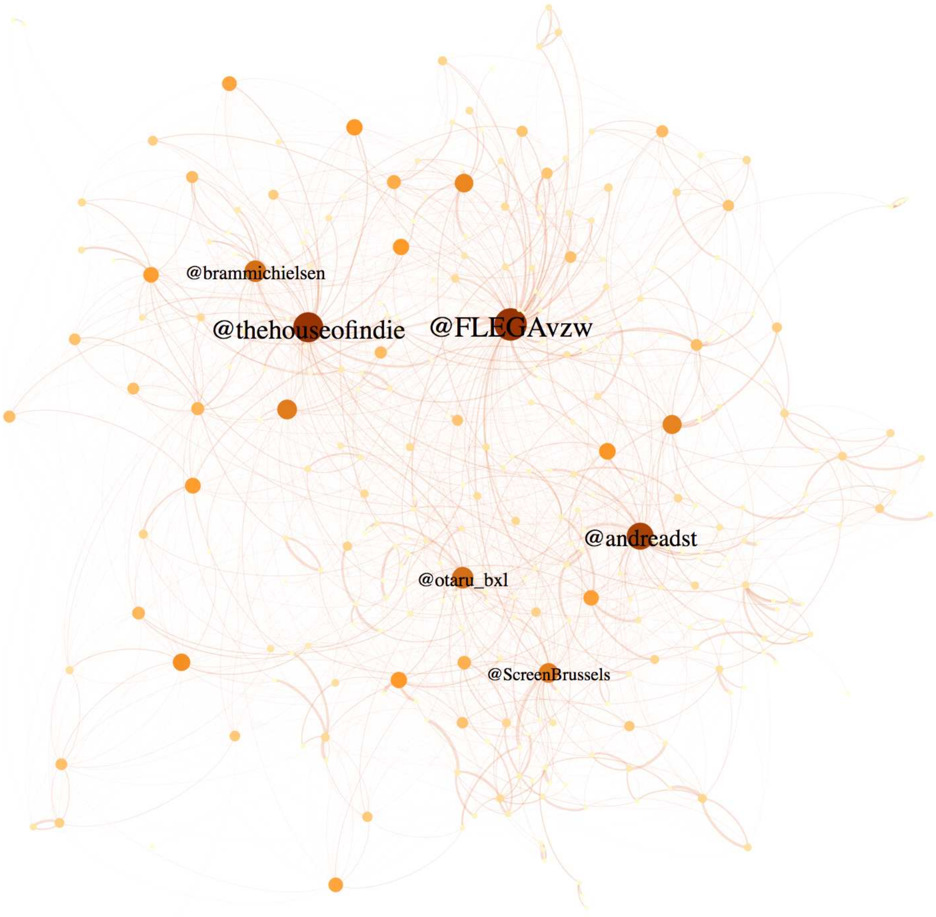


Figure 4: Visualization of the community of @otaru_bxl on Twitter with a node representing a Twitter account and an edge depicting the interaction between two accounts (the size of the nodes corresponds to their importance over the network).

Similar to the sharing of information within tweets, we could also identify frequent words that were used for planning events. Most of these tweets are not only used to plan events but also to inform. This includes again verbs such as "register", "attend", "meet", "welcome" and more. And finally, we could also identify words, that were used very often and that indicate that the tweet expresses the need for help. This includes words such as "searching", "looking for", "contact":

*@UrLabBxl: "There are still some places for the msp430 workshop this Saturday, do not hesitate to register"*

*@otaru_bxl: "Looking for 2d/3d graphic designers. Send us your resume"*

*"Hey Twitter! Screenshake needs your help! $1300 and 43 hours to go"*

*(2) Content analysis of all tweets in each community*: As a subsequent step in our analysis, we endeavour to find important topics that all followers and the communities of practice accounts discuss to see if the three practices are applied. Based on the topic modelling technique LDA introduced in Section 3.2, we have found three major topics that are discussed in all the communities. Table 8 shows the most salient terms that define each topic. We can see that Topic 1 is defined by the words "looking" and "need" with additional words that define the domain that people talk about such as "media". This shows that the first topic is about seeking help. Topic 2 is defined by words such as "discover" and "published" as well as words such as "press" and "journal". This indicates that within Topic 2, people discuss information they share. And lastly, Topic 3 is defined by words that request to "come" and "go" and a neighborhood in Brussels, "Ixelles". This indicates that people share and invite others within this topic to an event. Therefore, all

Table 7: Most frequent words used within the communities of practices' tweets in relation to a practice (words have been translated into English and the translation is provided by the authors).

| Information | Event | Help |
|---|---|---|
| "tonight", "Wednesday", "tomorrow", etc.<br>"presentation", "project".<br>"article", "info", etc.<br>"check out", etc.<br>"how", etc. | "register", "attend",<br>"welcome", "meet", "join", etc.<br>"debate", "presentation", etc. | "searching", "looking for", etc.<br>"contact", etc.<br>"please", "help", etc. |

practices of communities of practice are covered within the topics of the communities. Figure 5 depicts the distances among the three topics (denoted by circles) found in the corpus of tweets as well as their frequency of occurance of the topics. As described in Section 3.2, the visualization is made using the multi-dimensional scaling (MDS) method that is available in the Python package pyLDAvis. In Figure 5, the bigger the circle of a topic is, the more frequently the topic appears in the corpus of harvested tweets. Hence, tweets that seek help are the most prominent in the communities. Furthermore, the smaller the distance between topics is, the more similar these topics are. We can see that Topic 2 and 3 are more related to each other than to Topic 1. Namely, people are more likely to combine the sharing of information with invitations to events.

Table 8: Topics that relate to a practice extracted from the tweets of the communities of practice and their followers and the most relevant terms (with decreasing relevance level from left to right) (words have been translated partly into English and the translation is provided by the authors).

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **Topic 1: Help** | media | today | know | make | need | looking | work | people |
| **Topic 2: Information** | journal | new | social | press | discover | published | network | employment |
| **Topic 3: Event** | know | come | what | can | want | go | Ixelles | see |

If we look more closely at the different communities formed around the communities of practice accounts, we can see similar practices within the tweets. To exemplify, we can look more closely at the tweets of @VRTSandbox and its followers. Table 9 shows the three topics discussed in the tweets and Figure 6 shows the distances among the topics. @VRTSandbox's followers and their own account also tweet mostly around the topic of seeking for help with "looking" and "need" as the most salient terms in Topic 1. The salient terms in Topic 2 can be interpreted as the practice of planning and inviting for events with "come" and the Belgian city of "Gent". Topic three of the community of @VRTSandbox is depicted by the words "information" and content words such as "startup" and "data" showing that the community around @VRTSandbox is involved in innovation and the practices of the accelerator.

Table 9: Topics that relate to a practice extracted from the tweets of @VRTSandbox and its followers and the most relevant terms (with decreasing relevance level from left to right) (words have been translated partly into English and the translation is provided by the authors).

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **Topic 1: Help** | video | broadcast | production | looking | need | content | tech | network |
| **Topic 2: Event** | today | where | Gent | tomorrow | come | make | tonight | must |
| **Topic 3: Information** | take action | information | network | follow | data | IoT | business men | startup |

## 5. Implications and Final Considerations

Within this article, a novel mixed-methods approach was developed to measure the extension of communities of practice into the Twitter sphere. The approach integrated qualitative and quantitative methods

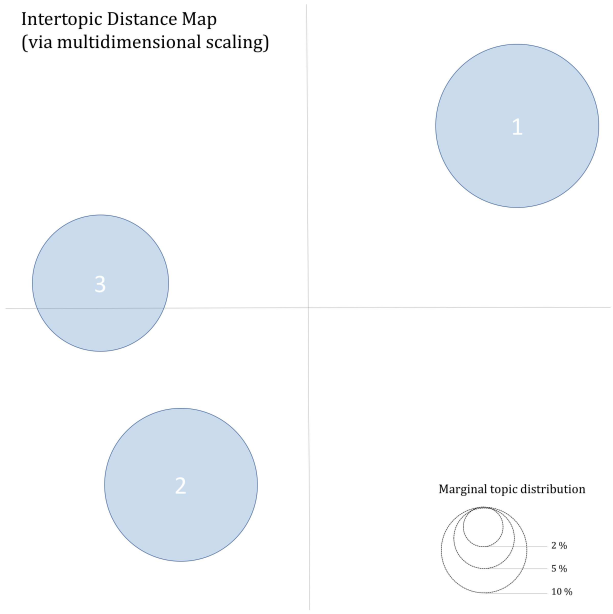


Figure 5: Visualization of the topics within the communities of practice using circles (The size of the circles are proportional to their prevalence, which is the marginal distribution across the corpus of tweets. Concretely, a 10% marginal topic distributon means that 10% of tweets are about this topic. The distance between two topics is expressed via the Jensen-Shannon divergence. If the distance between two topics is very small, we can observe similar words in these topics).

and extended existing research based on statistical measures, social network analysis, and content analysis. Following Wenger (1998), in order to have a community of practice in the virtual world, we suggested that three conditions need to be met: The Twitter community needs to have a (1) shared domain, (2) be a lively community, and (3) have shared practices. The mixed-method approach was applied on four Twitter communities that emerge around physical communities of practice in the Brussels' media industry. The research questions were: Do communities of practice use Twitter only to promote their physical activities of learning? Or, are the activities of the physical communities further extended into the virtual world meaning that virtual communities of practice emerge from them?

The analysis showed that members of the Twitter community indicate to (1) *share a domain*, as they have similar interests in technology and the media industry. The analysis also revealed that the members have interests in common that might not be so obvious from the beginning including interests in food and travelling or that they share certain occupations such as being an entrepreneur. The analysis further showed that all communities on Twitter are (2) *active and lively.* There are strong differences in how the communities of practice use Twitter in terms of frequency of tweeting and what kind of content they tweet. The social network analysis showed that the actions on Twitter have a strong correlation, thereby leading to interaction. The communities have highly influential members. This shows that the Twitter communities are horizontally organized, like the concept of communities of practice suggests. This means that the virtual communities are not only interacting top-down from the community of practice account to the followers but also horizontally, between followers. There are highly influential followers having leading roles in these communities. Finally, the analysis revealed that the tweeting on Twitter can be used to facilitate (3) *shared practices* including the sharing of information and new knowledge, the organization of events and the search for help. The topic modelling exercise revealed that within the communities of Twitter, the practice of seeking help is the most prominent. To summarize, the findings showed that within Twitter the communities meet all the three conditions, thereby leading to the conclusion that Twitter is used to extend the activities of the communities of practice into the digital world.

Based on the analysis, organizers of communities of practice can: (a) *Know their members better.* What

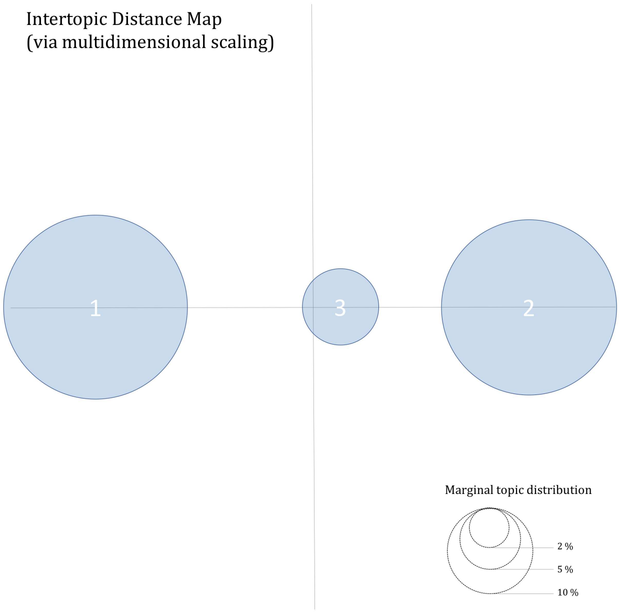


Figure 6: Visualization of the topics within the @VRTSandbox's Twitter community using circles (The size of the circles are proportional to their prevalence, which is the marginal distribution across the corpus of tweets. Concretely, a 10% marginal topic distributon means that 10% of tweets are about this topic. The distance between two topics is expressed via the Jensen-Shannon divergence. If the distance between two topics is very small, we can observe similar words in these topics).

interests do they have and what occupation? Where are they located and what languages do they speak? (b) *Understand their own behaviour on Twitter.* How often do I tweet and at what time and day? What kind of content could I tweet and how do I get the most feedback? (c) *Utilize the members with the most importance.* The analysis showed that often the Twitter account of the community is not even the most important member of the community. By knowing who these influential members are, collaborations can be established and more people reached. Twitter opens up many new opportunities for communities of practice. Eastmond (1998) highlights that an online environment can establish collaboration with others around a shared interest, interaction with issues that they deem to be important, application of their learning and the development of a sense of community. Based on the findings, we can establish three main opportunities that derive for communities of practice in the virtual world:

*(1) The geographical extension.* As shown in the analysis of the shared domain (see Section 4.1), the followers of the communities of practices Twitter accounts are not only located in Brussels but all over the world. While the physical events organized by the communities of practice are all taking place within Brussels, it becomes clear that only Brussels-based media professionals can participate regularly at the events and the knowledge exchange. However, the virtual world allows for an extension that is not limited by a physical location for the communities of practice anymore and activities such as sharing of knowledge, discussions and learning processes that are the goals of the communities of practice become dislocated into the virtual world. If you are interested in the activities of a certain community of practice, you can follow and engage through Twitter with this community even though you are not located where the events of the community take place. This allows communities of practice to extend their geographical reach and enrich their activities through world-wide members when using Twitter as a tool.
*(2) The temporal autonomy.* Physical communities of practice meet not only at a certain predefined place but also at a certain time slot. However, as shown in the analysis of the community in Twitter (see Section 4.2), the activity within the Twitter community is not restricted to certain times. The community of practice can tweet at any given time of the day and the followers can interact with the tweet, like, respond and re-

tweet it at any given time. This makes the activity of the community of practice on Twitter autonomous from time restriction. As media professionals have work obligations, they might not be able to attend the events organized by the communities of practice regularly, but the free access to the activities of them on Twitter, gives interested members the opportunity to interact and follow the community at any given time. Additionally, the member can interact with new knowledge with a time delay, which could be beneficial as it gives members the opportunity to "digest" and reflect on information before interacting with it.

*(3) The diversification of knowledge sharing.* Communities of practice highly rely on workshops and events. Invited speakers give presentations, the participants involve in discussions and face-to-face interactions. In the Twitter sphere pictures, videos, gifs, websites to more engaging content can be shared and new topics and practices developed (see Section 4.3). This not only allows discussions through responding and interacting with content but also gives the followers of the community of practice's Twitter account the possibility to first engage themselves with the content in detail and their own pace before interacting with it. Face-to-face contact is important in building networks and learning situations, however, Twitter and social media platforms in general are a great extension of such activities as they allow the diversification of how new knowledge can be shared and interacted with within the virtual sphere.

## References


Abascal-Mena, R., Lema, R., Sèdes, F., 2015. Detecting sociosemantic communities by applying social network analysis in tweets. Social Network Analysis and Mining 5 (1), 1–17.

Bird, S., 2006. Nltk: the natural language toolkit. In: Proceedings of the COLING/ACL on Interactive presentation sessions. Association for Computational Linguistics, pp. 69–72.

Blei, D. M., Ng, A. Y., Jordan, M. I., 2003. Latent dirichlet allocation. Journal of machine Learning research 3 (Jan), 993–1022.

Bruns, A., Stieglitz, S., 2013. Towards more systematic twitter analysis: metrics for tweeting activities. International Journal of Social Research Methodology 16 (2), 91–108.

Chuang, J., Manning, C. D., Heer, J., 2012. Termite: Visualization techniques for assessing textual topic models. In: Proceedings of the International Working Conference on Advanced Visual Interfaces. ACM, pp. 74–77.

Cohendet, P., Roberts, J., Simon, L., 2010. Créer, implanter et gérer des communautés de pratique. Gestion 35 (4), 31–35.

Cox, T. F., Cox, M. A., 2000. Multidimensional scaling. CRC press.

Cunha, E., Magno, G., Comarela, G., Almeida, V., Gonçalves, M. A., Benevenuto, F., 2011. Analyzing the dynamic evolution of hashtags on twitter: a language-based approach. In: Proceedings of the Workshop on Languages in Social Media. Association for Computational Linguistics, pp. 58–65.

Deller, R., 2011. Twittering on: Audience research and participation using twitter. Participations 8 (1), 216–245.

Eastmond, D. V., 1998. Adult learners and internet-based distance education. New directions for adult and continuing education 1998 (78), 33–41.

Gilbert, S., 2016. Learning in a twitter-based community of practice: an exploration of knowledge exchange as a motivation for participation in# hcsmca. Information, Communication & Society 19 (9), 1214–1232.

Grabher, G., 2002. The project ecology of advertising: tasks, talents and teams. Regional studies 36 (3), 245–262.

Honey, C., Herring, S. C., 2009. Beyond microblogging: Conversation and collaboration via twitter. In: System Sciences, 2009. HICSS'09. 42nd Hawaii International Conference on. IEEE, pp. 1–10.

Huberman, B. A., Romero, D. M., Wu, F., 2008. Social networks that matter: Twitter under the microscope.

Hughes, A. L., Palen, L., 2009. Twitter adoption and use in mass convergence and emergency events. International Journal of Emergency Management 6 (3-4), 248–260.

Janssens, R., 2008. Taalgebruik in brussel en de plaats van het nederlands. enkele recente bevindingen. Brussels Studies. La revue scientifique électronique pour les recherches sur Bruxelles/Het elektronisch wetenschappelijk tijdschrift voor onderzoek over Brussel/The e-journal for academic research on Brussels.

Karpovich, S., Smirnov, A., Teslya, N., Grigorev, A., 2017. Topic model visualization with ipython.

Komorowski, M., 2016. The seven parameters of media clusters: An integrated approach for local cluster analysis. International Journal of Media & Cultural Politics 12 (2), 171–191.

Komorowski, M., 2017. Report on data analysis: Brussels' media industry. Tech. rep., imec-VUB-SMIT.

Larsson, A. O., Moe, H., 2011. Who tweets? tracking microblogging use in the 2010 swedish election campaign. In: ECIS.

Lave, J., Wenger, E., 1991. Situated learning: Legitimate peripheral participation. Cambridge university press.

Lin, J., 1991. Divergence measures based on the shannon entropy. IEEE Transactions on Information theory 37 (1), 145–151.

Mendoza, M., Poblete, B., Castillo, C., 2010. Twitter under crisis: Can we trust what we rt? In: Proceedings of the first workshop on social media analytics. ACM, pp. 71–79.

Naaman, M., Boase, J., Lai, C.-H., 2010. Is it really about me?: message content in social awareness streams. In: Proceedings of the 2010 ACM conference on Computer supported cooperative work. ACM, pp. 189–192.

Newgarden, K., 2009. Annotated bibliography–twitter, social networking, and communities of practice. TESL-EJ 13 (2), 1–20.

Page, L., Brin, S., Motwani, R., Winograd, T., 1999. The pagerank citation ranking: Bringing order to the web. Tech. rep., Stanford InfoLab.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al., 2011. Scikit-learn: Machine learning in python. Journal of Machine Learning Research 12 (Oct), 2825–2830.

Perkins, J., 2010. Python text processing with NLTK 2.0 cookbook. Packt Publishing Ltd.

Ratner, B., Jun 2009. The correlation coefficient: Its values range between +1/1, or do they? Journal of Targeting, Measurement and Analysis for Marketing 17 (2), 139–142.
URL `https://doi.org/10.1057/jt.2009.5`

Roberts, J., 2006. Limits to communities of practice. Journal of management studies 43 (3), 623–639.

Romero, D. M., Tan, C., Ugander, J., 2013. On the interplay between social and topical structure. In: Seventh International AAAI Conference on Weblogs and Social Media.

Salton, G., McGill, M. J., 1986. Introduction to modern information retrieval.

Schmitz Weiss, A., Domingo, D., 2010. Innovation processes in online newsrooms as actor-networks and communities of practice. New media & society 12 (7), 1156–1171.

Sechelea, A., Do Huu, T., Zimos, E., Deligiannis, N., 2016. Twitter data clustering and visualization. In: Telecommunications (ICT), 2016 23rd International Conference on. IEEE, pp. 1–5.

Sievert, C., Shirley, K. E., 2014. Ldavis: A method for visualizing and interpreting topics. In: Proceedings of the workshop on interactive language learning, visualization, and interfaces. pp. 63–70.

Snyder, W., Briggs, X. N. D. S., 2003. Communities of Practice: A New Tool for Government Managers, November 2003 Series Collaboration. IBM Center for the Business of Government.

Sriram, B., Fuhry, D., Demir, E., Ferhatosmanoglu, H., Demirbas, M., 2010. Short text classification in twitter to improve information filtering. In: Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval. ACM, pp. 841–842.

Statista, jan 2017. Twitter: number of active users 2010-2016.
URL https://www.statista.com/statistics/282087/number-of-monthly-active-Twitter-users/

Stieglitz, S., Dang-Xuan, L., 2012. Political communication and influence through microblogging–an empirical analysis of sentiment in twitter messages and retweet behavior. In: System Science (HICSS), 2012 45th Hawaii International Conference on. IEEE, pp. 3500–3509.

Stieglitz, S., Krüger, N., 2011. Analysis of sentiments in corporate twitter communication–a case study on an issue of toyota. Analysis 1, 1–2011.

Sumner, C., Byers, A., Boochever, R., Park, G. J., 2012. Predicting dark triad personality traits from twitter usage and a linguistic analysis of tweets. In: Proceedings of the 2012 11th International Conference on Machine Learning and Applications - Volume 02. ICMLA '12. IEEE Computer Society, Washington, DC, USA, pp. 386–393.
URL http://dx.doi.org/10.1109/ICMLA.2012.218

Sylvain Plazy, G. P., 2015. Report on theoretical framework: Communities of practice in a media cluster context. Tech. rep., Media Clusters Brussels, dELIVERABLE 6.1.

Tiryakian, E. A., Gruzd, A., Wellman, B., Takhteyev, Y., 2011. Imagining twitter as an imagined community. American Behavioral Scientist 55 (10), 1294–1318.

Wasko, M. M., Teigland, R., Faraj, S., 2009. The provision of online public goods: Examining social structure in an electronic network of practice. Decision Support Systems 47 (3), 254–265.

Wenger, E., 1998. Communities of practice: Learning, meaning, and identity. Cambridge university press.

Wenger, E., 2006. Communities of practice-a brief introduction. retrieved february 26, 2010.

Wenger, E., McDermott, R. A., Snyder, W., 2002. Cultivating communities of practice: A guide to managing knowledge. Harvard Business Press.

Xing, W., Ghorbani, A., 2004. Weighted pagerank algorithm. In: Communication Networks and Services Research, 2004. Proceedings. Second Annual Conference on. IEEE, pp. 305–314.

Yardi, S., Boyd, D., 2010. Dynamic debates: An analysis of group polarization over time on twitter. Bulletin of Science, Technology & Society 30 (5), 316–327.

Zappavigna, M., 2011. Ambient affiliation: A linguistic perspective on twitter. New Media & Society 13 (5), 788–806.
URL https://doi.org/10.1177/1461444810385097